

JOURNAL OF INFORMATION SYSTEMS APPLIED RESEARCH

Volume 17, No. 2
August 2024
ISSN: 1946-1836

In this issue:

- 4. Insights for the next viral outbreak: An information systems applied research based on lessons from COVID-19**
Ivan D'Souza, Robert Morris University
Sushma Mishra, Robert Morris University

- 18. Decisional Guidance to Promote Motivation in Supply Chain Decision Making**
Russell Haines, Appalachian State University
Darin Hodges, Appalachian State University

- 31. Examining Factors Influencing the Acceptance of Big Data Analytics in Healthcare**
Abdul Sajid Mohammed, University of the Cumberland
Mary Lind, Louisiana State University in Shreveport

- 45. Tools for Success: Their Impact on Salaries in the Data Analytics Job Market**
Kathleen S. Hartzel, Duquesne University
Pinar Ozturk, Duquesne University

- 61. An Action Research Approach to Building an Enterprise-Specific Chatbot (ESCB)**
Zach Wood, University of North Carolina Wilmington
Geoff Stoker, University of North Carolina Wilmington

The **Journal of Information Systems Applied Research** (JISAR) is a double-blind peer reviewed academic journal published by ISCAP, Information Systems and Computing Academic Professionals. Publishing frequency is three issues a year. The first date of publication was December 1, 2008.

JISAR is published online (<https://jisar.org>) in connection with the ISCAP (Information Systems and Computing Academic Professionals) Conference, where submissions are also double-blind peer reviewed. Our sister publication, the Proceedings of the ISCAP Conference, features all papers, teaching cases and abstracts from the conference. (<https://iscap.us/proceedings>)

The journal acceptance review process involves a minimum of three double-blind peer reviews, where both the reviewer is not aware of the identities of the authors and the authors are not aware of the identities of the reviewers. The initial reviews happen before the conference. At that point papers are divided into award papers (top 15%) and other submitted works. The non-award winning papers are subjected to a second round of blind peer review to establish whether they will be accepted to the journal or not. Those papers that are deemed of sufficient quality are accepted for publication in JISAR. Currently the acceptance rate for the journal is approximately 35%.

Questions should be addressed to the editor at editor@jisar.org or the publisher at publisher@jisar.org. Special thanks to members of ISCAP who perform the editorial and review processes for JISAR.

2024 ISCAP Board of Directors

Jeff Cummings
Univ of NC Wilmington
President

Amy Connolly
James Madison University
Vice President

Eric Breimer
Siena College
Past President

Jennifer Breese
Penn State University
Director

David Gomillion
Texas A&M University
Director

Leigh Mutchler
James Madison University
Director/Secretary

RJ Podeschi
Millikin University
Director/Treasurer

David Woods
Miami University
Director

Jeffrey Babb
West Texas A&M University
Director/Curricular Items Chair

Tom Janicki
Univ of NC Wilmington
Director/Meeting Facilitator

Paul Witman
California Lutheran University
Director/2024 Conf Chair

Xihui "Paul" Zhang
University of North Alabama
Director/JISE Editor

Copyright © 2024 by Information Systems and Computing Academic Professionals (ISCAP). Permission to make digital or hard copies of all or part of this journal for personal or classroom use is granted without fee provided that the copies are not made or distributed for profit or commercial use. All copies must bear this notice and full citation. Permission from the Editor is required to post to servers, redistribute to lists, or utilize in a for-profit or commercial use. Permission requests should be sent to Scott Hunsinger, Editor, editor@jisar.org.

JOURNAL OF INFORMATION SYSTEMS APPLIED RESEARCH

Editors

Scott Hunsinger
Senior Editor
Appalachian State University

Thomas Janicki
Publisher
University of North Carolina Wilmington

2024 JISAR Editorial Board

Jennifer Breese
Penn State University

Dana Schwieger
Southeast Missouri State University

Wendy Ceccucci
Quinnipiac University

Anthony Serapiglia
St. Vincents College

Ulku Clark
Univ of North Carolina Wilmington

Michael Smith
Georgia Institute of Technology

Edgar Hassler
Appalachian State University

Jason Triche
University of Montana

Melinda Korzaan
Middle Tennessee State University

Karthikeyan Umapathy
University of North Florida

Lisa Kovalchick
PennWest University

Hayden Wimmer
Georgia Southern University

Li-Jen Lester
Sam Houston State University

Paul Witman
California Lutheran University

Jim Marquardson
Northern Michigan University

David Woods
Miami University Regionals

Muhammed Miah
Tennessee State University

Insights for the next viral outbreak: An information systems applied research based on lessons from COVID-19

Ivan D'Souza
isdst110@mail.rmu.edu

Sushma Mishra
mishra@rmu.edu

Department of Computer and Information Systems
Robert Morris University
Moon Township, PA 15108, USA

Abstract

The COVID-19 pandemic disrupted everything that people had taken for granted, specifically the freedom through which they could interact and engage with others. The COVID-19 vaccines provided the means to earn back the normalcy that people had lost. However not everyone was willing to receive the vaccine. Some wanted to take a chance on their health to wait and see. Others had other ways to defer the vaccine. One of the main reasons they credited for their hesitation was the communication process invoked by authorities at the onset of the pandemic and the uncertainty about the safety and effectiveness of the vaccines. The purpose of this information systems applied research paper is to conduct a phenomenological study and answer a research question that is geared toward understanding what vaccine hesitancy looks like in people with certain behavioral traits; some of these are established components of vaccine hesitancy and were known to underpin vaccine uptake during the 2009 H1N1 epidemic. Through the interview process and the ensuing thematic analysis, this paper hones into six themes with the intent to aid healthcare administrators and policy makers with a clearer understanding in planning effective campaigns against epidemics and pandemics in the future.

Keywords: COVID-19, vaccine hesitancy, vaccines, pandemic, qualitative analysis, phenomenology

Recommended Citation: D'Souza, I., Mishra, S., (2024). Insights for the next viral outbreak: An information systems applied research based on lessons from COVID-19, *Journal of Information Systems Applied Research*, 17(2), pp.4-17. <https://doi.org/10.62273/NWCI5504>

Insights for the next viral outbreak: An information systems applied research based on lessons from COVID-19

Ivan D'Souza and Sushma Mishra

1. INTRODUCTION

The COVID-19 pandemic has been an unprecedented experience for all. Since at its onset there were no viable vaccines for COVID-19, nations all over the world enforced lockdowns and social restrictions to curb the spread of the virus (Hale et al., 2021). Despite these measures the COVID-19 pandemic disrupted the way people lived, worked, and interacted with each other (Ciotti et al., 2020; Shieh-zadegan et al., 2021). People felt disconnected and isolated from the physical world and therefore relied on their social networks on social media to stay connected. As a result, digital activity surged (De' et al., 2020). In the United States of America (US or USA), the uncertainty of the virality of the COVID-19 infections and the constant pivoting of the federal and state governments on what should be done to curb this viral spread without an effective vaccine at the time, led to a deluge of conflicting information on social media (Reno et al., 2021). The World Health Organization (WHO) termed this phenomenon the 'infodemic' (WHO February 2020 Situation Report-13, 2020). While the world was struggling to keep pace with both the pandemic and infodemic, major pharmaceutical companies were in a race to be the first to produce an effective vaccine against COVID-19 (Lockey, 2020). Lopalco and Tan (2016) posited that vaccination is one of the main pillars of public health's response to a pandemic. Xia et al. (2020) supported vaccination as a safer way to accomplish herd immunity than building immunity from infection. According to the Mayo Clinic, once herd immunity is attained, the spread of the disease from person to person would become unlikely (Mayo Clinic Staff, 2022). Therefore, when the COVID-19 vaccines were rolled out, it was proposed that herd immunity in the US could be achieved if 90% of the population was fully vaccinated (Kricorian et al., 2022).

But, despite the wide availability of COVID-19 vaccines since November 2021, vaccine hesitancy, represented by those that were either avoiding or delaying COVID-19 vaccinations

(Khubchandani et al., 2021; Dubé & MacDonald, 2022), continued to represent a significant impediment to attain herd immunity in the US (Neely et al., 2022). In 2023, three years since the onset of COVID-19, life in the US for people has returned to some degree of normalcy, and the numbers of infections and deaths have reduced. But people are continuing to fall sick and die from COVID-19 (Centers for Disease Control and Prevention, 2023). As per the CDC through August 2023, due to COVID-19, hospitalizations had increased by 19% and deaths had increased by 18%. This showed that the drive to herd immunity through vaccination (Mayo Clinic, 2022) had still not achieved its goal. Also, there has been no letdown in social media activity with a userbase on social media platforms of 4.9 billion globally (Wong & Bottorff, 2023). Which begs the question; in 2023, are people in the US getting sick from COVID-19, due to their vaccine hesitant nature fueled by their activity on social media?

Blaming the spread of COVID-19 on the hesitant nature of humanity would be presumptive. Vaccine hesitant individuals are found in the middle of a continuum ranging from total acceptors to total refusers, and vaccine hesitancy could change over time (Larson et al., 2014; MacDonald et al., 2015, p. 4162). Vaccine hesitancy is not a recent phenomenon. People have demonstrated this behavior in the past toward other vaccines, for example: the routine influenza vaccine and the measles, mumps, and rubella (MMR) vaccines (Dubé et al., 2013). In 2012, the Strategic Advisory Group of Experts (SAGE) on Immunization identified three key factors among others that influenced vaccine hesitancy: confidence (do not trust vaccine or provider), complacency (do not perceive a need for a vaccine, do not value the vaccine), and convenience (access to a vaccine, barriers in acquiring a vaccine) (Larson et al., 2014, p. 2151). Based on their research on the 2009 H1N1 epidemic, Mills et al. (2020) found three additional factors: sources of information, social networks, and past vaccination behavior. Therefore, finding out what people are thinking now and what they plan to do in the future

based on their pandemic experiences would be invaluable for the US public health administration and the government in accommodating for vaccine hesitancy among the people in their methodical preparation against a future epidemic or pandemic.

The present study aims to explore vaccine hesitancy among people in the USA, based on their lived experiences during the COVID-19 pandemic. To achieve this aim, this study intends to answer the following research question (RQ): How does vaccine hesitancy induced by confidence, complacency, convenience, sources of information, social norms, and past vaccination behavior during the COVID-19 pandemic manifest in people in the USA?

2. LITERATURE REVIEW

Vaccine hesitancy

The World Health Organization (WHO) had estimated that vaccinations prevented at least 10 million deaths annually between 2010-2015 (Fridman et al., 2021). But vaccine hesitancy defined by the WHO as the “delay in acceptance or refusal of safe vaccines despite availability of vaccine services” and influenced by three key factors: vaccination confidence, vaccination complacency, and vaccination convenience (MacDonald et al., 2015) had continued to be a barrier to the effectiveness of vaccination programs globally (Betsch et al., 2018). Therefore, in 2019, a year before the COVID-19 crisis, vaccine hesitancy was listed among the top 10 health threats in the world by the WHO (Wilson & Wiysonge, 2020).

Vaccination confidence

The SAGE Vaccine Hesitancy Group defined vaccination confidence as “the trust in the effectiveness and safety of vaccines and in the system that delivers them, including the reliability and competence of the health services and health professionals and having trust in the motivations of the policymakers who decide which vaccines are needed and when they are needed” (WHO, 2014). Aw et al. (2021) evaluated COVID-19 vaccine hesitancy in high-income countries including the USA and found that the vaccine hesitant believed that COVID-19 vaccines were unsafe or ineffective. In 2020, the communication about the efficacy and safety of COVID-19 vaccines in the USA had to be executed under a cloud of uncertainty; people presumed that the vaccine development had a political impetus (“Operation Warp Speed”) in consideration of the upcoming Presidential

elections (Mills et al. 2020, p. 10). Troiano and Nardi (2021) analyzed the theme of vaccine hesitancy in the US during the COVID-19 pandemic and found that the vaccination hesitant had concerns that the vaccines were rushed, were too dangerous, and that the vaccines were useless because COVID-19 was harmless.

Vaccination complacency

For the SAGE Vaccine Hesitancy Group, vaccination complacency “exists where perceived risks of vaccine-preventable diseases are low, and vaccination is not deemed a necessary preventive action” (WHO, 2014). A study by Wolf et al. (2020) found that in March 2020 during the initial outbreak of COVID-19 in the USA, 25% of the participants were very worried about contracting the virus with around 13% not worried at all. For Fan et al. (2021), a perceived high risk of contracting COVID-19 was associated with individuals’ health-related attitudes and increased their likelihood in participating in preventive behaviors such as wearing a mask, washing hands, and keeping social/spatial distance. Aw et al. (2021) revealed that the vaccine hesitant believed that since they had already fallen sick from COVID-19 they were already immune from the disease from future infections.

Vaccination convenience

According to the SAGE Vaccine Hesitancy Group, vaccination convenience “affects the decision to vaccinate and is the quality of the service (real and/or perceived) and the degree to which vaccination services are delivered at a time and place and in a way that is considered appealing, affordable, convenient, and comfortable” (WHO, 2014). Aw et al. (2021) revealed that the differences in vaccine hesitancy rates across countries or regions in the world were linked to disparities in access, cost, and awareness of vaccines. The researchers also observed that the vaccine hesitant demonstrated a lack of trust in vaccination administrators.

Sources of information

Mills et al. (2020) presented multiple studies from several countries during the H1N1 pandemic to show the influence of sources of information on vaccination uptake. In one study they found that parents in the UK who watched national television news and in general proactively engaged in information-seeking behavior were more likely to vaccinate their children. In another study in the US, the researchers found that individuals who received their information about H1N1 from a health-care

provider or public health department were more likely to perceive the vaccine as safe. Furthermore, the surge in social media use during the COVID-19 pandemic (De' et al., 2020), and the resulting proliferation of anti-vaccination misinformation throughout social media, gave vaccine hesitancy a new urgency (Wilson & Wiysonge, 2020) for health care administrators and policy makers.

Social norms

A study found that parents who chose not to vaccinate their children had a much higher percentage of individuals (70%) in their social networks with similar attitudes than those who did vaccinate their children (13%) (Meleo-Erwin et al., 2017). Dubé et al. (2013) argued that the omnipresence of anti-vaccination content on the World Wide Web and Web 2.0 has allowed users to create and share this content using social networks. The researchers found that social norms of people who favored vaccine uptake intentions resulted in a collective stronger vaccine uptake behavior and concluded that social norms are a potentially powerful driver of vaccine acceptance. Even in communities that were isolated from media influence, social norms were an influential factor of vaccine acceptance as Henderson et al. (2008) found in their study of an orthodox Jewish community, in which word of mouth among the people was a potent source of rumors about vaccination dangers. Husain et al. (2021) and Yahaghi et al. (2021) supported this conclusion through their own findings that those around you, whom you respect, are being vaccinated themselves or having their children vaccinated, was a factor related to vaccine acceptance.

Past vaccination behavior

A systematic review of H1N1 vaccination uptake found that one of the strongest predictors for vaccination is past vaccination behavior (Mills et al., 2020, p. 12). Those who had previously been vaccinated against seasonal influenza were the most likely to opt for a pandemic vaccination (Bish et al., 2011; Jung et al., 2013). Past vaccine refusal, perceived risk of infection, and presence of comorbidities were confirmed as significant predictors of COVID-19 vaccine hesitancy (Reno et al., 2021).

This literature review showed that confidence, complacency, convenience, sources of information, social norms, and past vaccination behavior induced vaccine hesitancy among people during the COVID-19 pandemic and other outbreaks. Considering that herd immunity and vaccine hesitancy as an obstacle to it could be

issues to reckon with for generations, finding out what people think today about the vaccines and what they plan to do about it in the future could be leveraged by future research and/or future targeted public health messaging campaigns or strategies in the USA during future outbreaks of COVID-19 or other epidemics and pandemics.

3. METHODOLOGY AND ANALYSIS

Methodological approach

People experience life and share their experiences differently, whether as an individual or part of a family, group, community, or sub-culture. Qualitative research provides researchers with a lens to focus on how people live and experience the world by using one of its many approaches (Creswell & Poth, 2018). Phenomenology, a research design of inquiry, is a qualitative approach in which a researcher describes the lived experiences of individuals about a phenomenon, as described by participants of the study (Creswell & Creswell, 2018, p. 13). Phenomenology does not just focus on the life of an individual but rather on a research problem which focuses on a phenomenon and the essence of the lived experiences of people about that phenomenon (Creswell & Poth, 2018, pp. 121-122). The present study focuses on vaccine hesitancy as the phenomenon and adapts the design of phenomenological research using interviews to help elicit the essence of this phenomenon in the lived experiences of participants (Creswell & Creswell, 2018, p. 13).

Participant sampling and recruitment

The sampling frame of this study includes US residents, who were 18 years of age or older, used social media, and either directly or indirectly had experienced vaccine hesitancy. 15 participants were sampled through a combination of purposive and snowball sampling techniques. An initial set of five participants were selected from one of the researchers' social circle of friends. From these initial participants, referrals were obtained for ten additional participants who satisfied the requirements of the sampling frame. The participants were contacted for the interview either through email, a phone call, a mobile text message, or a personal face-to-face invitation. Through all these outreaches the purpose of the research and the interview, confidentiality of the interview process, and the duration of the interview were conveyed. Twelve of the fifteen participants agreed to a 'Google Meet' video-conference interview, and the remaining agreed to a face-

to-face meeting. The particulars of the meeting were emailed to the participants.

The interview

The researchers set up a free account with a third-party transcribing tool, 'otter.ai', for audio recording and transcribing the interview. At the beginning of the interview the researchers introduced themselves, and provided an overview of the research they were conducting, the format of the interview and its expected duration of forty minutes. The researchers also clarified the use of the transcription tool, 'otter.ai', to record the audio and transcribe the conversation. They assured confidentiality of the process and the measures they would take to safely and securely archive the information collected from the interview. Only after the participants had expressed their consent to continue and agree to the transcription of the interview, the researchers started the interview and activated the transcription of the interview. The interview protocol included in the Appendix was a semi-structured questionnaire made up of four demographic questions and 12 questions adapted from previous research. Each of the 12 questions attempted to gain a participant's perspective about an RQ component. For example: "Under what situations would you consider a COVID-19 vaccine and other vaccines in the future and why?" is an interview question that probed the participant's perception and future consideration or avoidance of vaccines. Based on how the conversation and the sharing of information progressed during the interview, the researchers decided on which questions would be used, altered, or skipped. At the end of the interview, the researchers thanked the participant for his/her participation and contribution to the study.

Reliability, validity, and ethical considerations

Credibility, transferability, dependability, and confirmability are four criteria through which trustworthiness of a qualitative study for accomplishing validity and reliability is assessed (Bell et al., 2022, p. 363-364). To attain credibility, the researchers reviewed the results of the study with three participants. To attain transferability the researchers included direct quotations from the interviews to highlight codes and themes. To attain dependability the researchers ensured that all data collected and generated was archived in a safe, secure, and accessible place. Finally, to attain confirmability the researchers appropriately documented all steps and conclusions to promote future audits of the study (Bell et al., 2022, p. 365).

Participation in the interview was voluntary. The researchers sent the transcription and recording to the participants who had requested a copy. The audio recordings were destroyed at the end of the study and all other data from the study would be retained for a maximum period of three years and destroyed afterward.

Data Analysis

The researchers used NVivo to organize and analyze the audio recordings and transcribed data collected by 'otter.ai' from each participant's interview. The researchers established classification sheets to organize the participants based on gender, marital status, level of education, and occupation. The researchers read and reread the text from the transcripts, and identified key words, combination of words, and sentences from each participant's transcript and from that process organized these into a set of thirteen unique categories of codes. Once organized into codes, the researchers repeated the review of the codes to form a unique set of six themes which formed the basis of the response to the research question and the findings of the overall study (Table 1).

4. RESULTS

15 participants, including five women and ten men with a lived experience of vaccine hesitancy, were chosen for the interview. There was one participant in the 18–29 age group, one in the 30–39 age group, five in the 40–49 age group, seven in the 50–59 age group, and two in the 60 years or older age group. Two were single and the remaining were married. One had high school education, two with some college, eight had undergraduate degrees, two with postgraduate degrees, and two had doctorates in the health care field. One participant was retired, another a home maker, three were business owners, while the rest were employed by others, with one in a part-time position. Among those employed, two were educators, two health care providers, one who worked in manufacturing, one in wealth management, one a software consultant, two in health care support, and one in religious ministry. The 15 transcripts recorded by 'otter.ai' yielded 555 significant sentences and phrases. The researchers arranged these sentences and phrases into clusters of 12 categories or codes aligned with the behavioral factors in the research question. A thematic analysis of these codes yielded six themes which are summarized in Table 1 and now discussed in more detail.

Theme 1: Natural immunity is more effective than vaccine induced immunity for the vaccine hesitant.

For most participants of the study, natural immunity was perceived as more effective than immunity induced by COVID-19 vaccines. As a participant quoted – “I’ve had COVID before. If I did get it my natural immunity would kick in and I wouldn’t be subject to it so hard if it was ever to happen again.” For another – “our immune system is designed so that when you do get a disease such as this, our bodies build up natural immunity to it. You know head colds are viruses. You don’t get the same head cold every year. You might get a head cold every year, but it’s not the one that you had last year or it’s a variant of the one that you had last year because our body has built up an immunity to that one.” Conversely there were participants who expressed their support for vaccines. For one – “I used to get so sick with the flu. Oh, I felt terrible for four or five days, but I went to work. I dragged myself. Then I got a flu vaccine. Guess what? I didn’t get the flu anymore.” And for another – “I’ve received a vaccine every time I go to my doctor. I get every vaccine that is offered. I trust my PCP.”

Theme 2: Freedom to choose to vaccinate or not is appreciated, but this was not the case in the past.

When the vaccines were rolled out in the US depending on their line of work people were mandated to vaccinate. A few took exception to this corporate strategy – “I think it’s just terrible, that you take away someone’s ability to provide for their family based on a decision that they make regarding their own health”, and “it should be your choice if you want to get it, but the mandate is the part that I had some concerns with.” Participants voiced their opinion that they would prefer to have the freedom to choose and enact their choice, as one quoted – “it should be up to ... the individual to choose whether they want certain shots or not” and another quoted – “all the freedom in the world ... to be able to choose whether you want to take the vaccine or not. I mean, that’s part of being an American.” Vaccination mandates were not enforced everywhere. According to one - “every place that I’ve gone the vaccine itself has been either optional or not required” and another “within our school it was strongly encouraged but it was not 100% mandated”. In 2023 vaccination mandates have either been relaxed or eliminated. One participant was grateful for her newfound freedom – “Right now I absolutely feel like I have complete freedom to get boosted or to continue along the path I am right now”.

Theme 3: Not all media is trustworthy, selectiveness of sources of COVID-19 information is prudent.

Since the onset of the COVID-19 pandemic in early 2020, people in the US were subject to an onslaught of information from the government, public health sites, and the news and social media. Due to the deluge of information, people who were not sure about being vaccinated found that the information was conflicting and inconsistently biased. For one participant – “I do find that most of the news is slanted, in my opinion ... political issues that the newspaper overwhelmingly pounds certain opinions, and I really don’t trust that.” For another – “there’s a lot of information that’s coming out recently that even executives who have made the COVID vaccine have come out and said, well vaccination really doesn’t restrict you from passing the disease on to somebody else.” Participants visited the CDC web site to see how things were progressing relative to the infections, fatalities, and vaccines but observed that information from certain qualified health care professionals, with alternative points of view, were blocked or suppressed. One shared – “Some of the medical professionals who are reliable were seen as unreliable by the media.” Based on their personal preferences, participants chose where they went to look for information about COVID-19. For one - “I like to look at the research. I like to get down into the sciences ... PubMed research, peer reviewed information”, and another “I look for sources of information like the CDC to help make that educated decision.”

Theme 4: Not everyone will believe in the same things. Diversity of thought should be respected.

The pandemic experience and the fallout from having to endure the social distancing, vaccination mandates, and differing opinions profoundly helped participants to respect the opinion of others. One quoted - “My son didn’t want to get ... vaccinated. It forced me to consider his perspective on it.” Participants also shared their interaction with family, friends, and acquaintances, and how these influenced their lived experiences about the pandemic. One said – “I have a brother-in-law who is a pharmacist, pushes for the vaccine, and we don’t. He called me and asked me if I would go for the vaccine. I said, no, for that was my choice. He let me be.” Another shared how a friend in the medical field provided him guidance in making a sound decision about his own health – “He’s very much an advocate for the vaccine. But he understood my reasonings as to not to get it. And he helped

#	Themes	Codes	RQ Factor	Times mentioned	Participants who mentioned it
1	Natural immunity or vaccine induced immunity	Health risks	Complacency	56	15
		Past vaccination	Past vaccination behavior	33	15
2	Freedom to choose to vaccinate or not	Freedom to choose	N/A	27	16
		Policy	N/A	32	14
3	Selectively trust media sources	Sources of information	Sources of information	153	15
4	Respect diversity of thought	Influence of people	Social norms	95	15
5	Past vaccination experience and full impact of COVID-19 vaccines	Access to vaccines	Convenience	19	12
		Efficacy of vaccines	Convenience	25	13
		Safety of vaccines	Confidence	49	15
		Speed to market	Confidence	22	13
6	High severity and fatality risk to vaccinate	Attitude	Vaccine hesitancy	26	13
		Will get the vaccine	Vaccine hesitancy	18	13

Table 1: Themes and Codes

me with extra precautions that I could take like, to wear a mask when I'm at work, and when I'm out in public places." A participant shared how she and her husband address their differences of opinion about vaccinating – "My husband and I agree to disagree about a lot of these things. He felt very strongly we should all be receiving these shots and I felt very strongly otherwise. So, we let our adult kids make their own decisions." A participant shared her opinion about the term "vaccine hesitant" – "I personally don't like the label vaccine hesitant. I feel like it diminishes my own intelligence and ability to do my own thinking to make decisions. I'm prochoice when it comes to vaccines. We're all about choice in every other aspect of life and death. So yeah, the term vaccine hesitancy, I think, isn't entirely accurate for me."

Theme 5: Past vaccination experience is key in determining future vaccination intention. Only time will tell the full impact of COVID-19 vaccines on the vaccinated.

Past poor and ineffective experience with vaccines affected the current stance of some of the participants of the study. One quoted – "I have received a few flu shots ... I had an adverse reaction to one ... so I don't get those anymore and that probably has affected my decision about the COVID vaccination." Participants

thought that the vaccines were rushed. For one - "I'm probably biased based on the sources that I have tended to pursue, but those have led me to believe that many of the trials were either rushed or some of the consequences were ignored or overlooked." Conversely for one - "The vaccine had already been made. So, I don't think it might have been rushed for the virus itself. Since the vaccine had been around long enough that it was perfectly safe to use and would not cause any adverse effects", and another expressed his confidence in vaccines in general - "No, I just like to say it's just my experience with how successful the flu shots have been." However, the future effects of the COVID-19 vaccines concerned a participant – "Just wonder in 10 or 20 years from now, if they ever find some type of tie back to heart attacks in young people, or maybe it's cancer or other issues ... I would say I worry about a little bit but at this point, you can't remove the vaccine. It's there, it has altered your DNA. And it is what it is." One recommended a study comparing the growth of children that have been vaccinated with those who have not been vaccinated. He believed that a lot can be learned from studies on the human anatomical effect of vaccines.

Theme 6: A higher fatality risk would be a factor to vaccinate for the vaccine hesitant. Those who rely on vaccines will continue to stay current on their shots when offered.

When participants were asked under which conditions they would consider being vaccinated or boosted at the end of their respective interviews, most of the participants thought that they would consider the vaccine only under life and death situations that affected someone at home or a close family member. One quoted – “if all of a sudden ... somebody near to me or in my house, you know, got cancer and they're undergoing chemo ... my thought process might change in that situation.” A participant was also looking at it from a severity or fatality perspective. According to this participant the rate of death due to COVID-19 in the US was under 2% and therefore was very low for him to consider the vaccine at this point. However, he would reconsider if the fatality rate would get higher than 20%. Other reasons that participants would consider a vaccine in the future were: a) getting vaccinated was the only viable option to build immunity against the disease, b) it was a must for their children to be vaccinated and they would also get it to support them, c) it was mandated at work, and d) if there was another lockdown and getting vaccinated was the way to get out of it. Some participants did not have any issues with the vaccination. They either were not against the vaccine or were for getting vaccinated whenever needed. One such participant quoted – “I am positive toward the vaccination ... I had all the confidence in the world that the COVID vaccine was nothing to be afraid of, that it was adequately tested, and that it should work and not harm me.”

5. DISCUSSION

Technology made the research process of inviting interview participation, conducting the interviews, and the analysis of the data collected from these interviews effective. Video conferencing through Google Meet made it possible for participants to be interviewed at their convenient time and place. Otter.ai automatically transcribed interviews and recorded the audio enabling researchers to stay focused on the interview instead of taking notes during the interview. Otter.ai was useful even for in-person interviews and those that required a phone call. NVivo, a qualitative analysis tool, helped the researchers in organizing the transcription and audio files from the interviews, classifying the content within these into twelve codes and six themes (Table 1). Based on these

six themes the researchers make the following six recommendations.

The unvaccinated were open to get the vaccine only if it was going to be a life and death situation close to home or the fatality rate was high. Some of the vaccinated were not sure if they would get the booster. They wanted to wait and see, since their experiences did not confirm the safety and effectiveness of the vaccine. Some of them had fallen sick from COVID-19 after they had been fully vaccinated and for others the side effects from the booster were severe. Conversely there were participants who wanted to follow the advice of their health care providers (HCP) and get the next dose whenever offered. Irrespective of their future considerations of whether to vaccinate or not, all participants wanted to continue to follow their self-prescribed precautions of good health and hygiene. They considered it their moral duty to protect and defend the wellbeing of their loved ones and the communities they lived and worked in. Public health administration and the government should account for this sentiment in their health crisis preparations to provide the needed resources to people who, despite being unanimous about not wanting to vaccinate, still have the innate desire to participate and contribute to the safety measures that are being put into place.

Most of the participants indicated that they had already fallen ill from COVID-19 and recovered from it. Some who couldn't recall being sick from COVID-19 reported that they were probably asymptomatic. Those who did not want the COVID-19 vaccine considered their natural immunity to be sufficient to fend off the next infection. Those who had been vaccinated previously, had fallen ill, or had severe side-effects to the booster were not sure if they would get the booster, although there were a few who wanted to follow the advice of their HCP and get boosted whenever recommended. Overall, participants through their individual pandemic experiences expressed that they had the know-how of what they would need to do if there was another outbreak. This would be a great opportunity for public health administrators and the government to add both the positive and negative experiences of the public in their safety measures rather than only from those who are in health care, pharmaceuticals, and the government.

Participants expressed frustration toward the public health administration and specifically the CDC for discounting the impact of natural immunity against COVID-19. For these

participants profit and political gain were the motivating factors for mandating the vaccine on the masses. There were also participants who trusted the advice of their HCP, and based on their positive experiences with vaccines, believed that vaccines were necessary and safe in developing immunity against a viral infectious disease. Health and government authorities should take note of both the natural and vaccine induced immunity when devising steps to prepare for the next viral outbreak. Investing in consistent messaging which primarily deflates rumors and misinformation on all media, and improved mechanisms for the surveillance, monitoring, response, and treatment beyond vaccination, would be instrumental in earning the trust and building confidence of people about the counter measures which will be put into place.

For all participants social media had no direct influence on their vaccination stance, however they were unanimous in reporting that the government and media could have done a better job in communicating truthful, unbiased information to the public. Conversely their social media experience affirmed that they had to rely on their own formed conscience in establishing opinions about the information they consumed and not let that affect their own judgment for or against vaccination. Reiterating the impact of consistent messaging irrespective of the medium through which it is propagated will be key for public health administration and the government to ensure that people always have reliable information. Knowing that there will always be people on both sides of the vaccine hesitancy continuum, tailored messaging will be ideally suited for them to take required safety precautions. Local support representatives equipped with this messaging would effectively address localized situations impacting the area.

Participants had all grown to develop mutual respect for others, specifically for those who had different opinions. Some expressed sorrow in broken relationships, others felt compelled to stand their ground but were willing to adapt and protect the wellbeing of the other, especially those with low immunity. A unified stance for or against vaccination was strongly embraced within a nuclear family unit rather than in an extended family and a friend network. There was one instance where spouses were on opposite sides but still found a common medium for the sake of their children. They were open to a dialogue and not let their differences in opinions about vaccination come in the way of their marriage. Having ready access to family data

from census and medical health insurance records, health and government authorities could create impactful targeted messaging and support services for the successful outcome of health and safety measures.

Looking back to the years of the pandemic, all participants expressed gratitude toward having the freedom of choice to either get the vaccine or not. They were also grateful that they did not have to make any difficult decisions related to their employment as others had to do. For all of them the desire to freely choose the best health option will serve in the best interests of not only their own health but also of their near and dear ones, and the community in which they work and live. Past vaccination experiences for the participants of the study had been mixed and was rooted in how they felt after the shot was administered. For one participant it was disconcerting to consider the long-term impact of a DNA altering vaccine and its potential link to future serious health issues. This would be another great opportunity for health and government authorities to channel and leverage these experiences. Citizens should be encouraged to share their experiences, apprehensions, and successes with their HCP so that the resulting analytics and intelligence could be adopted into future measures. These could also infuse care and empathy into future health and safety programs.

6. LIMITATIONS AND FUTURE CONSIDERATIONS

This study interviewed 15 participants, and all but one participant was from the same geographic location that the researchers lived in. For a subsequent or a future study, a social media extract from Twitter or responses from 15 additional participants from a wider geographic area would provide for a better sampling of experiences of the phenomenon. There was only one participant in the 18-29 age group. A better mix of participants perhaps in the younger age groups would represent people with heavier social media participation and interaction. All but three interviews were conducted over a video conference and among these only two participants did not have their camera on for the interview. Video of the interviews was not recorded, however being on camera could have inhibited participants from being fully candid about their experiences. Not all participants had the same social media understanding. Perhaps a better overview on social media, its various forms, and how people use them would have been beneficial to elicit more feedback from the

participants on the topic of whether social media had any influence on their respective vaccination stances. Adding coders to help review the analysis and help generalize the themes of this study could have helped eliminate biases, a beneficial consideration for a future study. This study has shown the effective use of technology tools such as Google Meet, otter.ai, and NVivo in the study of health care behavior which could be conducive to students in an educational program to repeat this study. Finally, a mixed methods study involving a larger sample of participants explained sequentially through a quantitative survey followed by a qualitative interview would provide a clearer and richer picture of COVID-19 vaccine hesitancy.

7. CONCLUSION

Vaccine hesitancy will continue to be a phenomenon long after COVID-19. Governments and health organizations all over the world will need to continue to develop ways to unite all people, despite their diverse beliefs and actions, in the defense against future disease outbreaks. There will always be people who will decide against vaccination measures, and therefore collaborative ways to protect the immunity of those that cannot naturally defend against the infection will have to be designed with careful thought. The themes from this study could offer needed clarity for those future programs. One key aspect that became clear from this study was that people wanted to be free to choose. Therefore, in addition to the consistency and genuineness in their messaging, it would be significant for governments and health organizations to a) devise disease prevention programs which consider people's freedom to choose as a key component, b) provide free access to these programs, and c) promulgate disease prevention education to people in their respective constituencies.

8. REFERENCES

- Aw, J., Seng, J. J. B., Seah, S. S. Y., & Low, L. (2021). COVID-19 vaccine hesitancy—a scoping review of literature in high-income countries. *Vaccines* 2021, Vol. 9, Page 900, 9(8), 900. <https://doi.org/10.3390/VACCINES9080900>
- Bell, E., Bryman, A., & Harley, B. (2022). *Business research methods* (5th ed.). Oxford University Press.
- Betsch, C., Schmid, P., Heinemeier, D., Korn, L., Holtmann, C., & Böhm, R. (2018). Beyond confidence: Development of a measure assessing the 5C psychological antecedents of vaccination. *PLOS ONE*, 13(12), e0208601. <https://doi.org/10.1371/JOURNAL.PONE.0208601>
- Bish, A., Yardley, L., Nicoll, A., & Michie, S. (2011). Factors associated with uptake of vaccination against pandemic influenza: A systematic review. *Vaccine*, 29, 6472–6484. <https://doi.org/10.1016/j.vaccine.2011.06.107>
- Centers for Disease Control and Prevention. (2023, March 15). CDC COVID Data Tracker: Vaccinations in the US. Atlanta, GA: U.S. Department of Health & Human Services. https://covid.cdc.gov/covid-data-tracker/#vaccinations_vacc-people-booster-percent-pop5
- Creswell, J. W., & Creswell, J. D. (2018). *Research design* (5th ed.). SAGE Publications.
- Creswell, J. W., & Poth, C. N. (2018). *Qualitative inquiry & research design: Choosing among five approaches* (Fourth). Sage publications.
- De', R., Pandey, N., & Pal, A. (2020). Impact of digital surge during Covid-19 pandemic: A viewpoint on research and practice. *International Journal of Information Management*, 55, 102171. <https://doi.org/10.1016/J.IJINFOMGT.2020.102171>
- Dubé, E., Laberge, C., Guay, M., Bramadat, P., Roy, R., & Bettinger, J. (2013). Vaccine hesitancy. *Human Vaccines & Immunotherapeutics*, 9(8), 1763–1773. <https://doi.org/10.4161/HV.24657>
- Fan, C. W., Chen, I. H., Ko, N. Y., Yen, C. F., Lin, C. Y., Griffiths, M. D., & Pakpour, A. H. (2021). Extended theory of planned behavior in explaining the intention to COVID-19 vaccination uptake among mainland Chinese university students: An online survey study. 17(10), 3413–3420. <https://doi.org/10.1080/21645515.2021.1933687>

- Fridman, A., Gershon, R., & Gneezy, A. (2021). COVID-19 and vaccine hesitancy: A longitudinal study. *PLOS ONE*, 16(4), e0250123. <https://doi.org/10.1371/JOURNAL.PONE.0250123>
- Hale, T., Angrist, N., Goldszmidt, R., Kira, B., Petherick, A., Phillips, T., Webster, S., Cameron-Blake, E., Hallas, L., Majumdar, S., & Tatlow, H. (2021). A global panel database of pandemic policies (Oxford COVID-19 government response tracker). *Nature Human Behaviour* 2021 5:4, 5(4), 529–538. <https://doi.org/10.1038/s41562-021-01079-8>
- Henderson, L., Millett, C., & Thorogood, N. (2008). Perceptions of childhood immunization in a minority community: Qualitative study. *Journal of the Royal Society of Medicine*, 101(5), 244–251. https://doi.org/10.1258/JRSM.2008.070363/ASSET/IMAGES/LARGE/10.1258_JRSM.2008.070363-FIG1.JPEG
- Husain, F., Shahnawaz, M. G., Khan, N. H., Parveen, H., & Savani, K. (2021). Intention to get COVID-19 vaccines: Exploring the role of attitudes, subjective norms, perceived behavioral control, belief in COVID-19 misinformation, and vaccine confidence in Northern India. *Human Vaccines and Immunotherapeutics*, 17(11), 3941–3953. <https://doi.org/https://doi.org/10.1080/21645515.2021.1967039>
- Khubchandani, J., Sharma, S., Price, J. H., Wiblehauser, M. J., Sharma, M., & Webb, F. J. (2021). COVID-19 vaccination hesitancy in the United States: A rapid national assessment. *Journal of Community Health*, 46(2), 270–277. <https://doi.org/10.1007/S10900-020-00958-X/TABLES/2>
- Kricorian, K., Civen, R., & Equils, O. (2022). COVID-19 vaccine hesitancy: Misinformation and perceptions of vaccine safety. *Human Vaccines & Immunotherapeutics*, 18(1), 1950504. <https://doi.org/10.1080/21645515.2021.1950504>
- Larson, H. J., Jarrett, C., Eckersberger, E., Smith, D. M. D., & Paterson, P. (2014a). Understanding vaccine hesitancy around vaccines and vaccination from a global perspective: A systematic review of published literature, 2007–2012. *Vaccine*, 32(19), 2150–2159. <https://doi.org/10.1016/j.vaccine.2014.01.081>
- Larson, H. J., Jarrett, C., Eckersberger, E., Smith, D. M. D., & Paterson, P. (2014b). Understanding vaccine hesitancy around vaccines and vaccination from a global perspective: A systematic review of published literature, 2007–2012. *Vaccine*, 32, 2150–2159. <https://doi.org/10.1016/j.vaccine.2014.01.081>
- Lockey, E. (2020). COVID-19: The race for a vaccine. *Journal of the Renin-Angiotensin-Aldosterone System: JRAAS*, 21(2). <https://doi.org/10.1177/1470320320926902>
- Lopalco, P. L., & Tan, L. (2016). Pandemic vaccines: Are we prepared for the next pandemic? *Future Virol*, 11(4), 253–258. <https://doi.org/10.2217/fvl-2016-0020>
- MacDonald, N. E., Eskola, J., Liang, X., Chaudhuri, M., Dube, E., Gellin, B., Goldstein, S., Larson, H., Manzo, M. L., Reingold, A., Tshering, K., Zhou, Y., Duclos, P., Guirguis, S., Hickler, B., & Schuster, M. (2015). Vaccine hesitancy: Definition, scope and determinants. *Vaccine*, 33(34), 4161–4164. <https://doi.org/10.1016/J.VACCINE.2015.04.036>
- Mayo Clinic Staff. (2022, April 20). Herd immunity and COVID-19: What you need to know - Mayo Clinic. Mayo Clinic. <https://www.mayoclinic.org/diseases-conditions/coronavirus/in-depth/herd-immunity-and-coronavirus/art-20486808>
- Meleo-Erwin, Z., Basch, C., MacLean, S. A., Scheibner, C., & Cadorett, V. (2017). “To each his own”: Discussions of vaccine decision-making in top parenting blogs. *Human Vaccines & Immunotherapeutics*, 13(8), 1895–1901. <https://doi.org/10.1080/21645515.2017.1321182>
- Mills, M., Rahal, C., Brazel, D., Yan, J., & Gieysztor, S. (2020). COVID-19 vaccine deployment: Behavior, ethics,

- misinformation and policy strategies. In *The Royal Society*.
<https://royalsociety.org/-/media/policy/projects/set-c/set-c-vaccine-deployment.pdf>
- Neely, S. R., Eldredge, C., Ersing, R., & Remington, C. (2022). Vaccine hesitancy and exposure to misinformation: A survey analysis. *Journal of General Internal Medicine*, 37(1), 179–187.
<https://doi.org/10.1007/S11606-021-07171-Z/TABLES/6>
- Reno, C., Maietti, E., Di Valerio, Z., Montalti, M., Fantini, M. P., & Gori, D. (2021). Vaccine hesitancy towards COVID-19 vaccination: Investigating the role of information sources through a mediation analysis. *Infectious Disease Reports*, 13(3), 712–723.
<https://doi.org/10.3390/idr13030066>
- Shiehzadegan, S., Alaghemand, N., Fox, M., & Venketaraman, V. (2021). Analysis of the delta variant B.1.617.2 COVID-19. *Clinics and Practice*, 11(4), 778–784.
<https://doi.org/10.3390/CLINPRACT11040093>
- Troiano, G., & Nardi, A. (2021). Vaccine hesitancy in the era of COVID-19. *Public Health*, 194, 245–251.
<https://doi.org/10.1016/J.PUHE.2021.02.025>
- WHO February 2020 Situation Report-13. (2020, February 2). Novel Coronavirus (2019-NCov).
https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200202-sitrep-13-ncov-v3.pdf?sfvrsn=195f4010_6
- Wilson, S. L., & Wiysonge, C. (2020). Social media and vaccine hesitancy. *BMJ Global Health*, 5(10), e004206.
<https://doi.org/10.1136/bmjgh-2020-004206>
- Wolf, M. S., Serper, M., Opsasnick, L., O’Conor, R. M., Curtis, L., Benavente, J. Y., Wismer, G., Batio, S., Eifler, M., Zheng, P., Russell, A., Arvanitis, M., Ladner, D., Kwasny, M., Persell, S. D., Rowe, T., Linder, J. A., & Bailey, S. C. (2020). Awareness, attitudes, and actions related to COVID-19 among adults with chronic conditions at the onset of the U.S. outbreak. *American College of Physicians*, 173(2), 100–109.
<https://doi.org/10.7326/M20-1239>
- Wong, B., & Bottorff, C. (2023, May 18). Top social media statistics and trends of 2023. *Forbes Advisor*, 1–1.
<https://www.forbes.com/advisor/business/social-media-statistics/>
- World Health Organization. (2014). Report of the SAGE working group on vaccine hesitancy.
https://www.who.int/immunization/sage/meetings/2014/october/1_Report_WORKING_GROUP_vaccine_hesitancy_final.pdf
- Xia, Y., Zhong, L., Tan, J., Zhang, Z., Lyu, J., Chen, Y., Zhao, A., Huang, L., Long, Z., Liu, N. N., Wang, H., & Li, S. (2020). How to understand “Herd Immunity” in COVID-19 pandemic. *Frontiers in Cell and Developmental Biology*, 8, 991.
<https://doi.org/10.3389/FCCELL.2020.547314/BIBTEX>
- Yahaghi, R., Ahmadizade, S., Fotuhi, R., Taherkhani, E., Ranjbaran, M., Buchali, Z., Jafari, R., Zamani, N., Shahbazkhania, A., Simiari, H., Rahmani, J., Yazdi, N., Alijani, H., Poorzolfaghar, L., Rajabi, F., Lin, C. Y., Broström, A., Griffiths, M. D., & Pakpour, A. H. (2021). Fear of COVID-19 and perceived COVID-19 infectability supplement theory of planned behavior to explain Iranians’ intention to get COVID-19 vaccinated. *Vaccines* 2021, Vol. 9, Page 684, 9(7), 684.
<https://doi.org/10.3390/VACCINES9070684>

Editor’s Note:

This paper was selected for inclusion in the journal as 2023 ISCAP Conference Distinguished Information Systems Applied Research Paper. The acceptance rate is typically 7% for this category of paper based on blind reviews from six or more peers including three or more former best papers authors who did not submit a paper in 2023.

Appendices and Annexures

Interview Protocol

The Script

At the start of the interview, the study participant would be welcomed using the following **script**:

Welcome and thank you for meeting with me today. I am a PhD student at Robert Morris University in the Information Systems and Communications program conducting a study that is exploring perceptions about COVID-19 vaccines. This interview will focus on your personal experiences with vaccines during the COVID-19 pandemic. This interview will last forty minutes and will consist of twelve questions following a few demographic related questions. I would like your permission to transcribe our conversation so I can accurately document the information you convey. If at any time during the interview you wish to discontinue the use of the transcriber or the interview itself, please feel free to let me know. All your transcribed responses will be kept confidential, archived in a safe place, and destroyed after the completion of this assignment. You will receive an email copy of the transcription after the meeting ends. Thank you. Do you have any questions or concerns before we begin? Then with your permission we will begin the interview.

The Interview

Demographics

1. What is your marital status?
2. What is your highest level of education?
3. What is your age group?
4. What is your occupation?

The Questionnaire

Note: Each question below shows the factor from the RQ it maps to in parentheses.

1. How worried are you about contracting COVID-19? (Vaccination complacency) (Mills et al., 2020)
2. Are you any less worried about contracting COVID-19, knowing that the COVID-19 vaccines are available? Explain. (Vaccination confidence) (Fan et al., 2021)
3. How do you stay in the know about COVID-19 and how do you consume this information? (Sources of information) (Reno et al., 2021)
4. Some people feel that the COVID-19 vaccines were created in a rush, are too dangerous for human consumption, and are ineffective against COVID-19. What do you think? Why? (Vaccination confidence) (Troiano & Nardi, 2021)
5. What has been your experience about vaccines and vaccinations, including COVID-19 vaccines? (Past vaccination behavior) (Mills et al., 2020)
6. What has been your experience in communities that you belong to including those on social media involving topics around the COVID-19 pandemic and COVID-19 vaccines? (Social norms)
7. How has your experience in communities on social media affected your trust of vaccines and in particular COVID-19 vaccines? (Social norms) (Fan et al., 2021)
8. What do the people close to you think about the COVID-19 pandemic and the COVID-19 vaccines? (Social norms) (Yahaghi et al., 2021)
9. What types of barriers or access to COVID-19 vaccines have you experienced and how have you overcome them? (Vaccination convenience) (Aw et al., 2021)

10. Describe your sense of freedom in choosing whether to get the COVID-19 vaccination or not? (Vaccination convenience) (Fan et al., 2021)
11. Under what situations would you consider a COVID-19 vaccine and other vaccines in the future and why? (Vaccine Hesitancy) (Yahaghi et al., 2021)
12. Would you have anything more to add relative to your pandemic experience?

Decisional Guidance to Promote Motivation in Supply Chain Decision Making

Russell Haines
hainesrp@appstate.edu

Darin Hodges
hodgesdc@appstate.edu

Computer Information Systems
Appalachian State University
Boone, NC, 28608-2037, USA

Abstract

Supply chain decision making can cause exaggerated fluctuations in inventory levels in spite of small changes in customer demand. While previous research has recognized the positive impact of information sharing on supply chain decision making, little attention has been given to how information is presented, such as through dashboards or information displays. This study applies bounded rationality theory, intrinsic motivation, and the knowledge-based view to address this gap, exploring the effects of providing information to supply chain managers, emphasizing how the information is presented rather than merely its availability. Via an experiment using a version of the beer game supply chain simulation, we find that the display of information about inventory and upstream incoming orders significantly impacted overall performance. Psychometric modeling indicates that knowledge acquisition and shared meaning are crucial in decision-makers' perceived performance. Moreover, information distribution among supply chain participants will likely contribute to cognitive overload and reduce motivation to improve decision-making.

Keywords: Information display, supply chain performance, bounded rationality, intrinsic motivation

Recommended Citation: Haines, R., Hodges, D.. (2024). Decisional Guidance to Promote Motivation in Supply Chain Decision Making. *Journal of Information Systems Applied Research*. 17(2), pp 18-30, <https://doi.org/10.62273/SUDK8361>

Decisional Guidance to Promote Motivation in Supply Chain Decision Making

Russell Haines and Darin Hodges

1. INTRODUCTION

In 2022, amid a post-COVID boom in retail sales, large retailers such as Target, Walmart, and Kohl's reported poor financial performance and cut profit forecasts (Repko, 2022). They blamed the poor performance on an inventory build-up and subsequent discounting of prior-season apparel and home items (Waldow, 2022). Analysts noted that the inventory build-up occurred because of cascading problems throughout the supply chain: (1) late deliveries of items manufactured in China because of back-ups of container ships at ports along the west coast of the United States, (2) the back-up of container ships at ports was the result of a shortage of truck drivers, and (3) the shortage of truck drivers was due to Covid-19 shutdowns (Saraiva, 2021).

In examining these supply chain disruptions at the system-wide level, supply chain experts saw a clear case of the bullwhip effect – the supply chain members overreacted to small changes in customer demand at the beginning of the Covid-19 lockdowns (Tan, 2021). This, in turn, led to large increases in orders, which ironically, meant that subsequent large deliveries would cause supply chain problems and take even longer to arrive (Moin, 2022). Similar ordering inefficiencies during Covid-19 lockdowns that had wide-reaching effects were observed in integrated circuit ordering at automobile manufacturers (Wayland, 2021) and even with toilet paper (Shih, 2021).

When criticizing the decision-making of these supply chain members, it is easy to point to the various warning signs that were available in the news. The more rational thing for them to do would have been to take the predicted effects of those warning signs into account and increase safety stocks, anticipating the supply disruption and being able to capitalize on it. However, in most circumstances, human decision-making is less than rational (Kahneman et al., 1982) especially when making decisions in supply chains (Sternman, 1989). In this paper, we explore the effects of providing information to supply chain managers on supply chain performance, which has been proposed to

mitigate supply chain inefficiencies such as the bullwhip effect (Croson & Donohue, 2003; Yang et al., 2021). We focus on *how* the information is presented rather than simply what information is available.

The greater information system and operations management literature have investigated supply chain inefficiencies and how to mitigate them quite extensively in the past and suggest increasing data sharing across the supply network, tightening supply chain slack, and decreasing overall risk in decision making (Chen et al., 2022; Hult et al., 2004; Kovach et al., 2015). Still others have suggested that information complexities in the supply chain decision matrix for organizations and individuals lead to strategic dilemmas. These dilemmas are where information ambiguity, information overload, and information asymmetry create failures within the decision-making process that bounds the decision makers' rational response mechanisms (Gunessee & Subramanian, 2020; Hodges & Salam, 2018; Patnayakuni et al., 2006; Vosooghidizaji et al., 2020).

Transaction cost economics (TCE) theory predicts that firms would wish to vertically integrate to alleviate any irrationality in the supply chain while at the same time attempting to reduce information ambiguity and asymmetries (Short et al., 2016), especially during times of environmental uncertainty. However, due to cost mitigations, merely monitoring the supply chain has become an acceptable form of supply chain governance, with data sharing being at the center, to help overcome transaction costs and reduce said information asymmetries and ambiguity.

While it is generally accepted that supply chain monitoring via upstream and downstream information sharing can lead to positive outcomes for supply chain operations and reduced bullwhip effect, little attention has been placed on how information is presented (or signaled) to supply chain participants. Few studies have focused in on information presentation strategies which can be used to design a bullwhip mitigation strategy, combating information overload via simplifying and

expanding the rational boundaries for decision makers as opposed to the general sharing of information. We attempt to fill this gap. This study focuses on how supply chain data is presented during experimental supply chain simulations to examine differences in supply chain performance outcomes due to *how* data is presented, and follow with theoretical tests to tease out how decision-makers incorporated the displayed information during the simulations. We will attempt to answer the following question: How can we diminish the perception of overwhelming environmental uncertainty for supply chain decision-makers?

2. THEORETICAL FOUNDATION

Bounded Rationality and Upper Echelons Theory

Bounded rationality focuses on the organization's or individual's lack of full information about alternate decisions that may be available even though the consequences of inaction about those decisions are recognized (Simon, 1972). Complex problems, such as supply chain sourcing or purchasing, are thought to have limited outcomes based on all possible reasonable and rational choices due to an individual or organization's ability to compute whether a choice is feasible or logical (Hodges & Salam, 2018). When considering time-sensitive decisions, Simon (1991) viewed linear models as too simplistic to anticipate and articulate optimized decisions for complex problems with large amounts of data. Upper echelons theory suggests that managers are constrained by their bounded rationality, which in turn influences how they process information regarding environmental uncertainties (Hambrick, 2007). Perceived environmental uncertainty is a subjective state experienced by individuals, such as managers, who perceive a lack of information about environmental factors that would affect their decision-making performance (Milliken, 1987), and hence would be limited by their bounded rationality.

Three main elements are considered for information processing to engage with an organization's main goal of lower costs and shortening the overall supply chain process (Hult et al., 2004), which are tied to individual and/or organizational rational boundaries. Grant (1996) stipulated that the ability to differentiate how organizations and individuals process information can lead to higher firm and organizational performance. In the age of big data, a large portion of this ability is placed on

algorithmic calculations. Once derived, these algorithms require decision makers to understand what data is relevant, what decisions needs to be integrated into the supply chain decision matrix, and then leverage that information into an operational decision. In any given environment of the supply chain or operational environment, organizations and individuals attempt to find meaning in the information within the operational environment and go through a sensemaking exercise to build a strategy of plausible actions which create knowledge and a willingness to act (Weick et al., 2005).

Sensemaking is a form of rational boundary creation based on available knowledge and understanding. When managing product flows through a supply chain, information about those flows is critical for decision-makers - with correct and timely reactions to occurrences of over- and under-ordering mitigated by making sense of shared information (Haines et al., 2017). Information about underlying consumer demand for the end product is thought to be critical to this process, with underperforming supply chain decision-makers blaming a lack of knowledge about consumer demand for their performance (Sterman, 1989). Providing information from the point-of-sale to upstream decision-makers has been shown to mitigate the bullwhip effect specifically (Croson & Donohue, 2003). We therefore hypothesize:

H1: Sharing consumer demand information with members upstream in the supply chain will increase supply chain performance.

Intrinsic motivation and information display

Intrinsic motivation refers to the internal drive and self-regulation that leads individuals to engage in activities or tasks for their inherent enjoyment, personal fulfillment, or satisfaction rather than relying on external factors or rewards. When intrinsic motivation is present, it can contribute to individual empowerment, particularly when individuals are engaged in tasks that provide a sense of accomplishment and where performing the task serves as a reward. In such cases, external reinforcement for task completion becomes unnecessary (Deci and Ryan, 1980).

Research on intrinsic motivation within the Information Systems (IS) discipline has focused on exploring the various factors influencing an individual's perceived competence to perform a task. Directly relevant to the supply chain

management context is that individuals' feelings of competence is related to their intrinsic motivation (Herath and Rao, 2009a); therefore, providing information in a way that reduces complexity and enhances decision-making performance could create a positive feedback loop where decision makers feel they are better able to make sense of the environment, perform better, and thereby feel that they can make sense of the changing decision environment.

However, it is important to recognize that the mere availability of information does not automatically enhance decision-making in supply chain management (Haines et al., 2017). Indeed, research has shown that sharing information may hurt performance when it leads to information overload and overwhelms the processing capacity of ordering managers (Tokar et al., 2012) by suppressing their ability to make sense of shared data. To improve decision-making and, therefore, intrinsic motivation, information needs to be effectively integrated into the decision-making process of supply chain managers. Therefore, it is essential to view supply chain management as a learning process that includes cognitive training (cf. Sterman, 1994; Wu & Katok, 2006). Within this framework, supply chain managers, as learners, can have varying learning outcomes based on their intrinsic or extrinsic motivation (Ghosh, 2016). Motivation also influences how individuals transfer their learning to different contexts (Ghosh, 2022).

While past research has proposed that additional information is the main way to improve decision-making, we focus instead on how display signaling encourages the use of the most relevant information, even in the presence of the overwhelming amount of information normally available to supply chain managers. We adopt a tangentially minimalist view in which the simplicity and elegance of the display are enhanced to provide the positive feedback loop, thereby helping make sense of the changing decision environment and ultimately increase intrinsically motivated abilities, competence in decision making, and relatedness to the task at hand. Prior research has also found that users rate information higher in usability when the characteristics of beauty, clarity, effectiveness, and simplicity are present and when those characteristics meets their expectations (Hill et al., 2018). We hypothesize:

H2: Providing information about inventory and the incoming supply line in an easy-to-process display will improve supply chain performance

versus traditional ways of making the information available.

Intrinsic Motivation, Bounded Rationality, and the Knowledge-based View

The organizational learning literature provides another source of support for understanding the relationship between knowledge and supply chain dynamics. Huber (1991) identified key learning elements such as knowledge acquisition and information distribution. Knowledge acquisition refers to the process of acquiring new knowledge or information. It involves gathering data, conducting research, learning from external sources, or training programs to expand an organization's or supply chain's knowledge base. Information distribution involves disseminating relevant knowledge or information to individuals or departments within the organization. It ensures that the acquired knowledge is shared and accessible to those who need it.

While these learning elements have been studied in marketing and decision sciences, their application within the supply chain context is crucial. Understanding how knowledge acquisition, information distribution, information interpretation, and organizational memory operate within the supply chain can provide insights into how knowledge is generated, shared, and leveraged to improve supply chain performance and outcomes.

Intrinsic motivation and the knowledge-based view of the firm are interrelated and mutually reinforced. When intrinsically motivated, individuals have a natural curiosity and desire to learn, explore, and engage in tasks that provide personal fulfillment, this intrinsic motivation leads them to actively seek new knowledge, acquire expertise, and develop unique abilities (Deci and Ryan, 2010). As a result, the firm should benefit from its employees' increased knowledge, capital, and capabilities within the supply chain environment. Additionally, this cooperation between intrinsic motivations and increased knowledge should expand the bounds of rational decisions and increase the alternatives available to individuals to make sense of supplied information. As individuals gain knowledge, an expectation of improved efficiency in each participant's role within the supply chain should increase chain performance (Grant, 1996a) and lead to positive feedback loops.

Knowledge is also acquired through simplifying cognitive loads through reductions in complexity

and is a way to increase learning for novices (Fathi et al., 2023). Developing knowledge about the supply chain is important for decision-makers in a supply chain in order to increase its performance (Hult et al., 2004). Thus, we assert that decision-makers who indicate higher levels of knowledge acquisition about the supply chain will perceive that they perform better than those with lower levels. We hypothesize:

H3: Knowledge acquisition activities will positively affect satisfaction with supply chain performance.

Hult et al. (2002) also noted that information distribution is linked to efficient operations of supply networks as supply chain members become more educated about supply partners and the patterns they articulate over time. Information distribution has also been shown to be a part of the knowledge development process (Kamel & Syed Awais Ahmad, 2019). Therefore, we would expect:

H4: Information distribution activities will positively affect satisfaction with supply chain performance.

Shared meaning has been analyzed within information processing literature (Daft and Weick, 1984; Gioia and Thomas, 1996; Thomas et al., 1993). In a supply chain context, shared meaning enables supply chain members to have a common understanding of goals, strategies, and anticipation of disruptions and/or opportunities. The sharing of information has been shown to positively affect the sensemaking ability of why and/or for whom decisions are made within the supply chain environment. Hult et al. (2004) concluded that shared meaning between supply partners created positive effects on reductions of supply cycle times. We therefore hypothesize:

H5: Shared meaning of information between supply network partners will positively affect satisfaction with supply chain performance.

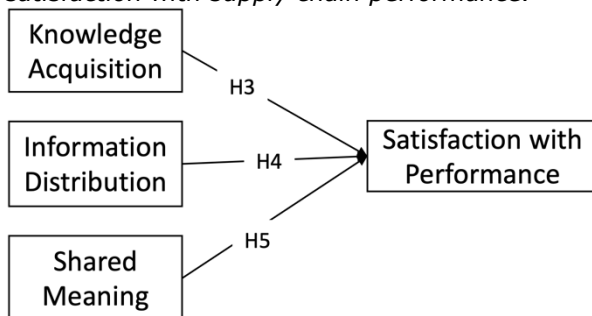


Figure 1: Research Model

The research model is shown in Figure 1.

3. RESEARCH METHODOLOGY

We used an experiment to examine how information display could influence supply chain performance. A version of the beer game (Sterman, 1989) called the "Distribution System Simulation" was used with bullets (•) representing cases of merchandise. Figure 2 shows the game screen in the control condition. Each supply chain made a total of 52 orders, lasting about 50 minutes total.

The simulation was set to have human players operate the wholesaler and distributor positions. Computerized players operated the retailer and factory positions to have a more predictable bullwhip pattern in response to incoming consumer orders.

Treatments

The demand information display was manipulated by randomly selecting supply chains to see a column on the record sheet with actual consumer demand listed (right hand side of Figure 3). Both human players in a supply chain saw the information display for consumer demand information. Showing consumer demand in this way was display manipulation because players could track this number by observing the retailer inventory level and amount shipped to consumer areas of the game screen.

The supply line information display was manipulated by randomly selecting supply chains to see "Your Inventory" and "Shipping to you" captions with their current inventory level and the number of cases that the player had in the supply line (lower left corner of Figure 4). Both human players in a supply chain saw the same supply line information - either with the inventory and supply line numbers in the order form or without (compare Figures 1 and 3). This was also a display manipulation because players without the information display could calculate both numbers via the game screen using their inventory, their upstream partners backlog, and the incoming shipments.

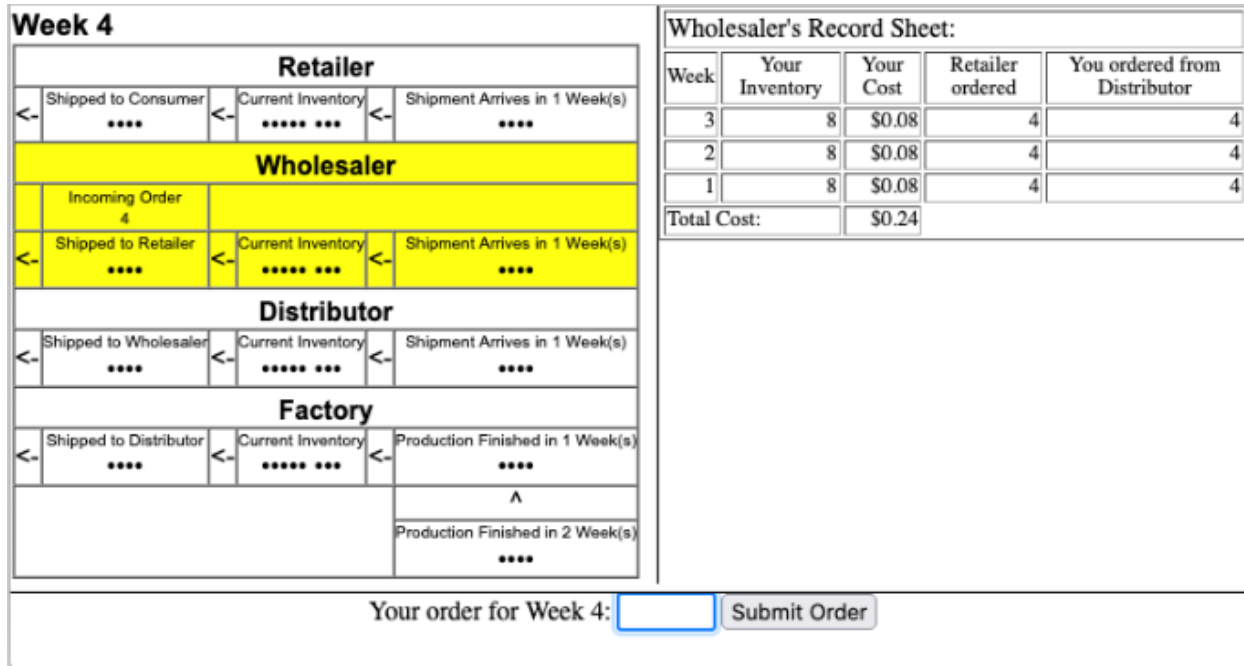


Figure 2: Game Screen in Control Condition

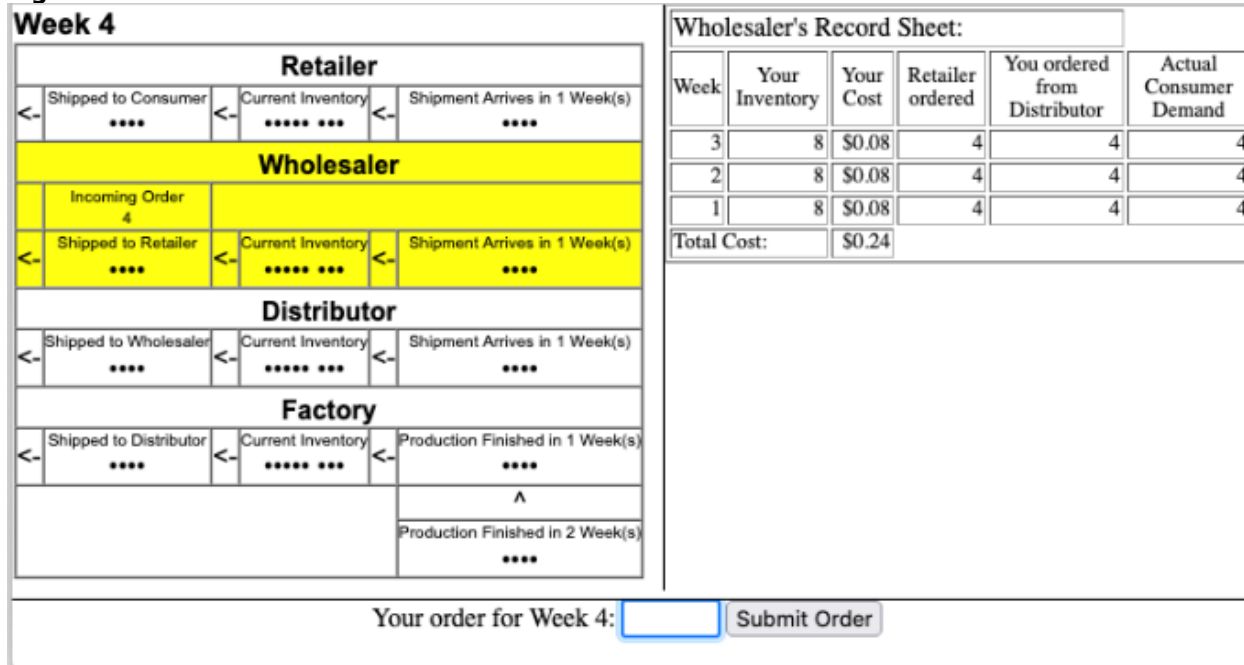


Figure 3: Demand Information Treatment

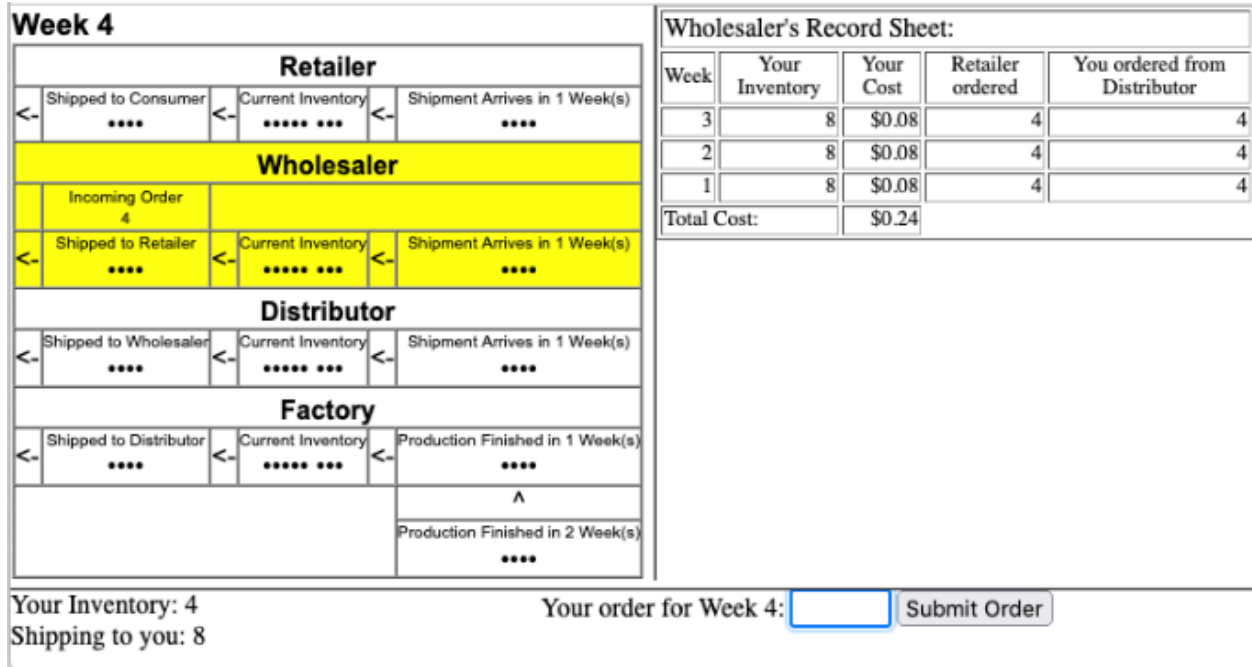


Figure 4: Supply Line Information Treatment

Overall, our contention is that information display can improve performance by enabling relevant information search and incorporating relevant information into the decision-making process. As shown in Figures 2, 3, and 4, the entire supply line was visible to all players in all positions in all conditions (cf. Croson & Donohue, 2003, 2006; Wu & Katok, 2006). Additionally, inventory backlogs were shown numerically in the "Current Inventory" cell when any supply chain member was backlogged (cf. Haines et al., 2017). Thus, information availability alone was not a factor in determining decision-making performance.

Dependent Variable

The dependent variable was the total cost for the supply chain, about which the players were explicitly instructed before the simulation began: "Your objective is to run the minimum total cost distribution system (retailer, wholesaler, distributor, and factory combined)". Costs were assessed for each position based on their current inventory level: \$0.01 per case per week of inventory on hand, and \$0.02 per case per week of unfilled orders. The supply chain with the lowest total cost won a prize of \$20 less their total cost. During the simulation, each participant's weekly and total costs were shown on their game screen (see Figures 1, 2, and 3).

Experimental Procedures

All players in an experimental session were present in the same classroom, where each player was seated in front of their computer. Supply chains and positions were randomly assigned, and participants were explicitly instructed that they could not talk or communicate with any other participant and could not look at the screen of any other participant. Each session began with the players indicating their informed consent. Next, the players completed a demographic and personality survey. Then, the players completed a 12-week trial session where they could familiarize themselves with the ordering process and on-screen displays. Computer players managed all the other positions during the trial. During the first three ordering periods of the trial, the administrator verbally guided the ordering process and instructed them about the various screen elements and how the costs were calculated. After those weeks, the players completed the remaining trial weeks. When all participants in a session had completed the trial, all inventories and supply lines were reset to the starting quantities, and the actual experiment began. At the end of the session, participants completed the psychometric questionnaire.

Psychometric Variables

The independent variables for the psychometric analysis were based on the questionnaire created and validated by Hult et al. (2004). They measured knowledge acquisition, information distribution, and shared meaning (for the exact

wording, see Appendix Table A1). The dependent variable was created specifically for this experiment and measured satisfaction with performance. We note here that satisfaction with performance had an expected significant, negative relationship with team total cost in a PLS model with all measures integrated ($b = -0.242$, $t = 1.983$, $p = 0.047$); however, the psychometric data was collected after the simulation was completed, and our intrinsic motivation approach recognizes that supply chain costs could have influenced the responses as much as costs were influenced by their decision-making perceptions.

4. RESULTS

A total of 124 students completed the simulation as a required class exercise in a junior-level operations management class required of all business majors at an Eastern U.S. university (62 total supply chains). Six supply chains were eliminated due to one or more members indicating that they misunderstood the ordering process via their ordering strategy, leaving 56 total supply chains included in the analysis. The average supply chain performance (total cost) and standard deviation for each treatment are summarized in Table 1.

Hypothesis Testing

The effect of showing information about consumer demand and/or a simplified presentation of supply line information

(Hypotheses 1 and 2) was tested at the supply chain level using the general linear model in the R statistical package. The dependent variable was supply chain total cost, while three dummy variables representing the three different experimental conditions were the independent variables (consumer demand displayed, inventory and pipeline information displayed, and both displayed). Thus, the control condition is no additional information display, with the three dummy variables testing the extent to which each experimental condition was greater than the control condition. Table 2 summarizes the results.

The results showed no significant differences in supply chain performance when consumer demand information was displayed to all supply chain members ($p = 0.4619$), providing no support for hypothesis one. There were significant differences in supply chain performance when inventory and the incoming supply line were presented in an easy-to-process display ($p = .0205$), which supports hypothesis two. The difference between supply chains that had both displays was marginally significant with respect to the no additional information (control) condition ($p = .0817$). Given this result, visual inspection of Tables 1 and 2 suggests that the additional consumer demand information may have hurt rather than helped decision-making versus an easy-to-process display of inventory and pipeline information alone.

	Pipeline Info Not Displayed	Pipeline Info Displayed	Overall
Consumer Demand Not Displayed	56.05 (29.04) n=13	33.55 (11.01) n=14	44.38 (24.10) n=27
Consumer Demand Displayed	49.19 (31.86) n=15	39.34 (19.90) n=14	44.43 (26.77) n=29
Overall	52.37 (30.22) n=28	36.44 (16.06) n=28	44.41 (25.29) n=56

Table 1: Mean (Standard Deviation) of Supply Chain Total Cost by Treatment Condition

	Estimate	Std. Error	t-value	p-value
Consumer Information Display	-6.864	9.260	-0.741	0.4619
Pipeline Information Display	-22.505	9.412	-2.391	0.0205
Consumer and Pipeline Information Display	-16.709	9.412	-1.775	0.0817

Table 2: Statistical Results of Treatments on Supply Chain Performance

	Knowledge Acquisition	Information Distribution	Shared Meaning	Satisfaction w/ Perf.
Knowledge Acquisition	0.838			
Information Distribution	0.679	0.821		
Shared Meaning	0.415	0.633	0.96	
Satisfaction w/ Performance	0.444	0.508	0.635	0.775

Table 3: Discriminant Validity Statistics. Correlation of Latent Variables, Square Root of AVE on Diagonals

	Chron. Alpha	Comp Rel.	AVE
Knowledge Acquisition	0.858	0.888	0.702
Information Distribution	0.837	0.851	0.674
Shared Meaning	0.914	0.915	0.921
Satisfaction Performance	0.663	0.777	0.601

Table 4: Reliabilities and Variance Extracted

The psychometric model and scale reliability and validity were tested at the individual level using SmartPLS (Ringle et al., 2022). The measurements all exhibited satisfactory reliability (Table 4), with Cronbach's alpha greater than the .6 cutoff for all constructs and composite reliability greater than the .7 cutoff for all constructs (Hair et al., 2011).

Discriminant validity was supported overall with correlations of latent variables all less than the square root of average variance extracted (Table 3) and item loadings on their own construct greater than cross-loadings on the other constructs for all items (Table A1 in the Appendix). Low levels of multicollinearity were shown with variance inflation factors (VIF) less than 2.569 (Table 5).

We note here that the loading for satisfaction with performance item three was 0.508 (see Table A1 in the Appendix), which, along with the relatively low alpha and AVE for that construct, indicates that rewording the item would increase the scale's reliability.

Hypoth.	Path Coef.	t-Value	p-value	VIF
KA-> SP (H3)	0.199	2.206	0.027	1.859
ID-> SP (H4)	0.037	0.315	0.753	2.569
SM-> SP (H5)	0.529	6.153	<0.001	1.671

Table 5: Path Statistics. t-value, p-value, Variance Inflation Factor

The measurement model is shown graphically in Figure 4, with path statistics summarized in Table 5. In the model, there are significant links from knowledge acquisition to satisfaction with performance (b=0.199, t=2.206, p=0.027) and shared meaning to satisfaction with performance (b=0.529, t=6.153, p<0.001), supporting hypotheses three and five. The link between information distribution and satisfaction with performance was not significant (b=0.037, t=0.315, p=0.753), providing no support for hypothesis four.



* p<.05, ** p<.01, ***p<.001

Figure 5: PLS Model Results

5. DISCUSSION AND CONCLUSIONS

This experimental task was completed by college students, which limits its generalizability versus contexts where supply chain professionals are making ordering decisions. However, the data

can be analyzed for psychological insights into the behavior of people who participate in supply chain exercises. Additionally, the results offer a theoretical grounding for future research about supply chain decision making by professionals.

The experimental results offer insights into the behavioral causes of the bullwhip effect and the broader research stream of information seeking and overload. Most importantly, the display treatment for inventory and upstream incoming orders had a significant impact on the overall performance of the supply chains. Prior research has suggested that underweighting of the upstream supply has the most effect on performance (Croson & Donohue, 2006; Yang et al., 2021); however, no treatment condition has confirmed this cause-and-effect relationship until now.

The relative simplicity of gathering upstream incoming order information has significant implications for information systems that support supply chain decision-making. Broad-based supply chain information systems have been proposed that involve all supply chain members (Smith et al., 2021). Because of power asymmetries, implementing supply chain information systems normally is determined by the most powerful member (Vanajakumari et al., 2021); however, our results suggest that individual supply chain members can develop and implement systems that aid inventory management and improve supply chain performance without involving other supply chain members. Such internal systems evoke intrinsic motivations while expanding the rational decisions available and allowing more purposeful sensemaking about shared information between partners.

The psychometric results suggest that knowledge acquisition and shared meaning impact decision-makers' perceived performance. Prior research has suggested generically that communication and coordination are able to improve supply chain decision-making (Wu & Katok, 2006; Yang et al., 2021). Our results suggest that information systems should encourage supply chain members to seek out relevant information that can be used to create a common understanding of what events are critical and how to react to them.

The insignificant link from information distribution to satisfaction with performance further supports the notion that additional information from upstream and downstream partners is as likely to add to cognitive overload

and reduce motivation as it is to improve decision-making (cf. Tokar et al., 2012). Tokar et al. (2012)'s findings, along with the insignificant impact of sharing consumer demand information, further support our contention that complicated supply chain management systems that integrate data from all levels of the supply chain are as likely to hurt as help supply chain decision-making.

6. REFERENCES

- Baum, C., Hauer, M., Joglekar, A., and Turco, A. (2023). "Thinking beyond Markdowns to Tackle Retail's Inventory Glut," May 8. (<https://view.ceros.com/leff-communications/mck-thinking-beyond-markdowns-ex1-alt>, accessed July 9, 2023).
- Chen, J., Yu, K., & Gong, J. (2022). Supply chain slack and sustainable development performance: The "fit-adjust" effect of objective and perceived environmental uncertainties. *Corporate Social Responsibility and Environmental Management*, 29(5), 1595-1604.
- Croson, R. and K. Donohue (2003). "Impact of POS data sharing on supply chain management: An experimental study." *Production and Operations Management* 12(1): 1-11.
- Croson, R., & Donohue, K. (2006). Behavioral causes of the bullwhip effect and the observed value of inventory information. *Management Science*, 52(3), 323-336.
- Daft, R. L., & Weick, K. E. (1984). Toward a model of organizations as interpretation systems. *Academy of Management Review*, 9(2), 284-295.
- Deci, E. L., & Ryan, R. M. (2010). Intrinsic motivation. *The Corsini Encyclopedia of Psychology*, 1-2.
- Fathi, R., Teresco, J. D., & Regan, K. (2022). Measuring Learners' Cognitive Load when Engaged with an Algorithm Visualization Tool. *In Proceedings of the EDSIG Conference* ISSN (Vol. 2473, p. 4901).
- Ghosh, B. (2016). Exploratory study of effects of elearning system acceptance on learning outcomes. *Journal of Information Systems Applied Research*, 9(2), 13.
- Ghosh, B. (2018). Exploratory Study of Organizational Adoption of Cloud based Big

- Data Analytics. *Journal of Information Systems Applied Research*, 11(3), 4.
- Ghosh, B. (2022). Grounded Theory Investigation into Cognitive Outcomes with Project-Based Learning. *Journal of Information Systems Applied Research*, 15(3).
- Gioia, D. A., & Thomas, J. B. (1996). Identity, image, and issue interpretation: Sensemaking during strategic change in academia. *Administrative Science Quarterly*, 370-403.
- Grant, R. M. (1996a). Prospering in dynamically-competitive environments: Organizational capability as knowledge integration. *Organization Science*, 7(4), 375-387.
- Grant, R. M. (1996b). Toward a knowledge-based theory of the firm. *Strategic Management Journal*, 17(S2), 109-122.
- Gunessee, S., & Subramanian, N. (2020). Ambiguity and its coping mechanisms in supply chains lessons from the Covid-19 pandemic and natural disasters. *International Journal of Operations & Production Management*, 40(7/8), 1201-1223.
- Haines, R., Hough, J., & Haines, D. (2017). A metacognitive perspective on decision making in supply chains: Revisiting the behavioral causes of the bullwhip effect. *International Journal of Production Economics*, 184, 7-20.
- Hair, J. F., Ringle, C. M., & Sarstedt, M. (2011). PLS-SEM: Indeed a Silver Bullet. *Journal of Marketing Theory and Practice*, 19(2), 139-152. <https://doi.org/10.2753/MTP1069-6679190202>
- Hambrick, D. C. (2007). Upper echelons theory: An update. In (Vol. 32, pp. 334-343): *Academy of Management Briarcliff Manor, NY* 10510.
- Hill, S., Wray, B., & Sibona, C. (2018). Minimalism in data visualization: Perceptions of beauty, clarity, effectiveness, and simplicity. *Journal of Information Systems Applied Research*, 11(1), 34.
- Hodges, D. C., & Salam, A. F. (2018). Machine Learning, Analytics and Strategic Decision in the Regulated Energy Industry.
- Hult, G. T. M., Ketchen Jr, D. J., & Nichols Jr, E. L. (2002). An examination of cultural competitiveness and order fulfillment cycle time within supply chains. *Academy of Management Journal*, 45(3), 577-586.
- Hult, G. T. M., Ketchen Jr, D. J., & Slater, S. F. (2004). Information processing, knowledge development, and strategic supply chain performance. *Academy of Management Journal*, 47(2), 241-253.
- Hult, G. T. M., Ketchen Jr, D. J., & Slater, S. F. (2004). Information processing, knowledge development, and strategic supply chain performance. *Academy of Management Journal*, 47(2), 241-253.
- Kamel, F., & Syed Awais Ahmad, T. (2019). Exploring the relationships of the culture of competitiveness and knowledge development to sustainable supply chain management and organizational performance [SSCM and organizational performance]. *Journal of Enterprise Information Management*, 32(6), 936-963. <https://doi.org/https://doi.org/10.1108/JEIM-06-2018-0129>
- Kovach, J. J., Hora, M., Manikas, A., & Patel, P. C. (2015). Firm performance in dynamic environments: The role of operational slack and operational scope. *Journal of Operations Management*, 37, 1-12.
- Moin, D. (2022). "Q3 Markdowns Rising, Margins Falling," *WWD*, August 30. (<https://wwd.com/business-news/retail/markdowns-third-quarter-outlook-price-promotions-1235304708/>, accessed July 9, 2023).
- Patnayakuni, R., Rai, A., & Seth, N. (2006). Relational antecedents of information flow integration for supply chain coordination. *Journal of Management Information Systems*, 23(1), 13-49.
- Repko, M. (2022). "Retailers' Biggest Holiday Wish Is to Get Rid of All That Excess Inventory," *CNBC*, November 13. (<https://www.cnn.com/2022/11/13/retailers-biggest-holiday-wish-is-to-get-rid-of-all-that-excess-inventory.html>, accessed July 9, 2023).
- Ringle, Christian M., Wende, Sven, & Becker, Jan-Michael. (2022). SmartPLS 4. Oststeinbek: SmartPLS. Retrieved from <https://www.smartpls.com>
- Saraiva, A., & Murray, B. (2021). Every step of the global supply chain is going wrong—all at once. *Bloomberg*, Nov, 22.
- Shih, W. (2021). "Toilet Paper Surplus, Ketchup And Semiconductor Chip Shortages – Supply Chain Forecasting Is Hard," April 19. (<https://www.forbes.com/sites/willyshih/2021>

- [1/04/19/toilet-paper-surplus-ketchup-and-semiconductor-chip-shortages--supply-chain-forecasting-is-hard/?sh=6ec3680271b2](https://www.researchgate.net/publication/3680271b2), accessed July 9, 2023).
- Short, J. L., Toffel, M. W., & Hugill, A. R. (2016). Monitoring global supply chains. *Strategic Management Journal*, 37(9), 1878-1897.
- Simon, H. A. (1972). Theories of bounded rationality. *Decision and Organization*, 1(1), 161-176.
- Simon, H. A. (1991). Bounded rationality and organizational learning. *Organization Science*, 2(1), 125-134.
- Smith, B., Xiang, J., & Medlin, D. (2021). Case Study of Blockchain Applications in Supply Chain Management Opportunities and Challenges. *J. Inf. Syst. Appl. Res*, 14, 22-50.
- Sterman, J. D. (1989). Modeling managerial behavior: Misperceptions of feedback in a dynamic decision making experiment. *Management Science*, 35(3), 321-339.
- Sterman, J. D. (1994). Learning in and about complex systems. *System dynamics review*, 10(2-3), 291-330.
- Tan, W. (2021). "Panic Ordering by Retailers Is Making the Supply Chain Crisis 'Even Worse,'" CNBC, October 20. (<https://www.cnbc.com/2021/10/20/retailers-manufacturers-over-ordering-amid-supply-chain-crisis.html>, accessed July 9, 2023).
- Thomas, J. B., Clark, S. M., & Gioia, D. A. (1993). Strategic sensemaking and organizational performance: Linkages among scanning, interpretation, action, and outcomes. *Academy of Management Journal*, 36(2), 239-270.
- Tokar, T., Aloysius, J. A., & Waller, M. A. (2012). Supply chain inventory replenishment: The debiasing effect of declarative knowledge. *Decision Sciences*, 43(3), 525-546.
- Vanajakumari, M., Mittal, S., Stoker, G., Clark, U., & Miller, K. (2021). Towards a Leader-Driven Supply Chain Cybersecurity Framework. *Journal of Information Systems Applied Research*, 14(2), 42.
- Vosooghidizaji, M., Taghipour, A., & Canel-Depitre, B. (2020). Supply chain coordination under information asymmetry: a review. *International Journal of Production Research*, 58(6), 1805-1834.
- Waldow, J. (2022). 'A Recipe for Disaster': Apparel Retailers, Crushed by Inventory Challenges, Ramp up Markdowns Ahead of the Holidays, *Modern Retail*, October 24. (<https://www.modernretail.co/marketing/a-recipe-for-disaster-apparel-retailers-crushed-by-inventory-challenges-ramp-up-markdowns-ahead-of-the-holidays/>, accessed July 9, 2023).
- Wayland, M. (2021). "How Covid Led to a \$60 Billion Global Chip Shortage for the Auto Industry," CNBC, February 11. (<https://www.cnbc.com/2021/02/11/how-covid-led-to-a-60-billion-global-chip-shortage-for-automakers.html>, accessed July 9, 2023).
- Weick, K. E., Sutcliffe, K. M., & Obstfeld, D. (2005). Organizing and the process of sensemaking. *Organization Science*, 16(4), 409-421.
- Wu, D. Y., & Katok, E. (2006). Learning, communication, and the bullwhip effect. *Journal of Operations Management*, 24(6), 839-850. <https://doi.org/10.1016/j.jom.2005.08.006>
- Yang, Y., Lin, J., Liu, G., and Zhou, L. 2021. "The Behavioral Causes of Bullwhip Effect in Supply Chains: A Systematic Literature Review," *International Journal of Production Economics* (236), p. 108120.

APPENDIX A:

	ID	KA	SM	SP
ID1: I was aware of trends as they moved through the supply chain	0.869	0.633	0.490	0.441
ID2: I could see what future needs in the supply chain would be.	0.872	0.568	0.606	0.469
ID3: I was aware of how satisfied the [other human player's position] was with me.	0.794	0.476	0.612	0.398
ID4: I could see when something important happened in the supply chain.	0.741	0.556	0.348	0.346
KA1: I tried to predict future needs in the supply chain.	0.623	0.878	0.348	0.401
KA2: I attempted to determine the effect that my orders had on the supply chain.	0.576	0.901	0.339	0.388
KA3: I made an effort to predict the effect of others' orders on the supply chain.	0.578	0.871	0.422	0.424
KA4: I tried to uncover faulty assumptions that I had about the supply chain.	0.502	0.685	0.262	0.241
SM1: The [other human player's position]and I developed a shared understanding of the available supply chain information.	0.610	0.387	0.958	0.600
SM2: The [other human player's position]and I developed a shared understanding of the implications of a supply chain activity.	0.605	0.409	0.961	0.619
SP1: I am satisfied with the performance of the supply chain.	0.497	0.402	0.629	0.892
SP2: Based on my knowledge of the supply chain, I think it was efficient and effective.	0.399	0.420	0.506	0.866
SP3: I feel that the performance of this supply chain could not be much better than it was.	0.235	0.138	0.263	0.508

Table A1: Loadings and Cross Loadings

Examining Factors Influencing the Acceptance of Big Data Analytics in Healthcare

Abdul Sajid Mohammed
amohammed5836@ucumberlands.edu
University of the Cumberland
School of Computer and Information Sciences
Williamsburg, KY

Mary Lind
mary.lind@lsus.edu
Louisiana State University in Shreveport
College of Business
Shreveport, LA USA

Abstract

This study investigated the factors influencing the acceptance of big data analytics in healthcare. Big data analytics can improve many aspects of healthcare, including diagnostics, service provision, and patient outcomes. A cross-sectional online survey administered to N = 132 professionals working in the U.S. healthcare industry used regression analysis to determine the extent of the predictive relationships between the variables. The findings support previous research linking big data analytics to performance improvements in healthcare by highlighting performance expectancy's significance as a predictor of behavioral intentions. The mixed results suggest that the Unified Theory of Acceptance and Use of Technology (UTAUT) has limited explanatory power when studying big data analytics adoption in healthcare settings. Future research should focus on developing a theory that explains big data analytics acceptance and use based on information security risks, implementation costs, and user aversion to technology.

Keywords: Big Data Analytics, health care, UTAUT, performance expectancy

Recommended Citation: Mohammed, A.S., Lind, M., (2024). Examining Factors Influencing the Acceptance of Big Data Analytics in Healthcare. *Journal of Information Systems Applied Research*, 17(2), pp.31-44. <https://doi.org/10.62273/QNDU3179>

Examining Factors Influencing the Acceptance of Big Data Analytics in Healthcare

Abdul Sajid Mohammed and Mary Lind

1. INTRODUCTION

Healthcare is making increasing use of big data analytics (Nazir et al., 2020). Big data analytics and innovative tools like electronic health records and centralized client-server architecture advance the delivery of healthcare services and improve patient outcomes (Galetsi et al., 2020). Applying big data analytics technologies in healthcare depends on healthcare practitioners' acceptance of these technologies (Aljarboa & Miah, 2020). Healthcare practitioners play a crucial role in adopting, implementing, and institutionalizing new technologies (Brock & Khan, 2019). Increasing healthcare practitioners' behavioral intentions to accept big data analytics increases the benefits of implementations.

This study investigated how factors associated with the Unified Theory of Acceptance and Use of Technology (UTAUT) influence big data analytics acceptance in healthcare settings i.e., performance expectancy, effort expectancy, social influence, and facilitating conditions.

2. THEORETICAL FRAMEWORK

This research relied on the UTAUT as the theoretical framework. Venkatesh et al. (2003) developed the UTAUT to show the factors influencing an individual's behavioral intention to accept and use new technology. The UTAUT model identifies four main factors influencing the acceptance of technology: performance expectancy, effort expectancy, social influence, and facilitating conditions (Venkatesh et al., 2003). Performance expectancy refers to an individual's level of trust in the ability of a system to help improve their performance. Effort expectancy refers to the potential user's evaluation of the ease of using and interacting with the new technology (Susanto et al., 2020). People's perceptions regarding whether or not the individual should accept and use a certain technology constitute social influence. Facilitating conditions refer to the objective environmental factors influencing an individual to accept and use new technology (Ayaz & Yanartas, 2020). These four factors were adopted as the study's independent variables, while the behavioral intention to accept big data analytics was the dependent variable.

In the UTAUT mode, the influence of the four independent variables (R1, R2, R3, and R4) on the dependent variables is moderated by age, gender, experience, and voluntariness of use. Additionally, the model includes use behavior as a dependent variable influenced by behavioral intention and facilitating conditions. The present study's theoretical model deviates slightly from the original UTAUT by eliminating the moderating variables from the model and focusing exclusively on behavioral intention. Figure 1 depicts the study's theoretical model adapted from the UTAUT.

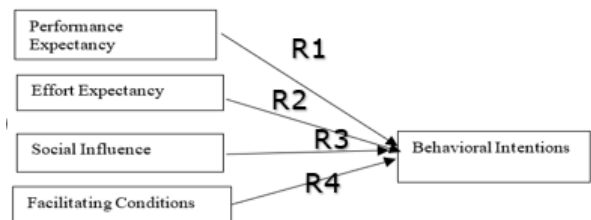


Figure 1 Research Model

The current state of the healthcare industry is shaped by the need to manage costs, increase quality, increase productivity, and function effectively in the face of complexity. The stakeholders in the industry, including payers, providers, managed healthcare organizations, pharmaceuticals, and patients, have needs that become more complicated over time. Healthcare organizations are expected to find ways of addressing current challenges, such as increased charges, inefficiency in delivery systems, increased rates of medication and medical error, and reliance on inaccurate information for decision-making (Zhu & Chen, 2021). One essential and promising solution to such challenges involves adopting information technology (IT), an enabler of healthcare transformation (Zhu & Chen, 2021). Integrating IT into healthcare operations is crucial in collecting, analyzing, and interpreting information to improve decision-making (Zhu & Chen, 2021). Healthcare organizations also risk losing competitiveness when they fail to proactively recognize the need for IT and carefully evaluate their potential contribution to healthcare operations.

IT integration in healthcare focuses on adopting a wide range of new technologies. One is telemedicine, which involves delivering healthcare services online. Unlike in the past, when patients would be required to visit hospitals upon the appearance of disease symptoms for diagnosis and treatment, telemedicine allows the exchange of clinical information between physicians and patients regardless of location using modern technologies (Han & Lee, 2021). Telemedicine combines various technologies to facilitate the non-face-to-face exchange of medical information. Another critical technology is mHealth, which involves the application of wearable devices and related applications in healthcare. Wearable devices are deployed to monitor a patient's activity levels, sleep patterns, and heart rate to help inform the prevention and treatment decisions made by the physician (Han & Lee, 2021). Healthcare organizations adopt these data-driven technologies to reduce costs, improve quality, foster productivity, and increase patient safety. These new technologies increase the digitization of health care.

Big data has emerged as one of the most famous developments in the public and private sectors. It is characterized by the data's high volume, velocity, variety, value, and veracity (Chasupa & Paireekreng, 2021). Big data refers to data that cannot be stored, processed, and computed using conventional data analysis techniques but requires advanced tools and methods. On the other hand, big data analytics involves collecting, organizing, and analyzing massive amounts of data to aid the discovery of patterns and other valuable insights. It encompasses the techniques and technologies to allow the disclosure of hidden insights from large datasets (Chasupa & Paireekreng, 2021). Big data analytics offers unique opportunities that can be exploited by society. Big data possesses the potential to change how the world, people, and organizations do things due to its role in increasing awareness and providing more profound insight.

Galetsi et al. (2020) notes that the healthcare industry is data intensive utilizing dynamic interactive platforms with innovative tools and technologies like electronic health records with centralized client-server architecture to improve patient outcomes and overall healthcare operations. Further Galetsi et al. (2020) observes that the industry captures and manages large volumes of data from various sources, such as laboratory information, library systems, and electronic health records. Big data

analytics in healthcare is characterized by the deployment of methods that enable the analysis of large amounts of electronic data relating to the delivery of care to patients (Zhan, 2019). Such data cannot be captured and analyzed using traditional techniques. Big data analytics in healthcare and medicine allows large and complex heterogeneous data to be integrated and analyzed, including telemedicine, biomedical, and electronic (Batko & Ślęzak, 2022). The application of big data analytics helps improve patient health by supporting long-term prediction about the health status of patients and informs the implementation of the proper therapeutic procedures.

Batko and Ślęzak (2022) noted that adopting big data analytics improves healthcare organizations' quality of care. The role played by big data analytics should cause the management and employees in these organizations to readily accept the deployment of big data analytics techniques to support their actions and decisions. However, accepting big data analytics techniques and tools in healthcare organizations still faces challenges, just like accepting other technologies. It is not uncommon for employees to resist and even oppose big data analytics deployment for personal and organizational reasons. For instance, employees will accept or reject a new technology depending on whether they possess the knowledge and skills required to utilize it (Lagzian & Pourbagheri, 2022). Individual, organizational, and social factors can influence the acceptance of new technology. Individual factors include individual innovation, knowledge, training, and previous experience. Organizational factors that influence technology acceptance include information security, supporting mechanisms, and the quality of the systems. Social factors include trust and available infrastructure.

3. CONTRIBUTION OF THIS RESEARCH

Most of the previous UTAUT studies reviewed either focused on the acceptance of technology at the organizational level or mixed the variables for organizations and individuals. Whereas, the current study explores individuals' evaluation that influences the acceptance of big data analytics in healthcare organizations. Employees play a crucial role in the acceptance of new technologies. No technology can ever be successfully implemented and diffused without the support and willingness of employees. Efforts made by the organization to introduce new technology, including introducing big data analytics in a healthcare organization, will not

produce the required results without the acceptance of employees (Panari et al., 2021). This study explores the individual-level factors influencing an employee's behavioral intention to accept new technology.

Much of the existing literature focused on specific job roles rather than taking a general approach to adopting big data analytics in healthcare. For instance, Cabrera-Sánchez and Villarejo-Ramos (2019) only collected data from the selected companies' CEOs and managers of departments. Managers' views may not represent those of all other employees in these organizations. Alternatively, Brock and Khan (2019) relied on the data collected from students enrolled in an IT program, who may not represent the view of all students in the university. Lastly, Ajimoko (2019) collected data from IT professionals, whose views may not represent those held by other professionals in the organization because IT professionals may focus more on the technical attributes of the technology. The present study sought to address these limitations by collecting a general sample of healthcare professionals to improve generalizability.

Most of the literature reviewed focused on accepting technology as a broader topic. For instance, Barkoczi and Lobontiu (2020) only investigated the factors influencing the acceptance of technologies in the telecommunications sector, specifically mobile computing. The study did not focus on any single technology, which was like research by Skoumpopoulou and Wong (2019), who only sought to understand the factors influencing the acceptance of new technologies in the workplace. It is important to note that employees will respond differently to different technologies. The behavioral intention to accept varies across technologies. Second, most of the studies reviewed focused on the workplace in general or other industries other than the healthcare industry. For instance, Cabrera-Sánchez and Villarejo-Ramos (2019) only explored the acceptance of big data analytics in companies, while Farias and Resende (2020) focused on technology acceptance in institutions of higher learning. After investigating the moderating role of resistance to change in adopting big data analytics in healthcare, Shahbaz et al. (2019) recommended further research on big data analytics adoption in healthcare. The healthcare industry differs from other industries regarding the technologies and workforce.

4. RESEARCH METHODS AND FINDINGS

Constructs and Survey Items from Venkatesh et al. (2003)

Performance Expectancy

If I use big data analytics...

1. I will increase my effectiveness on the job.
2. I will spend less time on routine job tasks.
3. I will increase the quality of output of my job.
4. I will increase the quantity of output for the same amount of effort.
5. My coworkers will perceive me as competent.
6. I will increase my chances of obtaining a promotion.
7. I will increase my chances of getting a raise.

Effort Expectancy

1. Learning to operate big data analytics would be easy for me.
2. I would find it easy to get big data analytics to do what I want it to do.
3. My interaction with big data analytics would be clear and understandable.
4. I would find big data analytics to be flexible to interact with.
5. It would be easy for me to become skillful at using big data analytics.
6. I would find big data analytics easy to use.

Social Influence

1. I use big data analytics because of the proportion of coworkers who use it.
2. The senior management of this business has been helpful in the use of big data analytics.
3. My supervisor is very supportive of the use of the system for my job.
4. In general, the organization has supported the use of big data analytics.

Behavioral Intentions

1. I have control over using big data analytics.
2. I have the resources necessary to use big data analytics.
3. I have the knowledge necessary to use big data analytics.
4. Given the resources, opportunities, and knowledge it takes to use big data analytics, it would be easy for me to use big data analytics.
5. Big data analytics is not compatible with other systems I use.

Facilitating Conditions

1. Guidance was available to me in the selection of big data analytics.
2. Specialized instruction concerning big data analytics was available to me.
3. A specific person (or group) is available for assistance with big data analytics difficulties.

Table 1 – Survey Items

This study addressed a gap in the literature regarding big data analytics adoption in the U.S. healthcare industry. The study examined four research questions: (1) To what extent does performance expectancy explain U.S. healthcare professionals' behavioral intentions to accept big data analytics in research-based healthcare organizations? (2) To what extent does effort expectancy explain U.S. healthcare professionals' behavioral intentions to accept big data analytics in research-based healthcare organizations? (3) To what extent does social influence explain U.S. healthcare professionals' behavioral intentions to accept big data analytics in research-based healthcare organizations? (4) To what extent do facilitating conditions explain U.S. healthcare professionals' behavioral intentions to accept big data analytics in research-based healthcare organizations? The UTAUT survey items are shown in Table 1.

Survey data were collected from healthcare professionals working in research-based healthcare organizations. This study focused on participants working at research-based healthcare organizations in the United States. Research-based healthcare organizations were selected as the research setting because these institutions generate and use vast amounts of data (Singh et al., 2018). A third-party survey provider called Pollfish facilitated participant selection and data collection. Power analysis using Gpower3 for regression indicated that a sample size of 120 was needed. Using Pollfish enabled affordable, timely, and anonymous data collection via random sampling.

The demographics for the respondents and sample size are shown in Table 2. The five UTAUT constructs had Cronbach Alpha scores of .70 or more demonstrating acceptable reliability. The constructs were not multicollinear.

	Percentage
Age	
18 to 24	85%
Over 54	15%
Gender	
Females	53%
Males	47%
Years of Work Experience	
<= 10 years	74%
> 10 years	26%

N = 132

Gpower3 recommended 120 sample size
22 million Healthcare Professionals in US

Table 2: Demographics

Analysis of the UTAUT constructs revealed that they were normally distributed. Also, they demonstrated a high degree of internal reliability, as shown in Table 3 of the descriptive statistics.

Construct/Item	M (SD)	<i>a</i>
Performance expectancy	5.61 (.989)	.73
Effort expectancy	5.45 (1.041)	.75
Social influence	5.38 (1.030)	.78
Facilitating conditions	5.56 (0.951)	.74
Behavioral intentions	5.74 (1.022)	.72

Table 3 Descriptive Statistics, Reliability

Research Question 1 addressed the relationship between performance expectancy and participants' behavioral intentions to adopt big data in U.S. research-focused healthcare organizations. In this study, performance expectancy referred to the participants' belief that big data analytics would facilitate enhanced performance for healthcare professionals working in research-focused organizations (Susanto et al., 2020). Behavioral intention refers to the participants' inclination to use big data analytics in the future (Handoko, 2019). The linear regression model analysis indicated that the relationship between performance expectancy and behavioral intentions was significant ($\beta = 0.368$, $t(127) = 3.968$, $p = .000175$). The significant finding was indicated by the p -value being lower than .05, the threshold for significance used in this study. The β value reflected a positive relationship between the variables. In support of Research Question 1, the positive β value meant that behavioral intentions increased when performance expectancy increased. This implies that doctors, nurses, and healthcare researchers who believed that using big data analytics would improve their work performance were more likely to use the technology than healthcare professionals who did not believe big data analytics would improve performance. The positive significant relationship between performance expectancy and behavioral intention was not unexpected. Working with big data is complicated because these datasets cannot easily be gathered, stored, managed, or analyzed using traditional database software tools (Sun et al., 2019). Cuzzocrea (2020) noted that big data analytics allows users to discover patterns, trends, and previously unidentified correlations. These discoveries often improve healthcare outcomes, supporting the link between performance expectancy and behavioral intention.

Research Question 2 addressed the relationship between effort expectancy and participants' behavioral intentions to adopt big data in U.S. research-focused healthcare organizations. Effort expectancy referred to the participants' belief that big data analytics would be easy to use (Yohanes et al., 2020). In the original UTAUT model, the expectation that a technology is easy to use was assumed to have a significant positive influence on users' behavioral intentions to accept and use that technology (Venkatesh et al., 2003). The linear regression analysis indicated that the relationship between effort expectancy and behavioral intentions was nonsignificant ($\beta = 0.019$, $t(127) = 0.204$, $p = .838$). The finding for Research Question 2 did not support the assumed relationship between effort expectancy and behavioral intentions in the UTAUT model. It is possible that ease of use is not closely associated with big data analytics because of the complexity of the technology (Sun et al., 2019).

Research Question 3 addressed the relationship between social influence and participants' behavioral intentions to adopt big data in U.S. research-focused healthcare organizations. Social influence refers to subjective norms and social factors in users' working environments that influence their attitudes toward using technology (Vallerie et al., 2021). Venkatesh et al.'s (2003) original UTAUT model assumed a positive link between social influence and the behavioral intention to adopt a technology. Linear regression model analysis indicated that the relationship between social influence and behavioral intentions was nonsignificant ($\beta = 0.008$, $t(127) = 0.089$, $p = .930$). The actual relationship between social influence and behavioral intention did not support Venkatesh et al.'s (2003) UTAUT model or the study's theoretical framework. Social influence has significantly influenced big data analytics use in some settings (Cabrera-Sánchez & Villarejo-Ramos, 2019). However, Queiroz and Pereira (2019) found that social influence was not always a driver of technology adoption among professionals. The present study's data analysis illustrated that social influence was not a significant driver of U.S. healthcare professionals' behavioral intentions to use big data analytics in research-focused organizations.

Research Question 4 addressed the relationship between facilitating conditions and participants' behavioral intentions to adopt big data in U.S. research-focused healthcare organizations. Facilitating conditions referred to objective factors within the work environment that supported participants' use of big data analytics (Handoko, 2019). Examples include specialized

software, organizational training programs, or onsite technical support. Linear regression analysis indicated that the relationship between facilitating conditions and behavioral intentions was significant ($\beta = 0.398$, $t(127) = 4.193$, $p = .000088$).

Research Questions	Regression Results
RQ1: To what extent does performance expectancy explain U.S. healthcare professionals' behavioral intentions to accept big data analytics in research-based healthcare organizations?	$\beta = 0.368$, $t(127) = 3.968$, $p = .000175***$
RQ2: To what extent does effort expectancy explain U.S. healthcare professionals' behavioral intentions to accept big data analytics in research-based healthcare organizations?	$\beta = 0.019$, $t(127) = 0.204$, $p = .838$
RQ3: To what extent does social influence explain U.S. healthcare professionals' behavioral intentions to accept big data analytics in research-based healthcare organizations? ($\beta = 0.008$, $t(127) = 0.089$, $p = .930$
RQ4: To what extent do facilitating conditions explain U.S. healthcare professionals' behavioral intentions to accept big data analytics in research-based healthcare organizations?	$\beta = 0.398$, $t(127) = 4.193$, $p = .000088***$

Table 4 Regression Results

The positive β value meant that behavioral intentions increased when participants had greater access to technology support structures. Doctors, nurses, and healthcare researchers who felt they had the support tools and infrastructure necessary to use big data analytics were more likely to use the technology than healthcare professionals who lacked the necessary infrastructure. The positive significant relationship between facilitating conditions and behavioral intentions was expected based on Venkatesh et al.'s (2003) original UTAUT model. Researchers like Lutfi et al. (2022) have noted that complexity can significantly affect technology adoption and use, and Sun et al. (2019) highlighted the complex nature of big data analytics. Using big data analytics in healthcare specifically involves analyzing and integrating large and complex datasets (Batko & Ślęzak, 2022). The present study's findings

regarding the significant relationship between facilitating conditions and behavioral intentions to use big data analytics support both the UTAUT model and existing research on the topic.

These findings are summarized in Table 4. Here are the research questions and regression results for each hypothesis.

5. SUMMARY, DISCUSSION, AND IMPLICATIONS

In the healthcare industry, big data analytics advance the delivery of healthcare services and improve patient outcomes (Zhu & Chen, 2021). Recent studies have illustrated the many benefits of adopting and implementing big data analytics in healthcare settings (Nazir et al., 2020). However, the effective use of new technologies depends heavily on user acceptance (Aljarboa & Miah, 2020), and realizing the benefits of big data analytics relies on the successful implementation of the analytics techniques and tools as well as support from all organizational members (Alghamdi & Alsubait, 2021). Big data analytics use in healthcare organizations is evolving (Shahbaz et al., 2019). Characterized by high volume, velocity, variety, value, and veracity, big data requires advanced analytic tools and methods (Chasupa & Paireekreng, 2021). Resistance to change is a leading cause of failed big data analytics implementations in healthcare settings (Zhang et al., 2021). Healthcare employees can be unwilling to accept new technologies, even after organizations have begun implementing them (Shahbaz et al., 2019). Healthcare organizations can use adoption models to support big data analytics implementations by identifying the most critical factors affecting acceptance and addressing them before introducing additional tools and techniques (Shahbaz et al., 2019).

Employees play a crucial role in adopting new technologies, and it is only possible to institutionalize big data analytics with the support of technology users (Boldsova, 2019). Establishing users' behavioral intentions to accept big data analytics before implementing these tools improves project success (Shahbaz et al., 2019). Attempting to implement big data analytics without gaining the support and willingness of employees leads to implementation failure (Brock & Khan, 2019). This study relied on Venkatesh et al.'s (2003) unified theory of acceptance and use of technology (UTAUT) model as a theoretical framework. The UTAUT constructs include performance expectancy, effort expectancy,

social influence, facilitating conditions, and behavioral intentions. The study aimed to determine how performance expectancy, effort expectancy, social influence, and facilitating conditions influence healthcare professionals' behavioral intentions to accept big data analytics techniques and tools.

This study was significant because its findings could be used to improve the acceptance of big data analytics in healthcare. A literature gap exists concerning the acceptance of technology in healthcare, including big data analytics (Alghamdi & Alsubait, 2021; Batko & Ślęzak, 2022). Technology managers and other healthcare organizations can refer to this study's recommendations when formulating policies and best practices that foster big data analytics acceptance among organizational staff (Lambay & Mohideen, 2020).

Healthcare organizations can benefit from several types of big data analytics, including descriptive, predictive, prescriptive, and diagnostic analytics (Fang et al., 2021; Kaur et al., 2021). Each analytics type can potentially support healthcare professionals' and healthcare organizations' performance. For example, Fang et al. (2021) used prescriptive analytics to illustrate how big data analytics could improve clinical decision-making. Hoque and Rahman (2020) developed a predictive analytics tool to support chronic disease prediction through machine learning techniques. As a result, they could predict patients' likelihood of developing health complications such as hypertension and heart disease more accurately than without prescriptive analytics.

The findings for Research Question 1 aligned with both theoretical and scholarly research. The UTAUT model assumes a positive relationship exists between performance expectancy and technology acceptance (Venkatesh et al., 2003). Research shows that big data analytics improves physicians' performance in predicting, preventing, and treating diseases (Fang et al., 2021). The present study's findings regarding performance expectancy align with and support these findings, suggesting that more research-based healthcare organizations should consider implementing big data analytics, especially if their goal is to improve performance.

Research Question 2 addressed the relationship between effort expectancy and participants' behavioral intentions to adopt big data in U.S. research-focused healthcare organizations. Effort expectancy refers to the ease or difficulty

users associate with a technology (Yohanes et al., 2020). Specifically, effort expectancy refers to the potential user's evaluation of the ease of using and interacting with the new technology (Susanto et al., 2020). The UTAUT assumes effort expectancy will influence technology adoption (Venkatesh et al., 2003). Thus, it was expected that participants who perceived big data analytics as easy to use would be significantly more likely to use the technology than participants who perceived it as harder to use.

The nonsignificant relationship between effort expectancy and the behavioral intention to adopt big data analytics did not support the study's theoretical framework. Evaluating a system's effort expectancy depends on the ease of use, design of the interface, ease of learning, and flexibility (Ayaz & Yanartas, 2020). Some studies have shown that the influence of effort expectancy on behavioral intention to use a technology decline in continuous and long-term use (Ayaz & Yanartas, 2020). The present study focused on healthcare professionals working in research-focused organizations. Such organizations would be more likely to have used big data analytics compared to rural clinics or other organizations primarily focused on patient care. Thus, their employees would have more familiarity with the technology, and it would seem easier to use.

Another similar explanation for the lack of a significant relationship between effort expectancy and the behavioral intention to adopt big data analytics in research-based healthcare organizations could be the combination of familiarity and voluntariness of use. As previously noted, researchers have found that effort expectancy influences users' behavioral intentions to use technology less over time (Ayaz & Yanartas, 2020). Effort expectancy has been known to significantly influence behavioral intentions in both mandatory and voluntary usage contexts, but this influence may only be significant the first time the technology is used (Ayaz & Yanartas, 2020). The influence of effort expectancy becomes insignificant when the system is used for a long time (Ayaz & Yanartas, 2020). Thus, it is possible that participants working in research-based healthcare organizations already had sufficient experience working with big data analytics (i.e., the ease of use of the technology) was not a concern.

Complex technologies can be harder to use, and in this way, complexity can significantly influence technology adoption. Lutfi et al.(2022)

reported that more complex technologies have lower adoption and acceptance than easy-to-use technologies. Batko and Ślęzak (2022) observed that big data analytics use in healthcare involves analyzing and integrating large and complex datasets. While the complexity of big data analytics use might seem like it would negatively affect behavioral intentions in the form of effort expectancy, the present study's findings did not support this conclusion. Rather, the findings suggest that other factors influence big data analytics adoption in this context. Researchers should look for other frameworks that might more accurately explain the drivers of big data analytics use in the U.S. healthcare industry.

Research Question 3 addressed the relationship between social influence and participants' behavioral intentions to adopt big data in U.S. research-focused healthcare organizations. Social influence refers to the effects that subjective norms and interpersonal factors have in working environments that influence users' attitudes toward using technology (Vallerie et al., 2021). In the present study, social influence described the influence of peers' and colleagues' attitudes and opinions that potentially affected U.S. healthcare professionals' intentions regarding the use of big data analytics in research-focused organizations.

The nonsignificant relationship between social influence and big data analytics did not support the study's theoretical framework. The UTAUT model assumes social influence will have a significant positive effect on technology adoption (Venkatesh et al., 2003). However, in the present study's context, this assumption was not validated. Participants were not significantly influenced by the opinions of their peers and colleagues. Despite contradicting the study's theoretical framework, the lack of a significant predictive relationship between social influence and behavioral intention to use big data analytics is not completely unexpected. Subjective norms and interpersonal factors can cause individuals to adopt technologies or adjust their technology use (Vallerie et al., 2021). However, the peer pressure associated with social influence is not typically perceived as a main driver of complex technologies like big data analytics (Queiroz & Pereira, 2019).

Cabrera-Sánchez and Villarejo-Ramos (2019) used the UTAUT model to study big data analytics adoption among CEOs and managers and found that social influence had a positive effect on the intention to use the technology. However, they also reported that social influence

was less likely to influence behavioral intention than factors like performance expectancy and facilitating conditions. In similar research, Queiroz and Pereira (2019) used the UTAUT model to examine big data analytics among Brazilian supply chain managers. Queiroz and Pereira found that social influence did not affect behavioral intentions in their population. The competitive nature of healthcare research may make social influence less of a concern when deciding whether to implement big data analytics in support of a research program.

Factors more likely to influence big data analytics adoption could include relative advantage, job fit, and organizational readiness (Lutfi et al., 2022). Relative advantage would be more relevant in environments where organizations are competing for market share or where competition is high for resources. Job fit may be more relevant in this research since healthcare researchers often deal with complex datasets that require specialized analysis tools. Many scholars have noted that big data analytics offers many benefits in the healthcare industry (Nazir et al., 2020; Shahbaz et al., 2019). These benefits can only be realized if the technologies are adopted, which relies on healthcare practitioners' acceptance of big data analytics (Aljarboa & Miah, 2020).

The lack of a significant relationship for Research Question 3 suggests that the original UTAUT model is not the most effective theoretical model when studying big data analytics in research-based healthcare settings. U.S. healthcare organizations seeking to understand their employees' intentions regarding big data analytics use should consider job fit, competitive advantage, and providing the necessary support infrastructure for the technologies to be effective. Support infrastructure is linked to facilitating conditions, the final variable examined as part of the UTAUT model. The following section addresses the results of the linear regression analysis regarding facilitating conditions.

Research Question 4 addressed the relationship between facilitating conditions and participants' behavioral intentions to adopt big data in U.S. research-focused healthcare organizations. Facilitating conditions referred to objective factors within the work environment that supported participants' use of big data analytics (Handoko, 2019). Examples include specialized software, organizational training programs, or onsite technical support. Vanduhe, Nat, & Hasan (2020) noted that training was essential to

improve technology acceptance and use because it creates a sense of self-competence among users. Facilitating conditions also reflect organizational readiness (Lutfi et al., 2022). Organizations without appropriate support have less successful technology implementations because the lack of support undermines performance and increases difficulty using the technology (Lutfi et al., 2022).

The significant relationship between facilitating conditions and behavioral intention to use big data analytics supported the UTAUT model as the study's theoretical framework. Facilitating conditions refer to objective factors like infrastructure, technical support, or institutional knowledge that make using complicated technology easier (Handoko, 2019). As noted previously, complexity can significantly influence technology adoption (Lutfi et al., 2022). Jadhav (2021) reported that the more complex a technology is to understand and use, the lower the adoption rate. Additionally, big data analytics in healthcare involves analyzing and integrating large and complex datasets (Batko & Ślęzak, 2022). These observations support the importance of organizational infrastructures that aid users in adopting technologies. This perspective highlights the inherent links between effort expectancy and facilitating conditions in the context of this study, as facilitating conditions can reduce the difficulties healthcare professionals face when using technologies like big data analytics.

The literature also suggests a link between facilitating conditions and performance expectancy. As previously noted, the large datasets associated with big data analytics are not easily gathered, stored, managed, or analyzed with traditional tools (Sun et al., 2019). Big data analytics uses technology to improve data processing and, by extension, patient outcomes (Philip et al., 2022). These considerations suggest an implicit link between facilitating conditions and the ability to improve job performance (i.e., performance expectancy) when adopting new technology. Thus, the study's findings related to facilitating conditions support the importance of performance expectancy but undercut the potential negative effects associated with effort expectancy.

6. LIMITATIONS OF THE STUDY

This study had some limitations that must be addressed when evaluating the findings. One limitation was associated with the decision to limit the target population to participants who

worked at research-based healthcare organizations in the United States. Research-based healthcare organizations were selected because these facilities generate and use vast amounts of data (Singh et al., 2018). Thus, big data analytics represents an effective way to analyze, evaluate, and manage data. The selection of the target population excluded individuals working at similar organizations in other countries and individuals working in different types of U.S. healthcare organizations. As a result, the findings should not be generalized to populations outside the United States or employees working in different types of healthcare organizations like rural hospitals or small clinics.

The study used an acceptable quantitative approach to data collection and analysis. However, the choice of methodology did limit the study's scope. Using a closed-ended survey instrument meant that participants did not have the opportunity to share personal insights or attitudes regarding their experiences with big data analytics. This limitation meant that the study could not account for individual differences between respondents, but the methods aligned with the positivist research paradigm (Alharahsheh & Pius, 2020).

Another limitation related to the research design was the decision to conduct a correlation study rather than an experimental or quasi-experimental study. Experimental and quasi-experimental research designs were eliminated because of the difficulty of obtaining access to individuals willing to participate in a study that required control groups and interventions. However, because the variables were not manipulated and the study did not include any intervention between groups, the findings only represent relationships or correlations between the variables. Thus, while changes in performance expectancy and facilitating conditions were significantly correlated with changes in behavioral intentions to use big data analytics, the changes in the independent variables cannot be said to have caused the changes in participants' behavioral intentions.

A final limitation resulted from selecting the UTAUT as a theoretical framework. The UTAUT is a common model for understanding technology adoption (Al-Fahim et al., 2021). However, focusing exclusively on UTAUT variables means that other factors contributing to big data analytics adoption and use could not be analyzed. Cost benefits, efficiency, security, and organizational culture are all considerations that

could influence big data analytics adoption in healthcare settings (Han & Lee, 2021). This limitation could be addressed through the development of a modified UTAUT model. The following section discusses the study's implications for future research.

7. IMPLICATIONS FOR FUTURE STUDY

The study's findings have implications for many stakeholders. Professionals in the healthcare industry can use the study's results to examine their perceptions, attitudes, and experiences using big data analytics in healthcare settings. While this study focused on research-based organizations, the literature identified the clinical benefits of using big data analytics (see Alghamdi & Alsubait, 2021; Philip et al., 2022). Thus, the present study's findings are likely generalizable to clinical settings. The findings and literature suggest that utilizing the benefits of big data analytics would improve outcomes for practitioners, healthcare organizations, and patients through cost reductions, improved diagnostics, and greater access to care.

Healthcare systems and data security regulations vary dramatically from country to country. These differences may influence the use of big data analytics in specific settings, limiting the study's generalizability to other settings. Conducting additional studies comparing the adoption and use of big data analytics in national healthcare systems would provide insights into national differences in healthcare provision. Specifically, differences could be explored between the use of big data analytics in countries with nationally funded healthcare systems (e.g., Canada, the United Kingdom) and countries with private healthcare systems (e.g., the United States, Switzerland).

Big data analytics is an emerging technology characterized by high levels of technical complexity (Lutfi et al., 2022). The present study's quantitative approach and use of closed-ended survey questions did not allow participants to share individual perceptions of big data analytics. Conducting a qualitative study of the barriers to big data analytics adoption would highlight the main obstacles individuals face as they implement new and complex technologies.

Designing an experimental or quasi-experimental study to examine individual attitudes toward big data analytics would allow researchers to determine causal relationships between variables. Additionally, longitudinal

research could be conducted to determine how behavioral intentions to use big data analytics changed over time. Some researchers have noted that factors like effort expectancy lose significance over time (Ayaz & Yanartas, 2020; Yohanes et al., 2020). Documenting these changes would contribute to discussing the UTAUT's efficacy in similar research settings.

Finally, only two constructs associated with the UTAUT were significant: performance expectancy and facilitating conditions. The significance of performance expectancy and facilitation conditions suggests that the UTAUT may not be the most effective theoretical framework when studying big data analytics adoption. Future research could focus on developing a theory that explains big data analytics acceptance and use based on information security risks, implementation costs, and user aversion to technology. Focusing on potential barriers would allow organizations to improve their chances of successfully using or implementing big data analytics.

8. SUMMARY

This study explored factors influencing user acceptance of big data analytics in research-based healthcare organizations in the United States. The study aimed to determine whether four independent variables associated with the UTAUT (i.e., performance expectancy, effort expectancy, social influence, and facilitating conditions) were significant predictors of the behavioral intention to adopt big data analytics. The linear regression analysis indicated that performance expectancy and facilitating conditions were significantly associated with behavioral intentions. Effort expectancy and social influence were not significant predictors.

This study's findings supported previous research linking big data analytics to performance improvements in healthcare and other industries. Additionally, the study supported the importance of organizational systems to facilitate the use of these complex technologies. As organizations become more dependent on data and technological advances, the importance of big data analytics tools will continue to increase. Recognizing how big data analytics can help healthcare professionals and research-focused organizations improve patient outcomes is important to effectively use these technologies. Future research should focus on identifying barriers to big data analytics adoption.

9. REFERENCES

- Ajimoko, O. J. (2019). Considerations for the adoption of cloud-based big data analytics in small business enterprises. *Electronic Journal of Information Systems Evaluation*, 21(2), 63–79. <https://academic-publishing.org/index.php/ejise/article/view/130>
- Al-Fahim, N. H., Abdulgafor, R., & Qaid, E. H. (2021). Determinants of banks' customer's intention to adopt Internet banking services in Yemen: Using the unified theory of acceptance and use of technology (UTAUT). In *2021 International Congress of Advanced Technology and Engineering* (pp. 1-8). IEEE. <https://doi.org/10.1109/ICOTEN52080.2021.9493448>
- Alghamdi, A., & Alsubait, T. (2021). Healthcare analytics: A comprehensive review. *Engineering, Technology & Applied Science Research*, 11(1), 6650-6655. <http://dx.doi.org/10.48084/etasr.3965>
- Alharahsheh, H. H., & Pius, A. (2020). A review of key paradigms: Positivism vs interpretivism. *Global Academic Journal of Humanities and Social Sciences*, 2(3), 39-43. <https://doi.org/10.36348/gajhss.2020.v02i03.001>
- Aljarboa, S., & Miah, S. J. (2020). Assessing the acceptance of clinical decision support tools using an integrated technology acceptance model. In *2020 IEEE Asia-Pacific Conference on Computer Science and Data Engineering* (pp. 1-6). IEEE. <https://doi.org/10.1109/CSDE50874.2020.9411594>
- Ayaz, A., & Yanartas, M. (2020). An analysis on the unified theory of acceptance and use of technology theory (UTAUT): Acceptance of electronic document management system (EDMS). *Computers in Human Behavior Reports*, 2(1), Article 100032. <https://doi.org/10.1016/j.chbr.2020.100032>
- Barkoczi, N., & Lobontiu, G. (2020). Theoretical aspects of the acceptance of new technologies on the smartphone market. In *IOP Conference Series, 200*, Article 012060. <https://doi.org/10.1088/1757-899X/200/1/012060>

- Batko, K., & Ślęzak, A. (2022). The use of big data analytics in healthcare. *Journal of Big Data*, 9, Article 3. <https://doi.org/10.1186/s40537-021-00553-4>
- Boldosova, V. (2019). Deliberate storytelling in big data analytics adoption. *Information Systems Journal*, 29(6), 1126–1152. <https://doi.org/10.1111/isj.12244>
- Brock, V., & Khan, H. U. (2019). Big data analytics: Does organizational factor matters impact technology acceptance? *Journal of Big Data*, 4, Article 21. <https://doi.org/10.1186/s40537-017-0081-8>
- Cabrera-Sánchez, J.-P., & Villarejo-Ramos, Á. F. (2019). Factors affecting the adoption of big data analytics in companies. *Revista de Administração de Empresas*, 59(6), 415–429. <https://doi.org/10.1590/s0034-759020190607>
- Chasupa, T.-I., & Paireekreng, W. (2021). The framework of extracting unstructured usage for big data platform. *2021 2nd International Conference on Big Data Analytics and Practices* (pp. 90–94). IEEE. <https://doi.org/10.1109/IBDAP52511.2021.9552131>
- Cuzzocrea, A. (2020). OLAPing big social data: Multidimensional big data analytics over big social data repositories. In *ICCBDC '20: Proceedings of the 2020 4th International Conference on Cloud and Big Data Computing* (pp. 15–19). Association for Computing Machinery. <https://doi.org/10.1145/3416921.3416944>
- Fang, X., Gao, Y., & Jen-Hwa Hu, P. (2021). A prescriptive analytics method for cost reduction in clinical decision making. *MIS Quarterly*, 45(1), 83–115. <http://dx.doi.org/10.25300/MISQ/2021/14372>
- Farias, J. S., & Resende, M. M. (2020). Impact of training on the implementation of a new electronic system and acceptance of new technologies in a federal institution of higher education. *Revista de Administração da UFSM*, 13(4), 773–791. <https://doi.org/10.5902/1983465932624>
- Galetsis, P., Katsaliaki, K., & Kumar, S. (2020). Big data analytics in health sector: Theoretical framework, techniques and prospects. *International Journal of Information Management*, 50(2), 206–216. <https://doi.org/10.1016/J.IJINFOMGT.2019.05.003>
- Han, J. H., & Lee, J. Y. (2021). Digital healthcare industry and technology trends. In *2021 IEEE International Conference on Big Data and Smart Computing* (pp. 375–377). IEEE. <https://doi.org/10.1109/BigComp51126.2021.00083>
- Handoko, B. L. (2019). Application of UTAUT theory in higher education online learning. In *ICEME 2019: Proceedings of the 2019 10th International Conference on E-business, Management and Economics* (pp. 259–264). Association for Computing Machinery. <https://doi.org/10.1145/3345035.3345047>
- Hoque, R., & Rahman, M. S. (2020). Predictive modelling for chronic disease: Machine learning approach. In *ICCD A 2020: Proceedings of the 2020 the 4th International Conference on Compute and Data Analysis* (pp. 97–10). Association for Computing Machinery. <https://doi.org/10.1145/3388142.3388174>
- Jadhav, D. (2021). *Understanding artificial intelligence adoption, implementation, and use in small and medium enterprises in India* [Doctoral dissertation, Walden University]. Walden Dissertations and Doctoral Studies. <https://scholarworks.waldenu.edu/dissertations/10655/>
- Kaur, A., Garg, R., & Gupta, P. (2021). Challenges facing AI and big data for resource-poor healthcare system. In *2021 Second International Conference on Electronics and Sustainable Communication Systems* (pp. 1426–1435). IEEE. <https://doi.org/10.1109/ICESC51422.2021.9532955>
- Lagzian, M., & Pourbagheri, M. (2022). An investigation into affecting factors on acceptance of e-government service counters as a service delivery channel: A case of developing country. In *ICEGOV '14: Proceedings of the 8th International Conference on Theory and Practice of Electronic Governance* (pp. 11–19). Association for Computing Machinery. <https://doi.org/10.1145/2691195.2691244>

- Lambay, M. A., & Mohideen, S. P. (2020). Big data analytics for healthcare recommendation systems. *2020 International Conference on System, Computation, Automation and Networking* (pp. 1-6). IEEE. <https://doi.org/10.1109/ICSCAN49426.2020.9262304>
- Lutfi, A., Alsyouf, A., Almaiah, M. A., Alrawad, M., Abdo, A. A., Al-Khasawneh, A. L., Ibrahim, N., & Saad, M. (2022). Factors influencing the adoption of big data analytics in the digital transformation era: Case study of Jordanian SMEs. *Sustainability*, *14*(3), Article 1802. <https://doi.org/10.3390/su14031802>
- Nazir, S., Khan, S., Khan, H. U., Ali, S., García-Magariño, I., Atan, R. B., & Nawaz, M. (2020). A comprehensive analysis of healthcare big data management, analytics and scientific programming. *IEEE Access*, *8*, 95714-95733. <https://doi.org/10.1109/ACCESS.2020.2995572>
- Panari, C., Lorenzi, G., & Mariani, M. G. (2021). The predictive factors of new technology adoption, workers' well-being and absenteeism: The case of a public maritime company in Venice. *International Journal of Environmental Research and Public Health*, *18*(23), Article 12358. <https://doi.org/10.3390/ijerph182312358>
- Philip, N. Y., Razaak, M., Chang, J., Suchetha, M., O'Kane, M., & Pierscionek, B. K. (2022). A data analytics suite for exploratory predictive, and visual analysis of Type 2 diabetes. *IEEE Access*, *10*, 13460-13471. <http://dx.doi.org/10.1109/ACCESS.2022.3146884>
- Queiroz, M. M., & Pereira, S. C. (2019). Intention to adopt big data in supply chain management: A Brazilian perspective. *Journal of Business Management*, *59*(6), 389-401. <http://dx.doi.org/10.1590/S0034-759020190605>
- Shahbaz, M., Gao, C., Zhai, L., Shahzad, F., & Hu, Y. (2019). Investigating the adoption of big data analytics in healthcare: The moderating role of resistance to change. *Journal of Big Data*, *6*, Article 6. <https://doi.org/10.1186/s40537-019-0170-y>
- Singh, G., Schulthess, D., Hughes, N., Vannieuwenhuysse, B., & Kalra, D. (2018). Real world big data for clinical research and drug development. *Drug Discovery Today*, *23*(3), 652-660. <https://doi.org/10.1016/j.drudis.2017.12.002>
- Skoumpopoulou, D., & Wong, A. (2019). Factors that affect the acceptance of new technologies in the workplace: A cross case analysis between two universities. *International Journal of Education and Development using Information and Communication Technology*, *14*(3), 209-222. <https://files.eric.ed.gov/fulltext/EJ1201573.pdf>
- Sun, Z., Strang, K., & Li, R. (2019). Big data with ten big characteristics. In *ICBDR 2018: Proceedings of the 2nd International Conference on Big Data Research* (pp. 56-61). Association for Computing Machinery. <https://doi.org/10.1145/3291801.3291822>
- Susanto, A., Tamimi, Z., Utami, M. C., Fitriyani, S. A., & Imam, S. (2020). Examining the implications of unified theory of acceptance and use of technology for national library navigation systems. In *2020 8th International Conference on Cyber and IT Service Management* (pp. 1-6). IEEE. <https://doi.org/10.1109/CITSM50537.2020.9268885>
- Vallerie, K., Fahira, N. I., Sebastian, V., & Ilimantara, N. (2021). Usage evaluation of beauty e-commerce with unified theory of acceptance and use of technology (UTAUT). In *2021 International Conference on Information Management and Technology* (pp. 429-433). IEEE. <https://doi.org/10.1109/ICIMTech53080.2021.9535051>
- Vanduhe, V. Z., Nat, M., & Hasan, H. F. (2020). Continuance intentions to use gamification for training in higher education: Integrating the technology acceptance model (TAM), Social Motivation, and Task Technology Fit (TTF). *IEEE Access*, *8*, 21473-21484. <https://doi.org/10.1109/ACCESS.2020.2966179>
- Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of new technology: Towards a unified view. *MIS*

Quarterly, 27(3), 425-478.
<https://doi.org/10.2307/30036540>

Computing Machinery.
<https://doi.org/10.1145/3318299.3318395>

Yohanes, K., Junius, K., Saputra, Y., Sari, R., Lisanti, Y., & Luhukay, D. (2020). Unified theory of acceptance and use of technology (UTAUT) model perspective to enhance user acceptance of Fintech application. In *2020 International Conference on Information Management and Technology* (pp. 643-648). IEEE.
<https://doi.org/10.1109/ICIMTech50083.2020.9211250>

Zhang, X., Yu, P., Yan, Y., & Spil, T. O. (2021). Using diffusion of innovation theory to understand the factors impacting patient acceptance and use of consumer e-health innovations: A case study in a primary care clinic. *BMC Health Services Research*, 15(71), 1-15.
<https://doi.org/10.1186/s12913-015-0726-2>

Zhan, G. (2019). Online forum authenticity: Big data analytics in healthcare. In *ICMLC '19: Proceedings of the 2019 11th International Conference on Machine Learning and Computing* (pp. 290-294). Association for

Zhu, T.-L., & Chen, T.-H. (2021). A patient-centric key management protocol for healthcare information system based on blockchain. In *2021 IEEE Conference on Dependable and Secure Computing* (pp. 1-5). IEEE.
<https://doi.org/10.1109/DSC49826.2021.9346259>

Tools for Success: Their Impact on Salaries in the Data Analytics Job Market

Kathleen S. Hartzel
hartzel@duq.edu

Pinar Ozturk
ozturkp@duq.edu

Palumbo-Donahue School of Business
Duquesne University
Pittsburgh, PA 15282, USA

Abstract

This research examines the data analytics job market, focusing on prominent tools in job advertisements and their salary implications. Analyzing a diverse range of postings for business analysts, data analysts, and other analytics roles, the most sought-after tools were identified: SQL, Tableau, Python, R and Power BI. The study reveals SQL's critical importance for business analysts, data analysts, and business intelligence analysts. Additionally, Tableau surpasses Power BI in popularity, while Python is in higher demand compared to R. The findings also indicate distinct salary trends across specializations. Data analysts witness salary increments for all top five tools. However, for system analysts, these tools do not tend to impact salaries. Data scientist roles prioritize programming, with SQL and Python leading to salary increases. By understanding the current tool trends and their salary implications, stakeholders can strategically position themselves in the data-driven landscape.

Keywords: data analytics, business analytics, analytics tools, salary, online job postings

Recommended Citation: Hartzel, K.S., Ozturk, P., (2024). Tools for Success: Their Impact on Salaries in the Data Analytics Job Market *Journal of Information Systems Applied Research*. 17(2), pp. 45-60, <https://doi.org/10.62273/JPTA5240>

Tools for Success: Their Impact on Salaries in the Data Analytics Job Market

Kathleen S. Hartzel and Pinar Ozturk

1. INTRODUCTION

In the era of rapidly growing data collection, companies find themselves grappling with the challenge of harnessing the wealth of information available. The demand for insightful data analysis and informed decision-making has never been greater. As organizations strive to derive value from the vast amounts of data they accumulate, the role of analytics professionals has become increasingly crucial. The Bureau of Labor Statistics projects analytics jobs will grow by about 35% by 2031, much faster than the average position (Seattle Times, 2023).

Notably, the landscape of analytics is evolving, driven by cutting-edge technologies such as machine learning, generative AI, and other AI tools (Lampropoulos 2023; Ooi et al., 2023). These advancements have revolutionized the way data is processed and insights are derived, enabling organizations to unlock previously untapped potential in their data. As a result, the demand for professionals proficient in these AI-driven techniques has surged, further compounding the need for skilled professionals who can leverage such tools effectively.

In response to this increasing demand, colleges and universities have rushed to establish a diverse range of analytics and data science graduate programs (Batistic & van der Laken, 2019). These educational institutions face the formidable challenge of adapting their curricula rapidly to keep pace with the dynamic landscape of analytics and AI tools. The goal is to produce graduates who can meet the ever-changing needs of the data-centric workplace and possess the skills to harness statistics, programming, machine learning, generative AI, and other AI-driven approaches to derive valuable insights.

Despite these endeavors, the shortage of skilled professionals remains a pressing concern, emphasizing the urgent need to align analytics programs with industry requirements effectively. By cultivating a pool of analytics talent well-versed in highly sought-after analytics tools and techniques, educational institutions can play a pivotal role in empowering organizations to make strategic decisions that leverage the full potential of their data. This research aims to

explore the evolving requirements for high-demand analytics tools and competencies, providing valuable insights for employers, aspiring analytics professionals and curricula designers seeking to excel in the data-driven world.

2. LITERATURE REVIEW

Studies on analytics knowledge, skills and abilities (KSAs) needed in industries and the extent to which universities meet these needs are not new. However, as the landscape of available tools and technologies evolve and widespread adoption of these capabilities increases the level of competition among organizations, continued assessment of industry needs is essential to ensure skilled employees will be available. Using actual postings from online job advertisement portals, such as Career-BUILDER, Dice.com, Glassdoor, Indeed, LinkedIn and Monster.com, is the pervasive approach to measuring real employer demands for data analytics KSAs.

Cegielski and Jones-Farmer (2016) as well as Chiang et al. (2012) discuss the development stages of a business analytics (BA) curriculum and identify critical skills, tools, and languages. Their study reveals that technical skills in SAS, structured query language (SQL), and Excel are essential for entry-level jobs. Watson (2012) suggests that communication skills, SQL and query skills, data mining and data warehousing skills, statistical skills, data visualization, text mining, NoSQL skills, and awareness about emerging topics in analytics are also important.

A study, using 2013-2014 data from LinkedIn, Indeed and Monster.com, analyzed 924 data analyst job advertisements and found that SQL, Tableau, and Python were mentioned in only 11.1%, 4.4% and 3.8% of the listing at that time (Luo, 2016). In close temporal proximity, Leon et al. (2018) analyzed 958 U.S.-based positions advertised during 2014-2015 on Monster.com and Indeed.com and found that different analytics job positions require different skill sets. Their work shows how the expected skill sets differ between four groups of job positions, namely business analyst, data analyst, data scientist, and data analytics manager. Data

scientist positions need depth in technical skills, whereas the other positions need more business-oriented skills with a broad knowledge of the technical tool sets in the business analytics area.

Subsequently, Verma et al. (2019) examined a sample of online job ad postings, from 2016-2017, related to professions such as business analyst (BA), business intelligence analyst (BIA), data analyst (DA) and data scientist (DS) using content analysis. They concluded that structured data management skills are relatively more important for BIA compared to BA (SQL in 30% of BA ads versus 75% of BIA ads). There is also a higher demand for skills in statistical software in case of the BIA compared to the BA, with Excel being the most sought-after tool (60% of BIA ads). DS jobs, on the other hand, require significant programming tools: R in 60%, Python in 45% and SAS in 40% of DS ads.

Using 14,495 online job postings from 2019, seven distinct topics, types of jobs delineated by skill sets, emerged when Almgerbi et al. (2022) applied an LDA algorithm. The search terms used to select the job ads were big data, data science, business intelligence, data mining, machine learning and data analytics. The resulting types of jobs were 1) market analyst, 2) business intelligence (BI) analyst, 3) project manager, 4) software developer, 5) data analytics, 6) data engineer, and 7) machine learning. Of particular interest in this paper is the BI analyst whose skill set includes BI, SQL, dashboard, Tableau, database and finance and the data analyst skill set with statistics, data science, mathematics, python, prediction and optimization.

Johnson et al. (2020) scraped 5,257 business analyst entry-level job postings from Indeed in 2018 and extracted the KSAs listed in the job postings. The results were subsequently validated by surveying experts and focus groups. Their results confirmed that SQL, Python and R were key tools. Similarly, Seal et al. (2020) found that SQL is, by far, the most prominent tool for the database programming competency based upon 3,500 U.S.-based business analytics-related job ads posted in 2018-2019. Their analysis further revealed that R and Python meet most of the demand for programming language competency.

Dong and Triche (2020a) analyzed 9,163 job ads from Indeed for jobs posted for data analysts, business analysts and business intelligence analysts from 2014 to 2018. Among other skills,

they found an increase between 2014 and 2018 in the percent of ads with demands for SQL (from 10.4% to 17.8%), Tableau (from 5.3% to 18.9%), Python (from 2.3% to 10.7%), R (from 4.5% to 11.8%) and Power BI (from .1% to 1.9). Demands steadily increased each year with the exception of R which had a small decrease in 2015 (Figure 1). Furthermore, looking at 746 Indeed job posts for business intelligence and analytics positions in 2019, they reported a SQL requirement in 56.6%, Python in 31.5% and R in 13.7% of the ads. Thus, showing a continuing increase in the demand for tool skills over time. Tableau and Power BI, though not explicitly mentioned, would have been included in the generalized categories of data visualization and analytical (Dong & Triche, 2020b).

Growth in Tool Demand (%)

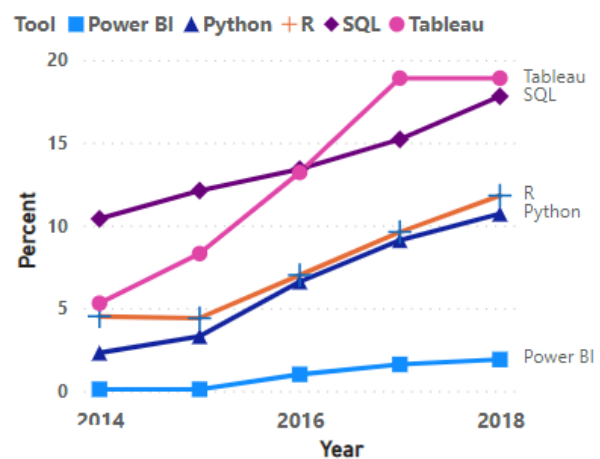


Figure 1: Growth in Tool Demand (adapted from Dong and Triche, 2020a)

Increases in demand could be tied to tool availability. Python was released in 1991, R was introduced in 1993, Tableau was founded in 2003 and Power BI was released as a stand-alone tool in July of 2015.

Stanton and Stanton's (2020) study of over 191,000 LinkedIn job posting during 2019 included a report on the software skill requirements for entry-level positions for data science, data analytic and business analytic jobs. SQL, Tableau, Python and R were found in the top eight most frequency mentioned tools in each job category (Table 1). In each type of position, either SQL or Python had the highest number of occurrences.

	Data Science Positions	Data Analytic Positions	Business Analyst Positions
SQL	23.5%	31.8%	29.9%

Tableau	5.5%	12.0%	12.3%
Python	25.7%	28.1%	24.1%
R	11.4%	16.3%	12.8%

Table 1: 2019 Tools Requirements (Stanton & Stanton, 2020)

A recent LinkedIn analysis of 2,512 ads in 2021, found that SQL, Tableau, Python Power BI and R were the most sought-after data analytics tools. These results highlight the value of resume building skills for emerging data analysts (Zhang et al., 2023).

3. DATA OVERVIEW

The dataset contains 30,569 job advertisements, from Indeed.com, for business analyst-type positions. A Python-based web crawler was used to scrape the web site. All observations were posted during January 2021 and the specific parameters for inclusion were 1) the job ad contains the words 'business' and 'analyst' in the job title and / or detailed description fields and 2) the address listed for the organization is located within 25 miles of a university designated as an R1 research institution.

The dataset includes location information, job titles, company names, job summaries, detailed job descriptions, the length of the job description and when provided salary information

Search Term	Job Category	Number of Ads	Number of Ads with Salary
data analyst	data analyst	764	106
business analyst	business analyst	1,035	132
bi analyst or business intelligence analyst	bi analyst	254	21
systems analyst	systems analyst	170	23
developer	developer	719	84
data scientist	data scientist	482	43
architect	architect	461	59

Table 2: Job Categories

Each job description is categorized as a specific analyst-type position by using keywords. After converting all job titles to lower-case, a conditional statement was used to place each

job title into one category based upon the presence or absence of specific words. Distinctive non-analyst titles, such as architect, data scientist, developer, and systems analyst, are also categorized to distinguish among different professions and functions requiring data analytics skills. The search terms and resultant job categories are shown in Table 2.

Out of the 30,569 advertisements 3,143 jobs contain salary data. This salary data is unstructured where some entries are listed as hourly, weekly, monthly or yearly figures. Some of the listings provide a range of salary figures assumedly dependent on the experience of a candidate. The salary information was normalized by converting all figures to a yearly rate based upon the high-end of the salary range when present. Salaries ranged from \$16,000 to \$500,000 where many of the lower-end jobs are for internship type positions and many of the higher-end jobs are for executive level positions. Although the actual mean salary in the dataset is \$94,333, the average across the different job categories, as shown in Figure 2, is \$110,300. This difference is attributable to the fact that more salary amounts are provided in the lower paying categories. The highest paying category is architect (\$158K), that role is responsible for the data architecture. The data scientist (\$125K) has the next highest salaries listed. These individuals are not typically alumni of business school programs. These professions tend to demand more mathematical and technical skills than the analyst categories. Next in salary level are the systems analyst (\$108K), developer (\$101K) and business intelligence (\$100K) ads which describe jobs that traditional information systems and technology programs, often found in the business school, and computer science departments would prepare students to pursue.

The salary difference between the newer analytics roles of business analyst (\$98K) and data analyst (\$82K) of \$16,000 suggests the importance of domain knowledge in understanding data. The data analyst role does not imply the same level of domain expertise. This suggests that tool skills may be necessary to create business value, but not necessarily sufficient. Business experience and understanding are critical in finding insight and value in data analytics. Figure 2 summarizes the job category salary information.

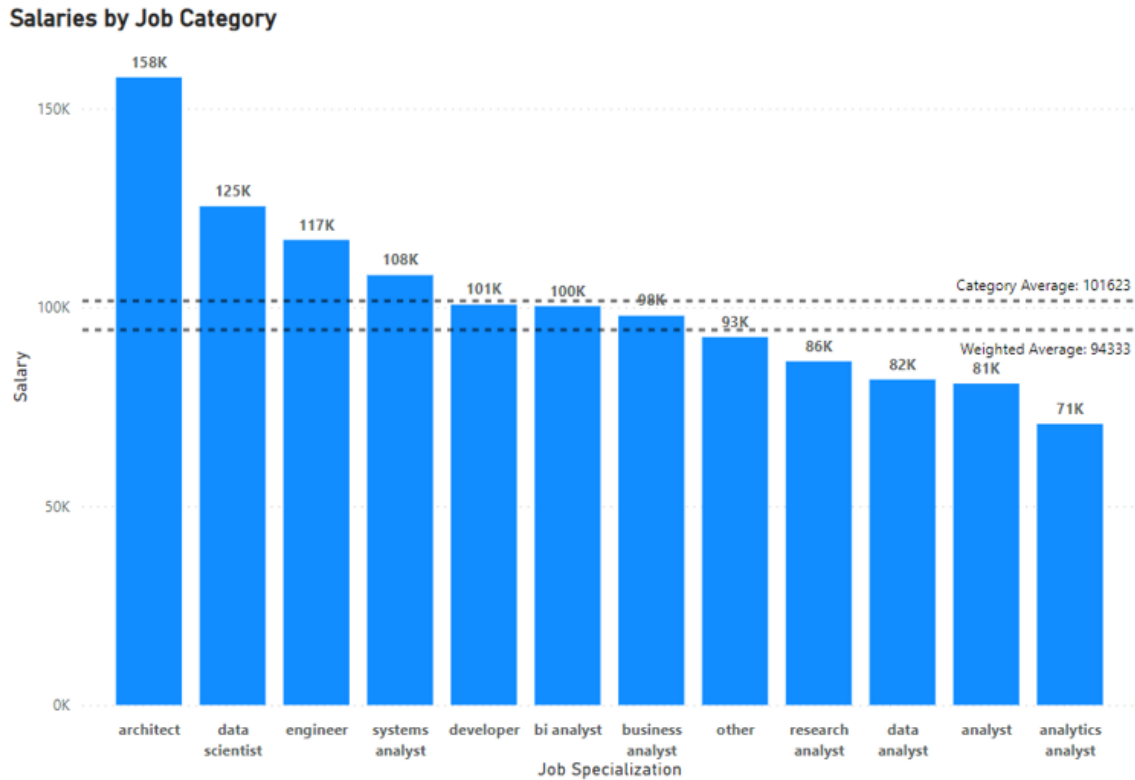


Figure 2: Salaries by Job Categorization

In addition to the salary information, the required (desired) business analytic tools were extracted from the detailed description when available. Required software tools skills are explicitly mentioned in 13,281 (43.4%) of the listings. No specific tools were mentioned in the other 17,288 job ads.

4. DATA ANALYTICS TOOL SETS

The analytics-related tools are presented in Appendix A. A total of 76 different tools are referenced in the dataset. However, these 76 tools are found in 2,238 combinations throughout the 13,281 observations. Some listings mention only one tool and others more. One job ad listed 18 different tools.

SQL, alone and in combination with other tools, is by far the most frequently occurring tool with 7,414 observations. The next four most frequently occurring tools are Tableau (4,682) Python (4,192), R (2,926) and Power BI (1727).

To understand what 'toolsets' are in high demand, the frequency of unique combinations of tools is analyzed. The 20 most frequently occurring single tools, pairs of tools and sets of three co-occurring tools are listed in Table 3.

The five most frequently occurring tools, SQL, Tableau, Python, R, and Power BI, are becoming more commonly taught in business school programs. The next few tools, JAVA, C, JavaScript and Perl, are more geared toward development activity rather than analytics, as are the top five tools.

Counting the occurrences of two tools coupled within a single job ad reveals 67,955 distinct tool-pairs. Examining the tool-pairs, the top seven most frequently occurring tool-pairs are composed of tools found in only the five most frequently occurring tools (SQL, Tableau, Python, R, and Power BI). Suggesting that despite 76 different tools, a limited number of tools and combinations of tools dominates the dataset. Furthermore, despite a large number of tools explicitly mentioned by employers, a smaller core of associated tools meets the vast portion of employers' stated requirements. This is an important pattern for employers, individuals and educators, who are responsible for developing the workforce, to consider when they are assessing both the relative importance among tool skills and also the level of expertise required for each tool.

Single Skills		Paired Skills		3 Skill Sets	
Tool	frequency	Tool	frequency	Tool	frequency
sql	7414	python, sql	3157	python, r, sql	1847
tableau	4682	sql, tableau	3039	python, sql, tableau	1520
python	4192	python, r	2375	python, r, tableau	1206
r	2926	r, sql	2154	r, sql, tableau	1163
power bi	1727	python, tableau	1799	power bi, sql, tableau	756
java	1698	r, tableau	1419	java, python, sql	681
c	1424	power bi, sql	1169	java, javascript, sql	576
javascript	1030	java, sql	1072	power bi, python, sql	551
perl	967	power bi, tableau	1017	python, spark, sql	538
spark	936	java, python	941	hadoop, python, sql	516
sap	907	java, javascript	940	power bi, python, tableau	477
hadoop	810	hadoop, sql	686	qlik, sql, tableau	412
qlik	669	spark, sql	677	power bi, python, r	411
sas	616	python, spark	631	java, javascript, python	406
vba	506	c, sql	626	power bi, r, sql	398
hive	444	power bi, python	622	power bi, r, tableau	372
nosql	421	javascript, sql	593	c, python, sql	367
spss	420	hadoop, python	572	c, java, python	356
ms access	358	qlik, tableau	548	hadoop, spark, sql	356
informatica	319	c, python	518	c, java, sql	345

Table 3: Most Frequently Occurring Tools

Ordinal Tool Acquisition

Merely examining the frequency count of co-listed tool requirements is limiting, because it doesn't facilitate the prioritization of tool skills or the degree of expertise one should possess with tools both individually and conjointly. In developing a program of study to address the demands of the business analytics job market, an individual or institution needs to establish a practical pathway toward identifying and building robust tool skill sets. This section of the paper, presents a series of tree diagrams, using a data-driven approach, that illustrates the actual number of most frequently occurring four-tool sets. There are five diagrams, each one starting with a node representing one of the top five most frequently occurring tool mentions. Given Microsoft Excel's ubiquity in organizations and business schools' curriculum, Excel has been excluded from this analysis.

Each tree diagram models the number of occurrences of specific tools combinations within the data. The number under the tool node represents how many times that tool and any preceding tools on a particular branch occurs in the dataset. For example, in Figure 3, the SQL tree, SQL is found in 7,414 ads. In the second tier, SQL and Python occur together in 3,157 ads. The third tier indicates that SQL, Python and R occur together in 1,874 ads. Finally, the fourth tier shows that SQL, Python, R and

Tableau are jointly listed in 1,017 job ads.

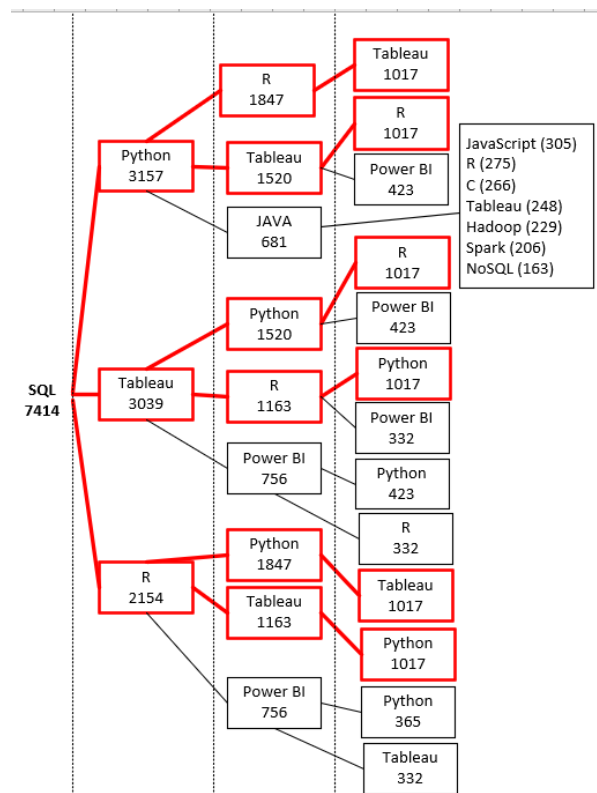


Figure 3: SQL Tree

The heuristic for inclusion on the tree diagrams

requires that the number of occurrences for a node on a specific branch must be at least 20% of the number of occurrences of the previous tier's node. For added clarity, the branches are assembled so that within a specific branch, the more frequently occurring tools are shown higher on the diagram.

As shown in the Figure 3, SQL was found 7,414 times in the job ads. Thus, SQL is mentioned in 47% of the job ads requiring tools. Following the 20% rule, the tools paired with SQL must occur at least 1,483 times ($7,414 * .2$) for inclusion as a second-tier node. Python, Tableau and R meet this criterion. In the third tier, SQL, Python, Tableau and R, continue to dominate. Java and Power BI are the only major new entries onto the tree diagram. The only new tools shown on the fourth-tier stem from the Java branch. Given the SQL-Python-Java branch is relatively low in observations (681), the 20% heuristic requires only 136 observations to be included. JavaScript, R, C, Tableau, Hadoop, Spark, and NoSQL are added to this branch. The six highlighted branches each occur 1017 times. These branches are identical in content except for the order in which the tools (SQL, Python, Tableau and R) are introduced into the model.

Figure 3: SQL Tree

As shown in the other tree diagrams, Figures 4 through 6, the Python, R, and Tableau trees exhibit virtually the same pattern except for the order by which the tools are introduced in the initial seeding of the tree. The tree seeded by Power BI (Figure 7), exhibits a different pattern. In the Power BI tree, tier two contains the same four dominate tools observed in the previously discussed diagrams (SQL, Tableau, Python, and R). However, the early commitment to Power BI, a less dominant tool in the 2021 dataset, leads to a significantly a smaller number of occurrences of tool combinations on all subsequent tiers. The only additional tool meeting the 20% requirement is Qlik, a business analytics platform.

Moving beyond the five most frequently occurring tools, the pattern in the data changes. For example, the sixth tool on the most frequently referenced list is Java (Figure 8). The tree diagram appears to be more geared toward big data and development activity, than business analytics skills that would be most characteristic of business professionals.

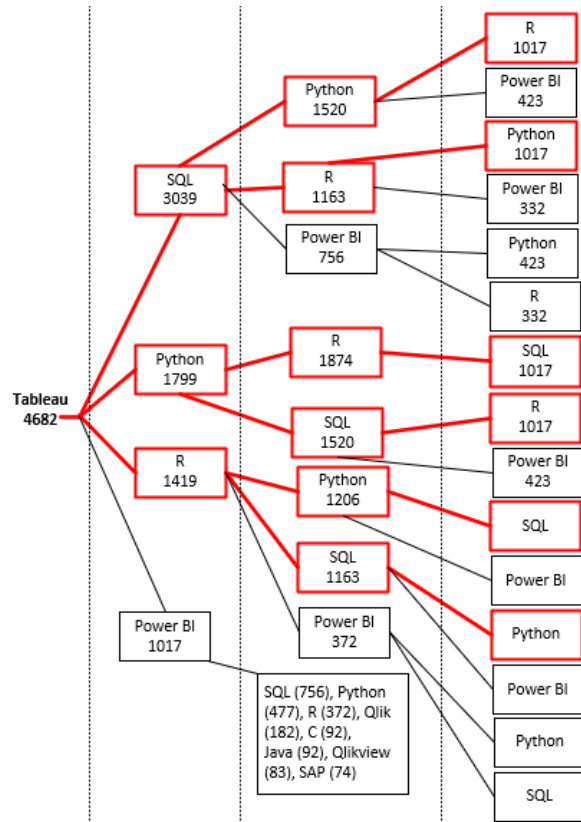


Figure 4: Tableau Tree

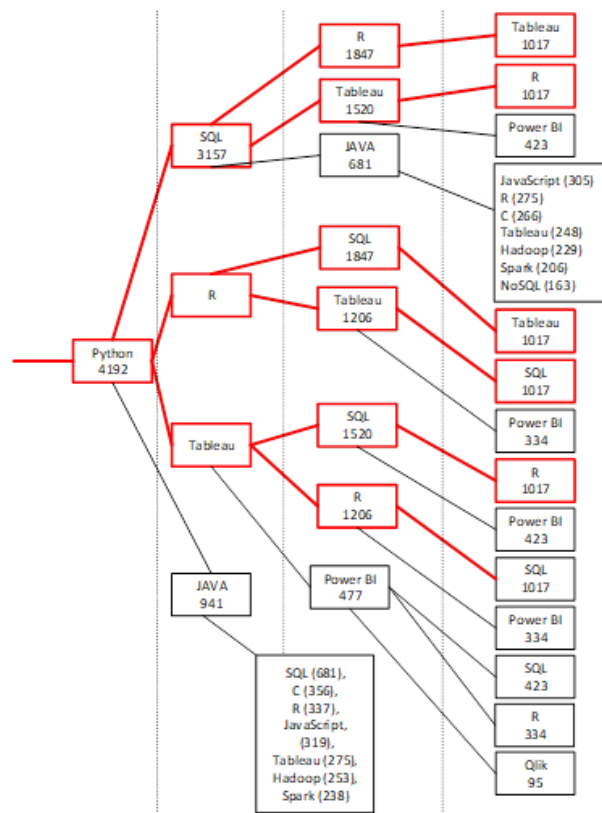


Figure 5: Python Tree

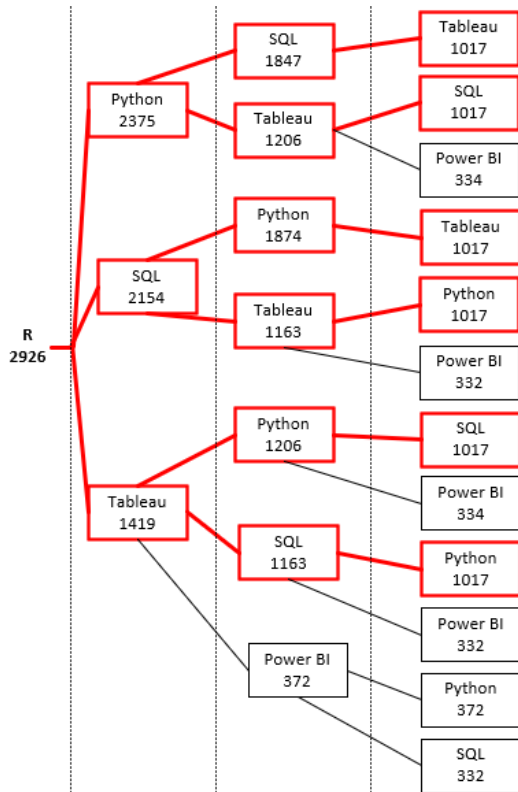


Figure 6: R-Tree

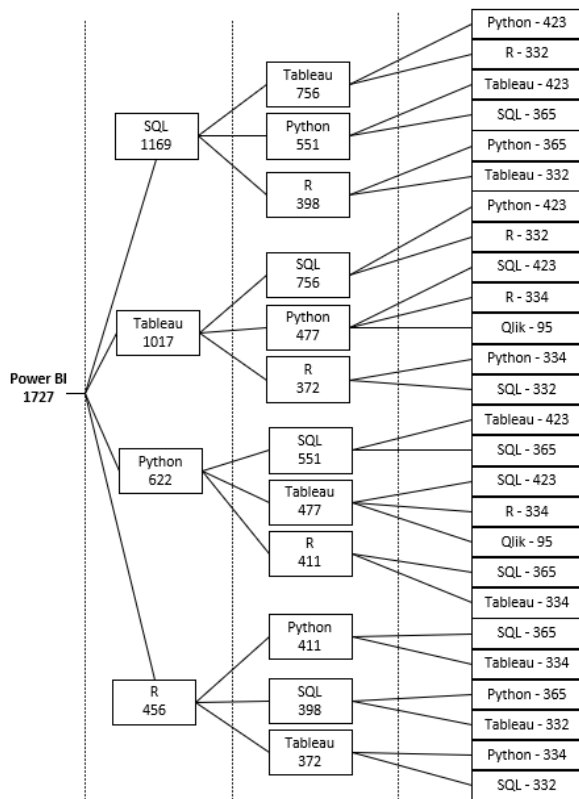


Figure 7: Power BI Tree

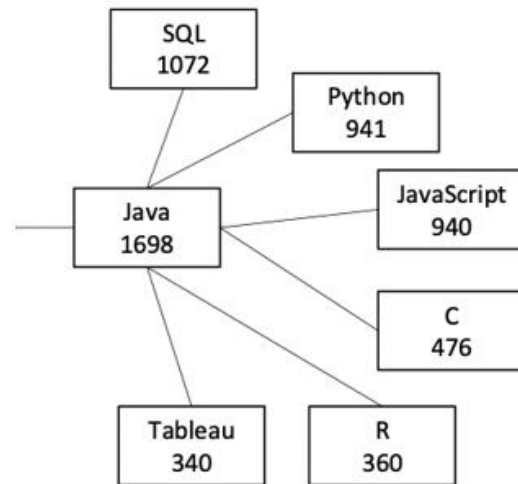


Figure 8: Java Tree

Rank	4-Tool Set	# of Job Ads
1	python-r-sql-tableau	1017
2	power bi-python-sql-tableau	423
3	power bi-python-r-sql	365
4	power bi-python-r-tableau	334
5	power bi-r-sql-tableau	332
6	java-javascript-python-sql	305
7	hadoop-python-r-sql	289
8	hadoop-python-spark-sql	281
9	java-python-r-sql	275
10	c-java-python-sql	266
11	python-r-spark-sql	265
12	java-python-sql-tableau	248
13	hadoop-java-python-sql	229
14	python-qlik-sql-tableau	221
15	hadoop-python-sql-tableau	215
16	java-python-spark-sql	206
17	hadoop-hive-spark-sql	203
18	hadoop-hive-python-sql	201
19	hive-python-spark-sql	193
20	c-python-r-sql	187

Table 4: Most Frequency Occurring 4-Tool Sets

Given the 20% heuristic, this tree's expansion is stunted at two tiers. Java's second tier includes SQL, Python, JavaScript, C, R and lastly Tableau. Tableau, with the smallest number of

Salary Amount #Ads (Ads w/ Salary)	SQL	Tableau	Python	R	Power BI
SQL	\$105.9K 7414 (633)	\$108.1K 3039(243)	\$109.4K 3157(234)	\$115.5K 2154(191)	\$111.1K 1169(114)
Tableau		\$102.1K 4682(374)	\$112.7K 1799(137)	\$117.5K 1409(124)	\$111.2K 1017(108)
Python			\$109.1K 4129(298)	\$112.3K 2375(190)	\$121.0K 622 (62)
R				\$110.0K 2926(249)	\$127.6K 456(52)
Power BI					\$101.2K 1723(171)

Table 6: Average Salaries with Two Most Common Tools

occurrences, is the only exclusive business intelligence, big data, visualization tool in this tool group. SQL is a query language and the other four nodes are all programming languages.

Absence any context, the acquisition of skills for a four-tool set, SQL, Python, R, and Tableau, seems to be the clear choice. As shown in Table 4, this most frequently occurring 4-tool set occurs at a frequency of 2.4 times more than the second most frequently occurring 4-tool set. However, the five most frequently occurring 4-tool sets contain only the five most frequent tools in the data set in every combination possible (SQL, Tableau, Python, R, and Power BI). As observed in the tree diagram analyses, the trees seed with the SQL, Tableau, Python, and R node had little variation. However, the tree originating with the Power BI has more branches signifying the pervasiveness of the tool across a more diverse platform of tools.

The next six most frequent 4-tool sets do not contain Tableau or Power BI. These sets each continue to include python. Additionally, they introduce one or two lower-level programming languages such C or Java. This suggests that the required skill sets may be more technical or big data skills and less applied business analytics.

5. TOOLS AND SALARY

Out of the 3,143 ads with salary data, 1,230 ads list the tools that are required for that position. The average salary for ads with tools is \$97,446 which is \$5,115 higher than the average of \$92,331 calculated for the 1,913 jobs with no tools specified. The inclusion of requirements for the five most frequently mentioned tools seems to have a positive effect on the posted salary level.

Table 5 shows the average of the salary posted when only one of the top five tools is mentioned. Although interesting, the exceptional low

number of instances with salary where only Python or R is mentioned suggests that this information alone is not definitive.

Stand Alone Tool Mentions	Avg Salary	n	n with Salary
SQL	\$98.1K	1215	109
Tableau	\$93.4K	838	76
Python	\$96.2K	92	4
R	\$55.3K	78	3
Power BI	\$71.2K	225	18

Table 5: Average Salaries for One Tool Ads

Table 6 shows how the top tools contribute to overall salary averages when they are combined with another tool. Example, if SQL is included in an ad regardless of any number of other required tools, the average for the ads is \$105.9. However, if both SQL and Tableau are mentioned in the ads, the average increases to \$108.1K. In all cases, the inclusion of two of the five most frequently mentioned tools is greater than the average of ads referencing only one of the top five skills. This indicates that any combination of the two tools should increase both potential salary offers and the number of ads that match a candidate's skill set.

Tool and Job Category

To understand the distribution patterns of the tool usage and salary trends, the data is sliced based on the job categories that were shown in Table 2. Figures 9 through 15 show how the average salary for each job category varies with and without the mention of the five most frequently required tools. These also show how many job ads request a tool and how many of those that carry salary information for each job category and tool.

Salary information is available for 132 of the

1,035 business analyst ads in the dataset. Figure 9 shows the average salary per requirement varies from \$77K for ads requiring R to \$108K for ads requiring Power BI. Ten (117-107=10) of the 132 ads with salary information specifically ask for Power BI experience. Those 10 ads looking for Power BI experience carry a \$11K premium moving the average from \$97K without Power BI mention to \$108K with Power BI mention. SQL and Tableau also seem to have a positive impact on salary level. SQL registers a \$6K gain and Tableau yields \$11K. On other hand, Python and R are associated with a decline in the salary offers for business analysts of \$20K and \$23K respectively.

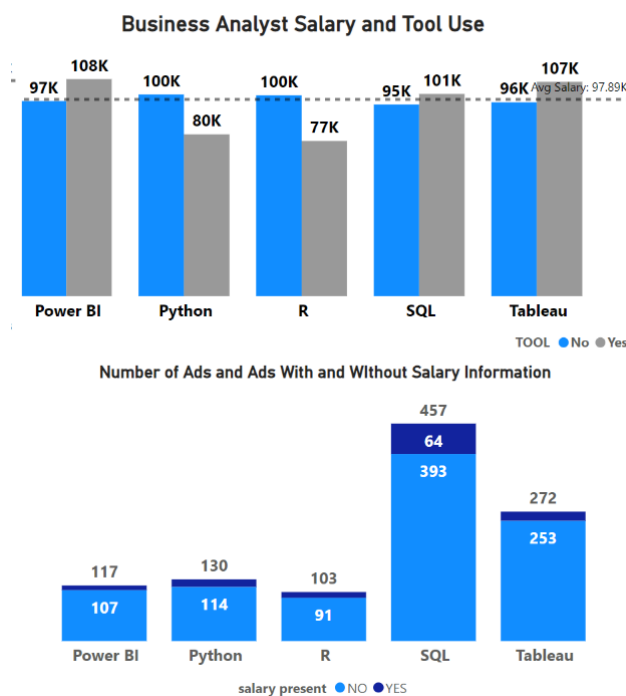


Figure 9: Business Analyst Overview

The BI analyst category occurs a total of 254 times with 21 of those ads carrying salary data. Power BI, Python and Tableau, occur with salary data nine, five and 11 times, with observed decreases of \$17K, \$17K, and \$13K respectively. R, with only three ads with salary data, and SQL, with 18 salary observations, exhibited increases in the average salary of 2K and 33K (Figure 10).

For the data analyst category (Figure 11), there are 106 ads with salary data out of 764 job ads for the category. Each of the five tools is associated with a positive impact on salary. Tableau in the lowest salary number at \$85K. At the high-end of the continuum, both Power BI

and R average compensation of \$90K.

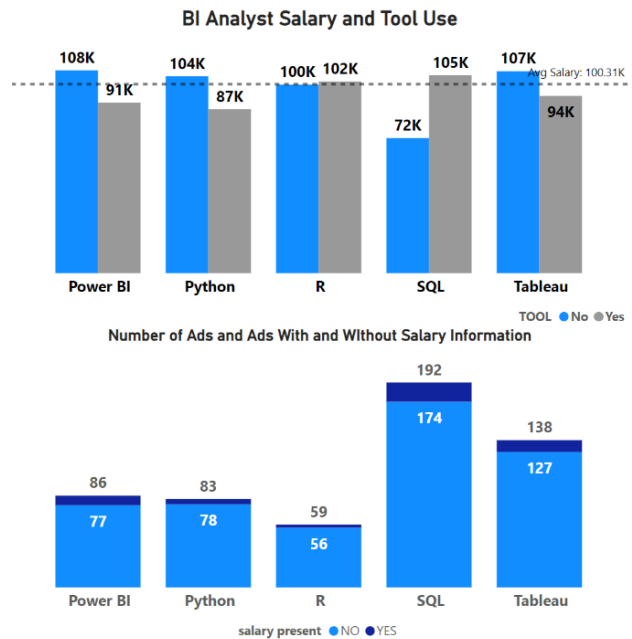


Figure 10: BI Analyst Overview

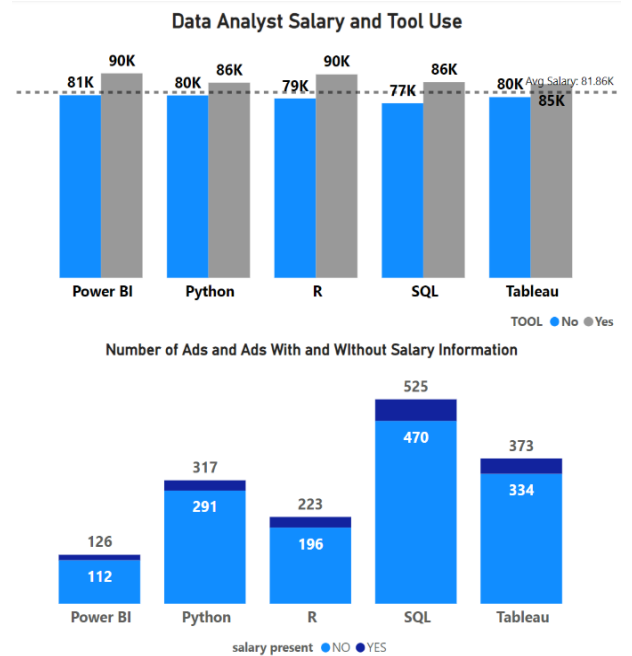


Figure 11: Data Analyst Overview

The 482 job ads in the data scientist category contain 43 listing with salary data (Figure 12). In this category, Power BI (5 ads with salary data and 21 ads without salary data), R, and Tableau are associated with lower salary levels. Python and SQL do carry premiums. Although the Python differential is significant, \$22K, the

SQL difference is marginal at \$2K.

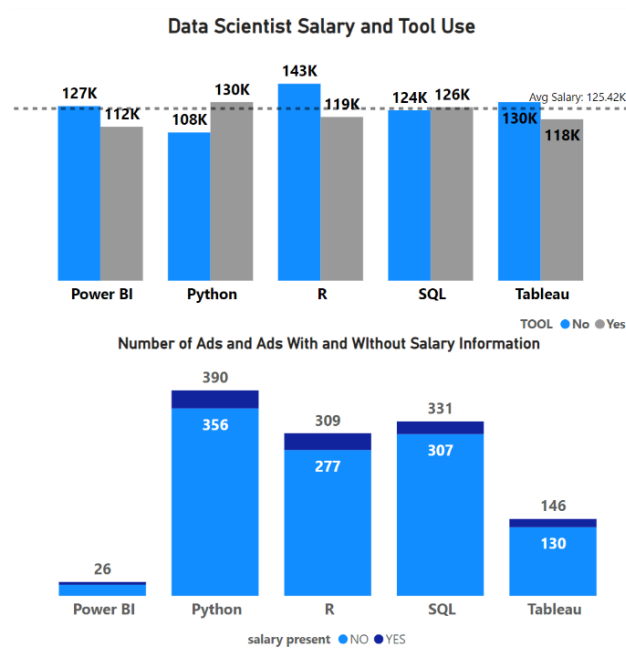


Figure 12: Data Scientist Overview

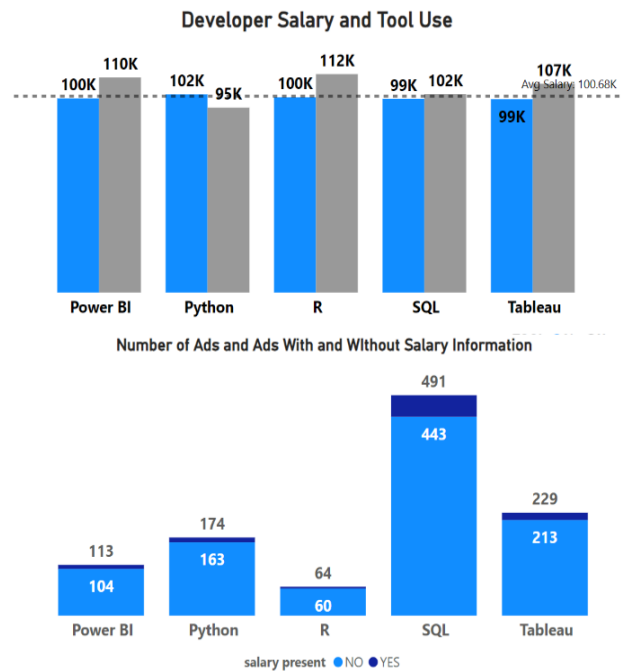


Figure 14: Developer Overview

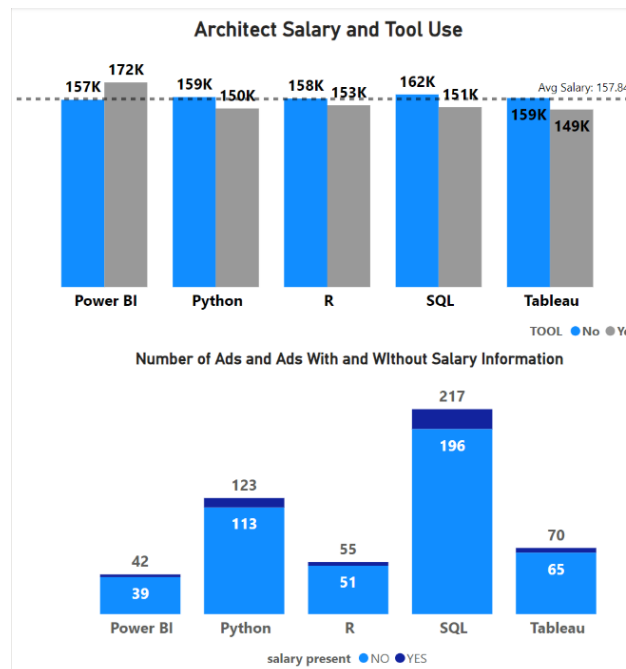


Figure 13: Architect Overview

For the data architect, 59 of the 461 job ads carry salary data (Figure 13). Only the Power BI tools yield a premium. The \$15K boost is calculated on only 3 salaried observations. This suggests that data analytics tools are not important to the data architect.

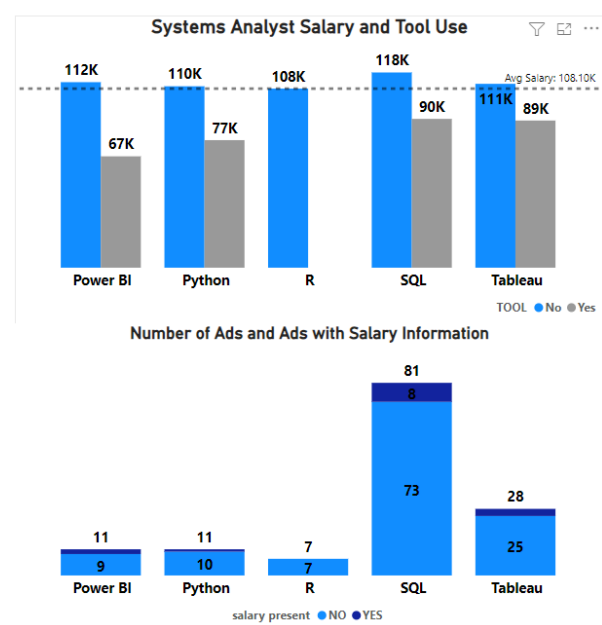


Figure 15: Systems Analyst Overview

Figure 14 shows that Power BI, R, SQL and Tableau all lead to higher salary averages for the developer. Python alone did not suggest a premium for the systems developer.

Finally, the systems analyst category (Figure 15) exhibited a pattern in this dataset that suggests

data analytics tools are not valued in that type of position.

Job Category	Without AI-ML		With AI-ML	
	Salary	n	Salary	n
business analyst	\$98K	129	\$90K	2
BI analyst	\$100K	19	\$108K	2
data analyst	\$81K	97	\$88K	7
data scientist	\$83K	4	\$130	25
developer	\$102K	80	\$83	4
systems analyst	\$108K	23	-	0
architect	\$151K	52	\$210K	7

Table 7: Salary and AI-ML Mentions

The job descriptions were also searched for any mention of artificial intelligence, machine learning or deep learning. Table 7 shows the averages with and without the AI-ML mention for each job category. The business analyst and developer averages were higher when AI-ML was not present. The BI analyst, data analyst, data scientist and architect averages were higher when AI-ML was present in the description. The number of data scientist ads mentioning AI-ML is higher than the number of ads not mentioning the technology. This is the only instance where a specific skill is mentioned in more ads than it isn't. Also, the systems analyst category is not in the table because this category contains no mentions of artificial intelligence, machine learning or deep learning.

6. DISCUSSION

Job ads for business analysts and data analysts are the most common positions in the dataset. The dataset also included ads for positions in various roles, such as BI analysts, data scientists, data architects, systems analysts, and software developers.

With the present emphasis on analytics, it is intriguing to observe how the stated demands for the most requested tools vary across the job titles such as business analyst, data analyst, BI analysts and data science (Figure 16). SQL emerges as the dominant tool with the highest percentage of mentions in each of the job categories, except for the data scientist, where Python is mentioned in 81% of the ads with tools. Tableau was in higher demand than Power BI in all job categories. The data scientist had twice as many mentions of Python and R than any other job category. This aligns with the expectation that the data science would involve

the use of more powerful, versatile tools that are less user friendly.

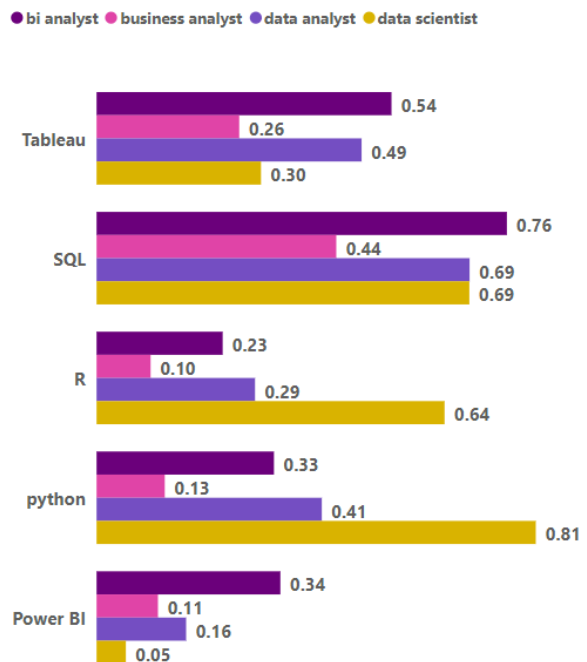


Figure 16: Analytic Roles - Top 5 Tools

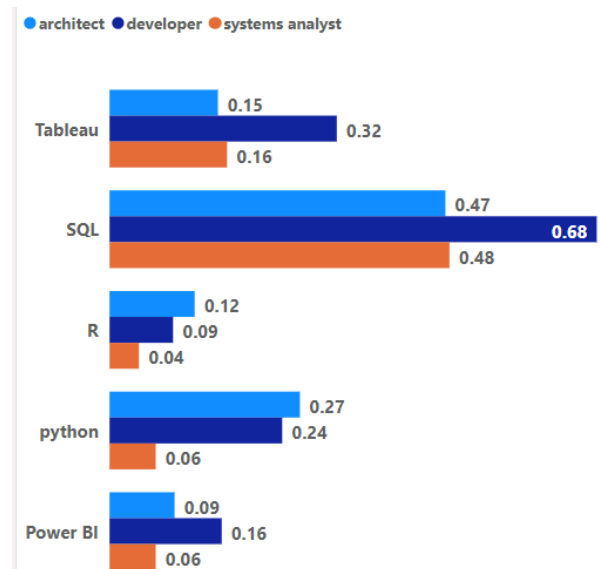


Figure 17: Non-Analytic Roles - Top 5 Tools

The non-analytic roles of architect, developer and systems analyst exhibit a generally lower percentage of jobs requesting these specific tools (Figure 17). However, SQL stands out as a notable exception with continued high demand for SQL skills.

The next four most frequently listed tools are Java, C, JavaScript and Perl. These tools are

primarily geared toward development rather than analytics. Figures 18 and 19 show that demand for these tools is higher for the non-analytical categories of jobs. The demand in the analytics job categories is under 10% of the ads in all instances except for the data scientist with a Java requirement.

Among the non-analytics job categories, developer has the highest mentions of these tools. Among the programming languages, Java and JavaScript were more in demand than C. Perl was only mentioned in 3% to 4% of these listings.

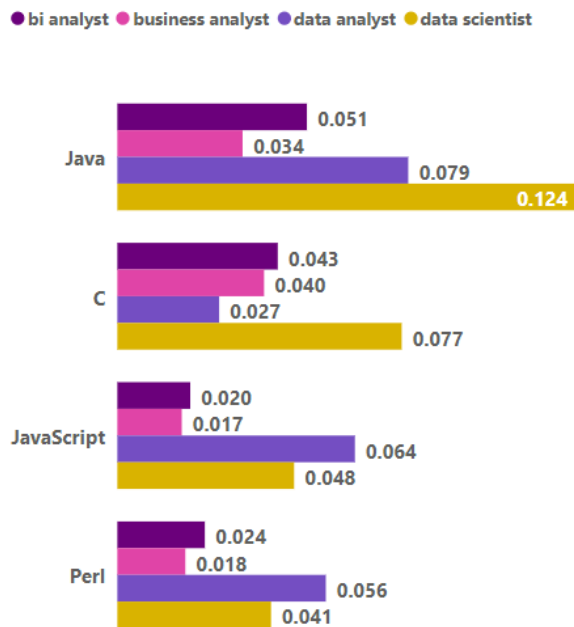


Figure 18: Analytic Roles – Tools 6 to 9

Generally, the mention of tools tends to raise salary expectations. However, this effect is moderated or sometimes reversed when the job category is introduced. Table 8 shows that all five most frequently mentioned tools correlate with an increase in salary for the data analyst exclusively. despite having the lowest salary of \$82,000.

Within an average salary of \$98,000, the business analyst does not benefit from ads listing R or Python. Likewise, the BI analyst (\$100,000) and data scientist (\$125,000) only have 2 tools which seem to positively impact their salaries.

It is important to note that the absence of a tool mention in an ad does not imply that a candidate with that tool on their resume would

not be desirable or valued if selected. It simply means that a specific tool is not the primary skill sought for that particular role. However, as salary increases it may be more likely that skill sets may be assumed and not explicitly stated in the ad, with a candidate’s knowledge and domain experience taking precedence.

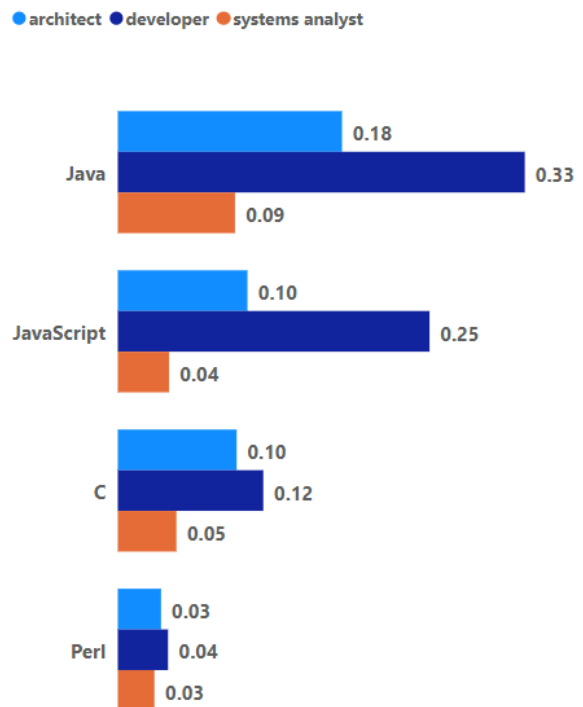


Figure 19: Non-Analytic Roles – Tools 6 to 9

	SQL	Power BI	Tableau	R	Python
Data Analyst	+	+	+	+	+
Business Analyst	+	+	+	-	-
BI Analyst	+	+	-	-	-
Data Scientist	+	-	-	-	+
Developer	+	+	+	+	-
Data Architect	-	+	-	-	-
Systems Analyst	-	-	-	-	-

Table 8: Salary Impacts

Given the salary difference between the business analyst and data analyst categories (\$98,000 vs \$82,000), the importance of business context cannot be overstated. Business analysts must be prepared to both ask and answer insightful business questions, while the data analyst category may lean more toward providing analysis (and visualization) as a service.

7. CONCLUSION

This study's foundation assumes that job advertisements reasonably represent human resource demand in organizations and provide insights into the specific skills and competencies required in the marketplace. However, it's essential to recognize that not all job advertisements may accurately reflect the true skills needed. Organizations may not always fully understand the precise competencies required, leading to the inclusion of multiple competencies in a single position with the hope of finding a candidate possessing some subset of those skills.

A comprehensive examination, of the data analytics tools found in job advertisements is performed to gain valuable insights into the job market for various type jobs, such as business analysts, data analysts, and more. By examining a wide range of job postings, the study addresses questions, such as the prevalence of tools across different job categories and their potential association with higher salaries.

The findings revealed the most common tools associated with each job category, providing valuable guidance for aspiring professionals seeking to enter or advance within the analytics field. Moreover, the analysis sheds light on the potential relationship between specific tools and higher salary offerings in ads. This crucial insight can assist job seekers in understanding which skills may yield greater earnings potential and can aid employers in tailoring their hiring strategies to attract top talent. These findings are also valuable to educators when considering analytics program design.

This dataset, with over 30,000 listing from Indeed.com, shows that SQL is a critical skill for business analyst, data analyst and business intelligence analyst. Additionally, Tableau is more popular than Power BI, and Python is more sought after than R.

As the job market and analytics domain continue to evolve, it is essential for both job seekers and employers to stay informed about the latest trends in tools and their impact on job roles and compensation. This research serves as a valuable resource, providing actionable information for both employers and candidates to make informed decisions in the dynamic landscape of analytics careers, fostering a thriving analytics ecosystem.

Future research is necessary to observe how the

demand and monetary value of specific tools are changing over time. Further research should also assess whether certain tools listed on a resume might signal someone could be overqualified.

8. REFERENCES

- Almgerbi, M., De Mauro, A., Kahlawi, A., & Poggioni, V. (2022). A systematic review of data analytics job requirements and online-courses. *Journal of Computer Information Systems*, 62(2), 422-434.
<https://doi.org/gnnz4k>
- Baticic, S., & van der Laken, P. (2019). History, evolution and future of big data and analytics: A bibliometric analysis of its relationship to performance in organizations. *British Journal of Management*, 30(2), 229-251.
<https://doi.org/10.1111/1467-8551.12340>
- Cegielski, C. G., & Jones-Farmer, L. A. (2016). Knowledge, Skills, and Abilities for Entry-Level Business Analytics Positions: A Multi-Method Study. *Decision Sciences Journal of Innovative Education*, 14(1), 91-118.
<https://doi.org/10.1111/dsji.12086>
- Chiang, R. H., Goes, P., & Stohr, E. A. (2012). Business Intelligence and Analytics Education, and Program Development: A Unique Opportunity for the Information Systems Discipline. *ACM Transactions on Management Information Systems*, 3(3), 1-13.
<https://doi.org/10.1145/2361256.2361257>
- Dong, T., & Triche, J. (2020a). A longitudinal analysis of job skills for entry-level data analysts. *Journal of Information Systems Education*, 31(4), 312.
- Dong, T., & Triche, J. (2020b). Aligning Business Intelligence and Analytics Curriculum with Industry Demand. *AMCIS 2020 Proceedings* 20.
- Johnson, M. E., Albizri, A., & Jain, R. (2020). Exploratory analysis to identify concepts, skills, knowledge, and tools to educate business analytics practitioners. *Decision Sciences Journal of Innovative Education*, 18(1), 90-118.
<https://doi.org/10.1111/dsji.12195>
- Lampropoulos, G. (2023). Artificial Intelligence, Big Data, and Machine Learning in Industry 4.0. In *Encyclopedia of Data Science and Machine Learning* (pp. 2101-2109). IGI Global.
<https://doi.org/mch8>

- Leon, L. A., Seal, K. C., Przasnyski, Z. H., & Wiedenman, I. (2018). *Skills and competencies required for jobs in business analytics: A content analysis of job advertisements using text mining*. In *Operations and Service Management: Concepts, Methodologies, Tools, and Applications* 880-904. IGI Global. <https://doi.org/mch6>
- Luo, W. (2016). Responsibility and Skills Requirements for Entry Level Analytics Professionals. *Journal of Organizational and End User Computing*, 28(4), 1-14. <https://doi.org/10.4018/JOEUC.2016100101>
- Ooi, K. B., Tan, G. W. H., Al-Emran, M., Al-Sharafi, M. A., Capatina, A., Chakraborty, A., & Wong, L. W. (2023). The potential of generative artificial intelligence across disciplines: perspectives and future directions. *Journal of Computer Information Systems*, 1-32. <https://doi.org/mch7>
- Seal, K. C., Leon, L. A., Przasnyski, Z. H., & Lontok, G. (2020). Delivering Business Analytics Competencies and Skills: A Supply Side Assessment. *INFORMS Journal on Applied Analytics*, 50(4), 239-254. <https://doi.org/10.1287/inte.2020.1043>
- Seattle Times. (2023, April 26). High demand surges for data analytics personnel | provided by Northeastern University Seattle Campus. The Seattle Times. <https://www.seattletimes.com/sponsored/high-demand-surges-for-data-analytics-personnel/>
- Stanton, W. W., & Stanton, A. D. A. (2020). Helping business students acquire the skills needed for a career in analytics: A comprehensive industry assessment of entry-level requirements. *Decision Sciences Journal of Innovative Education*, 18(1), 138-165. <https://doi.org/10.1111/dsji.12199>
- Verma, A., Yurov, K. M., Lane, P. L., & Yurova, Y. V. (2019). An investigation of skill requirements for business and data analytics positions: A content analysis of job advertisements. *Journal of Education for Business*, 94(4), 243-250. <https://doi.org/gmjbdtd>
- Watson, H. J. (2012). The necessary skills for advanced analytics. *Business Intelligence Journal*, 17(4), 4-7.
- Zhang, J., Le, T., & Chen, J. (2023). Investigation of Essential Skills for Data Analysts: An Analysis Based on LinkedIn. *Journal of Global Information Management (JGIM)*, 31(1), 1-21. <https://10.4018/JGIM.326548>

APPENDIX A
Business/Data Analytics-Related Tools

apache storm	informatica	obiee	ruby
c	java	octave	sap
ca7	javascript	oracle analytics	sas
caffe	jmp	oracle analytics cloud	scala
cart	json	oracle r	scikit
databricks	julia	otbi	shiny
datawatch	kafka	pandas	spark
django	keras	perl	spss
domo	knime	php	sql
gis	map reduce	pig	sqoop
green plum	matlab	postgresql	swift
hadoop	matplotlib	power bi	tableau
hana edw	microsoft access	python	tensorflow
hcm cloud	modelbuilder	pytorch	teradata
hdfs	mxnet	qlik	trillium
hdinsights	netezza	qlikview	vba
hive	nodejs	r	vmware
hyperion essbase	nosql	rapidminer	weka
ibm watson	numpy	rdbms	yarn

An Action Research Approach to Building an Enterprise-Specific Chatbot (ESCB)

Zach Wood
zkw8126@uncw.edu

Geoff Stoker
stokerg@uncw.edu

Congdon School
University of North Carolina Wilmington
Wilmington, NC 28403 USA

Abstract

Organizations are increasingly turning to chatbots to provide customer support via computer-generated, conversational, natural language answers to human queries. This paper describes a technique for creating an enterprise-specific chatbot (ESCB). We conducted an action research study to investigate the possibility of creating an ESCB with a local policy document knowledge base using readily available software tools, a basic level of programming competence, and user community feedback. The applied research on this chatbot leverages the power of Artificial Intelligence (AI), Natural Language Processing (NLP), and proprietary local data to transcend the typical limitations of conventional chatbots. Utilizing three quick-turn action research cycles, we evolved the chatbot to demonstrate high accuracy and relevance in its responses. The results indicate that our chatbot is becoming increasingly efficient in interpreting user queries, extracting necessary information, and formulating appropriate responses. The work underscores the significant potential of AI-powered chatbots for data interaction and the affordability of AI implementation, paving the way for organizations with limited resources to leverage the power of AI in their local operations.

Keywords: Chatbot, Action Research, Proprietary Data, Large Language Model, AI, Semantic Search

Recommended Citation: Wood E., Stoker, G., (2024). An Action Research Approach to Building an Enterprise-Specific Chatbot (ESCB). *Journal of Information Systems Applied Research*. 17(2), pp 61-73. <https://doi.org/10.62273/RAON2946>

An Action Research Approach to Building an Enterprise-Specific Chatbot (ESCB)

Zach Wood and Geoff Stoker

1. INTRODUCTION

In the 7th decade of the Information Age, it appears that the precept characterized by the famous adage “knowledge is power” (Bartleby, 2023) is evolving to one perhaps better characterized as *applied knowledge is power*. With the exponential growth in data generation, collection, and availability across most domains, organizations are inundated with information. While having access to this data can be crucial, a significant challenge lies in understanding, managing, and effectively using this information for impactful outcomes (Patel & Trivedi, 2020). This is where Artificial Intelligence (AI) and Natural Language Processing (NLP) can come into play, specifically in the form of enterprise-specific chatbots (ESCB).

Chatbots have increasingly become indispensable customer service tools across industries, from commerce to financial services to healthcare. They are designed to improve service delivery, enhance operational efficiency, and offer personalized assistance in an automated yet human-like manner. However, the traditional chatbot implementation can fall short in adaptability and flexibility, particularly when handling local information and dealing with the varied phrasings of user queries (Nuruzzaman & Hussain, 2018).

Commercial website chatbots often adhere to a rigid question-answer pathway, enabling them to process inquiries only within predefined frameworks set by the company. This limited flexibility can result in less-than-optimal user experiences, with chatbots being unable to process or respond to queries outside their programmed parameters effectively (Ayanouz et al., 2020).

Data today can be as diverse as it is abundant, necessitating the ability to effectively harness and process disparate information. Organizations often possess large volumes of proprietary data, within which lie valuable insights regarding organization-specific knowledge. However, this data can be underutilized due to the challenges of integrating and processing it. This situation requires tools, techniques, and procedures to

effectively navigate this data, providing relevant insights and answers, irrespective of query complexity (Ayanouz et al., 2020).

Motivated by a desire to find answers to local organization policy questions more efficiently, we present our action research of the design and development of an ESCB capable of effectively querying local data and interpreting questions posed in various ways. By harnessing the power of AI and NLP to curate proprietary data, we demonstrate an approach for enterprises to unlock the potential of their organization-specific information. This approach can empower users who possess a basic competence with technology, regardless of parent organization, to engage with and interpret complex information, thereby transcending some of the limitations of conventional chatbots.

Our paper outlines using NLP techniques to provide enterprises with an efficient and cost-effective method to query their data. This can potentially avoid the prohibitive time and financial investment often required to train a Large Language Model (LLM) on proprietary datasets (Kalla & Smith, 2023). The advantages of our proposed approach include:

1. Cost-efficiency and time-savings by eliminating extensive training requirements
2. Immediate updates to the underlying knowledge base, allowing for real-time data interaction
3. The ability to pose abstract queries stemming from a wide range of knowledge backgrounds by leveraging an existing LLM
4. Allowing customer service representatives to dedicate their efforts more effectively by automating responses to simple queries

We utilized OpenAI’s Generative Pre-Trained Transformer (GPT) LLM through its Application Programming Interface (API), which serves as an intelligent filter. This enabled us to focus on the applications of our proposal rather than the process of creating an entirely new system.

With regular advancements in the LLM models that OpenAI offers for public use, we can accomplish more with less (Kalla & Smith, 2023). Our approach emphasizes the importance

of only sending the necessary proprietary data to the LLM. By asking the LLM to provide an answer based on a specific chunk of text, we can respond to questions phrased in various ways, including those that the chatbot is not explicitly programmed to answer. This methodology leverages the advanced understanding of the LLM and paves the way for more efficient and intuitive data querying.

The remainder of this paper is organized as follows: Section 2 provides a literature review; Section 3 describes the action research method we followed as well as the major components of the ESCB; in Section 4, we present some results; Section 5 discusses the implications of the results; and Section 6 concludes.

2. LITERATURE REVIEW

Though the precursor for the term Chatbot dates to ~1994 when "ChatterBot" was coined (Mauldin, 1994), computer programs trying to interact with humans via natural language pre-date the term by several decades. ELIZA (Weizenbaum, 1966), probably the best-known early example, was intended as a project to explore how humans and computer programs might communicate. ELIZA (e.g., <https://www.masswerk.at/elizabot/>) used a typing interface comment-response process to imitate the reflection techniques used by Rogerian psychotherapists (Landsteiner, 2005) and occasionally fooled people into believing they were communicating with a human therapist.

In the decades after ELIZA's arrival, the field of NLP developed unevenly but gave rise to increasingly sophisticated chatbots. Some more recent iterations of chatbots, like Mitsuku/Kuki ("Hi, I'm Kuki," 2023), Cleverbot ("Cleverbot," 2023), and IBM Watson ("IBM Watson," 2023), leverage advances in AI and machine learning (ML) to engage in more complex and human-like conversations (Nuruzzaman & Hussain, 2018).

The development of NLP Libraries such as spaCy (Partalidou et al., 2019) has been instrumental in the evolution of chatbots. SpaCy, for instance, provides functionalities like part-of-speech tagging, entity recognition, and dependency parsing, which are crucial for understanding and processing human language. This has significantly enhanced the ability of chatbots to understand and generate responses to user queries.

Semantic searching has also been increasingly utilized in the development of chatbots.

Semantic searching enhances the chatbot's ability to understand user queries by considering the context and intent behind the words rather than just the literal meanings. This results in more accurate and relevant responses (Wei et al., 2008). A key semantic innovation demonstrated in 2013 (Mikolov et al.) was that vector representations of words (also called word embeddings), learned by a recurrent neural network language model, captured relationships to other words that allowed NLP applications to distinguish subtle semantic connections, including the now relatively famous example:

$$v(\text{"King"}) - v(\text{"Man"}) + v(\text{"Woman"}) \approx v(\text{"Queen"})$$

Finally, the advent of AI models like GPT-3 by OpenAI ("GPT-3", 2023) has revolutionized the field of chatbots. These transformer-based models can generate human-like text by predicting the likelihood of a word given the previous words used in the text. They are trained on a diverse range of internet text but cannot explicitly recall or understand information due to their generative nature. They can generate creative, contextually relevant responses, making them ideal for use in chatbots (Nath et al., 2022).

This paper explores ESCB development by combining the capabilities of the spaCy library, semantic searching using scikit-learn (Pedregosa et al., 2011), and the GPT-3.5 model by OpenAI. This approach aims to leverage the strengths of these technologies to create a chatbot that can generate more accurate, contextually relevant, human-like answers to enterprise-specific queries.

3. METHODOLOGY

This paper explores the possibility of constructing an ESCB with readily available tools, basic programming competence, and user community feedback. As we are "addressing questions in one's immediate work environment, with the goal of solving an ongoing problem in that environment" (Leedy & Ormrod, 2010, p. 44), we take an action research approach in this investigation. We follow the canonical action research process model (Susman & Evered, 1978) in Figure 1 (Davison et al., 2004), making three quick turns to refine the development technique and evolve the ESCB. Following this model helps ensure systematic rigor is applied to the problem. Steps include:

- Diagnosis – conduct a thorough examination of the current organizational circumstances
- Planning – the diagnosis results directly

- inform all planning; intended actions should be specified before being undertaken
- Action – planned actions are implemented in the order specified (if any)
- Evaluation – once planned actions are complete, outcomes are compared to project objectives and expectations
- Reflection – explicitly reflect on the activities taken and the outcomes achieved; decide whether to exit the cycle or iterate

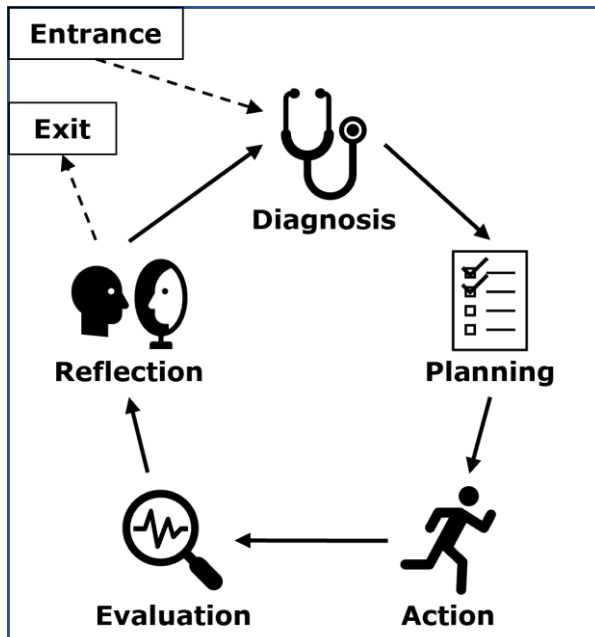


Figure 1: A canonical action research process model (Fig. 1. Davison et al., 2004)

Diagnosis and Planning

In our local organization, finding answers to common enterprise-specific questions most often involves searching for and reading through static web pages, linked portable document format (PDF) policy files, and/or knowledge base articles. It seemed that an ESCB could provide a more dynamic customer-service-oriented user experience. We hypothesized that the state and availability of current AI/NLP tools had evolved to the point that a functional ESCB could be constructed with these tools and the locally available programming talent.

To explore ESCB development, we planned to iterate through cycles that involved tool evaluation, coding, user interaction/feedback, and explicit results reflection to see what we had learned about the evolving chatbot, and the process involved in its development. Our goal was to outline a technique that others interested in creating an ESCB could follow for their own organizational-specific use.

We started with a high-level conceptual sketch of the expected business process for user-chatbot data exchange and refined it as we iterated through the action research cycle. The final conceptual form of the process is depicted in Figure 2. Key components of the process are shown in rounded rectangles at the top, the order of significant steps is listed top-to-bottom, and arrows indicate information flow/exchange between key components at each significant step. Details related to Figure 2 are provided later in the paper. They will necessarily differ somewhat from the simplifying, high-level conceptual sketch, which abstracted away some of the more complicated aspects of the process.

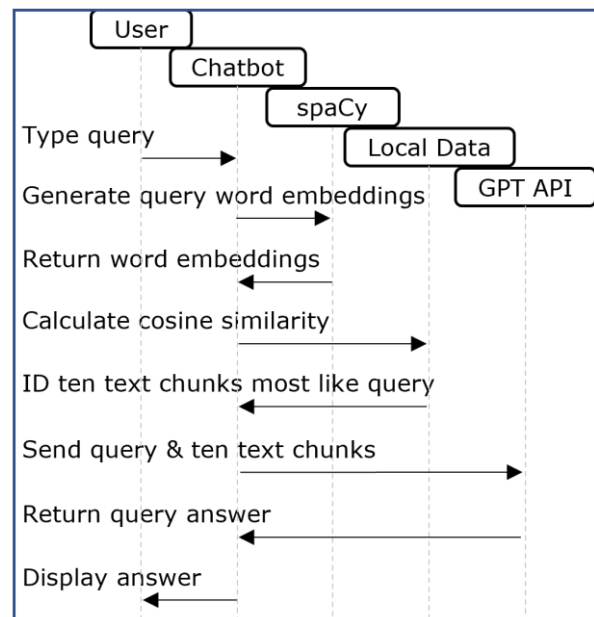


Figure 2: A logical depiction of significant steps in user-ESCB information exchange (top-to-bottom) and key components involved

Action – spaCy and Word Embeddings

Words need to be converted to vectors of numbers to be used in a user query and the text contained in the corpus of local policy documents (CLPD). After a user submits a query, the ESCB loads a medium-sized English-language model from spaCy. This model has been pre-trained on a massive text dataset and has ready-to-use vectors corresponding to thousands of words. To process the user query, the text is first broken into component words or word parts (i.e., tokenized – see Figure 3). The individual words (tokens) are then matched and assigned vectors from the spaCy model. A centroid vector representing the entire query is then calculated from the vectors of the individual

words. One (1) token is equivalent to about four (4) characters in English, while 100 tokens are roughly equivalent to 75 words (Raf, n.d.b).

To process the CLPD, each document is first broken into text chunks of a specified number of tokens (we first used 100, then later 200 tokens), and then each text chunk is transformed into a single vector in the same manner as the user query. After this step, word embeddings exist for the user query and for every text chunk in the CLPD.

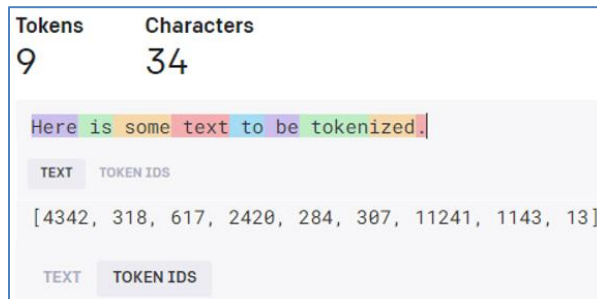


Figure 3: An example of text tokenization
(<https://platform.openai.com/tokenizer>).

Action – scikit-learn and Cosine Similarity

With the user query converted to a single vector and the CLPD converted to many vectors, semantic search can begin. This is done by using scikit-learn to calculate the cosine similarity between the embeddings of the user query and each of the embeddings for the text chunks from the CLPD. Cosine similarity values range from zero (0) to one (1), with zero being no relation and one being a perfect match. Our system sorts the results in ascending order and then selects the 10 chunks with the highest cosine similarity scores. These chunks are the ones that are most semantically like the user query according to the spaCy language model. Thus, at the end of this process, the ESCB has the portions of the CLPD most relevant to the user query and most likely to contain the text that can answer the user's query.

Action – GPT API

With retrieval of the most semantically relevant information from the CLPD complete, we next integrate the GPT API. We bundle together the original plaintext of the user query, the 10 plaintext chunks resulting from the previous step, and the plaintext from the chatbot session history (if any) and make a call to the GPT API. Including the session history enhances the chatbot's ability to understand and respond to user queries in various ways due to the ability to expand on previously answered queries. After

the GPT API returns an answer, it is displayed for the user.

Evaluation

To evaluate the ESCB, we leveraged three different opportunities across the three quick-turn, action research cycles to have users interact with and provide feedback on the chatbot. The first evaluation opportunity was at a local annual Information and Technology Exchange (ITX) event hosted at our university. We set up a table and engaged with the student, faculty, staff, and local IT professional attendees. The second evaluation opportunity was during project presentations in a 400-level course. The third chance to get feedback on the ESCB was at a semi-annual computing showcase featuring student and faculty projects. Across the three presentation opportunities, there were 50 people exposed to the ESCB who either directly interacted with the chatbot or provided feedback on its observed performance. Additional details regarding evaluation and feedback will be provided later.

Reflection

While the reflection phase of action research is enumerated last, it is really an ongoing process. Throughout this applied research activity, we endeavored to use a deliberate reflection activity to both consider the success of the technique for developing an ESCB and to determine whether to proceed with an additional action research cycle. These reflection activities led us to conduct three quick turns through the action research cycle before concluding the ESCB applied research.

4. RESULTS

ESCB Development and Deployment

At the end of our third action research cycle, we had a quite capable ESCB built by a local programmer using readily available AI tools and drawing on information from a 194-PDF-document CLPD. Local community user feedback guided ESCB development, and the chatbot was ultimately able to respond with a reasonable degree of accuracy to questions regarding local policies.

The ESCB was built with a React JS front end and a Python-Flask back end with code organized, as depicted in Figure 4. Key parts include:

- *DocRWidget.jsx* is the front-end parent file.
- *file_read.py* reads the 194 PDF policy docs.
- *chunk_check.py* uses the spaCy language model to create word embeddings for the

user query and 200-token-long chunks of the policy docs, then uses *scikit-learn* to calculate Cosine similarity scores.

- *openai_call.py* interfaces with the user and formats the *text-davinci-003* GPT API call.

***Code details available upon request**

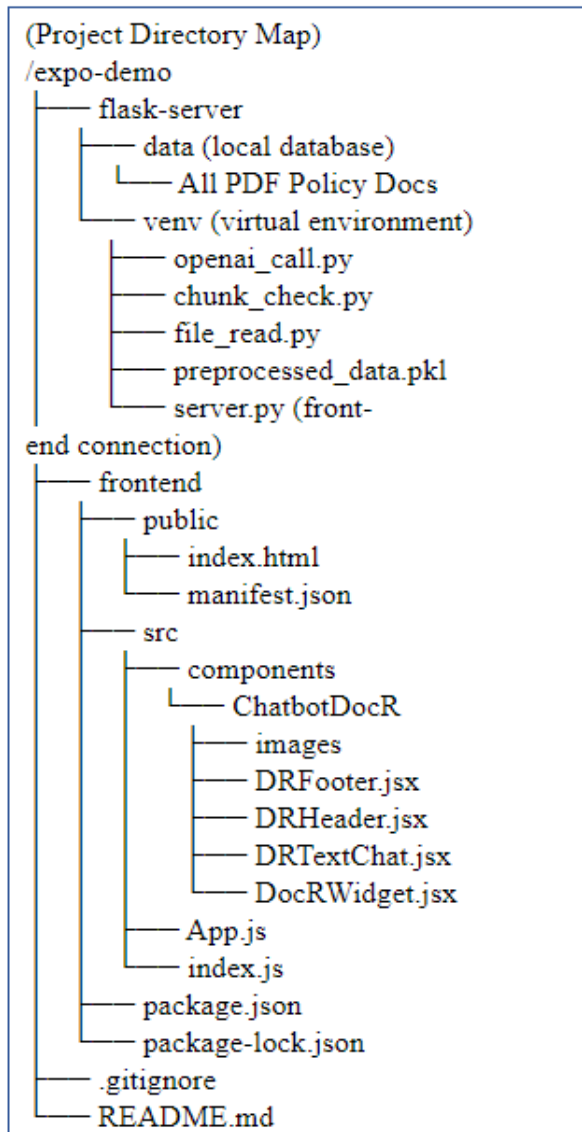


Figure 4: Project directory map

5. DISCUSSION

Several notable points emerge from the development, use, and subsequent evolution of the ESCB. In this section, we discuss the choice of GPT models, the considerations around CLPD text chunk size, and reflect on user community interactions with the ESCB. We also make some observations related to this exploration and

consider aspects for the future. Figure 5 identifies how these key elements relate to each action research cycle iteration.

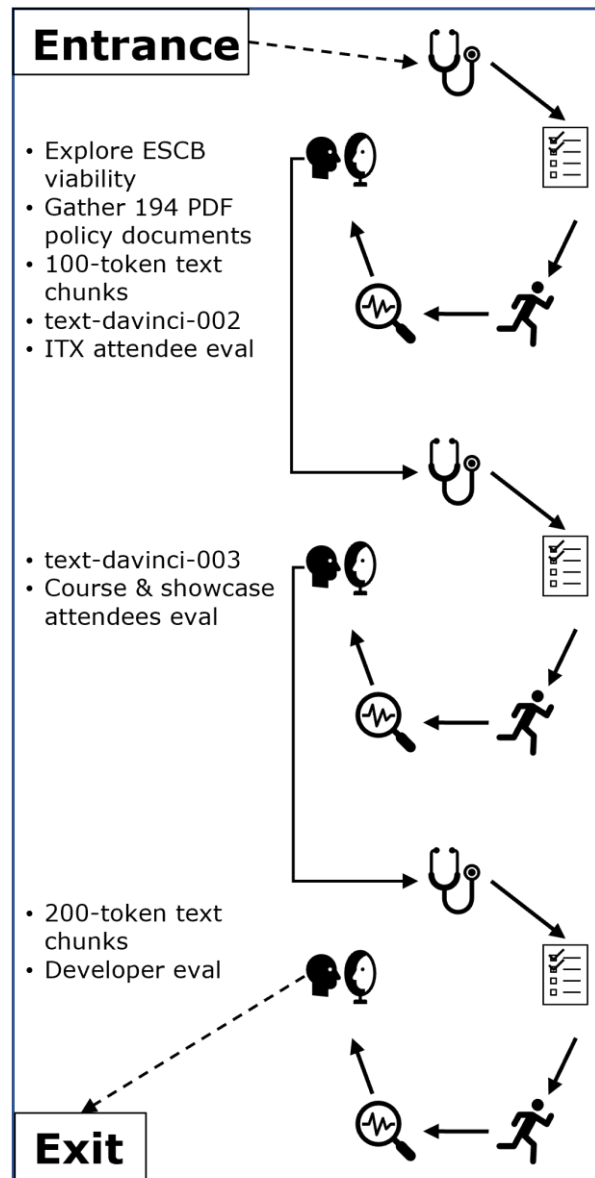


Figure 5: Highlights from the three action research cycle iterations

Language Model

OpenAI consistently evolves its line of large ML foundation models and has thus far produced GPT-1, GPT-2, GPT-3, GPT-3.5, and GPT-4 ("Generative pre-trained transformer," 2023). Models based on GPT-3.5 include *text-davinci-002*, released in March 2022 (Bavarian et al., 2022), and *text-davinci-003*, released in November 2022 (Laike, 2022). The initial ESCB version made use of *text-davinci-002*. This choice was based on the capabilities of the

davinci model line compared to the others (ada, curie, babbage), the relative newness of the model, and the high availability of articles discussing it and coding examples demonstrating its use.

During the evaluation phase of the first action research cycle, it became clear that the ITX participants received the best answers when they wrote queries as specifically as possible. Vague language periodically resulted in API responses that did not fully satisfy the query. During reflection, we decided to investigate the next version of the model, *text-davinci-003*. OpenAI indicated that improvements included producing "higher quality writing," handling "more complex instructions," and being "better at longer form content generation" (Raf, n.d.a). As a result of this information and testing results, we changed the API call code to use the new model.

In subsequent user community evaluations at the course project presentations and the semi-annual computing showcase, there were clear improvements in query response quality. With the model transition, the chatbot began providing better answers that addressed user queries more accurately, even when user phrasing was vague. This switch improved the quality of user-chatbot interactions and demonstrated a key benefit of using these AI tools. Since the model is specified with each API call, it was quite easy to edit the call to use the new model. This bodes well for the future maintainability of an ESCB.

Text Chunks

Key ESCB design considerations included the size, in tokens, of the text chunks to create from the CLPD and to semantically compare to user queries, as well as the number of text chunks to send to the GPT API along with the user query. The two factors involved were GPT token size limitations and cost.

Both *text-davinci-002* and *text-davinci-003* have API request limits of 4,097 total tokens for the combined prompt (user query + text chunks sent) and completion (answer received) (Models - OpenAI API, n.d.; Raf, n.d.b). Assuming a user query length of about 20 tokens and adding the 18 tokens required for prompt formatting (i.e., labels for Question, Relevant Data, Chat History, User, and Response), our initial choice to send 10 CLPD text chunks, each 100 tokens in length, would mean that the GPT API call would be 1,038 tokens and the answer could be a maximum size of 3,059 tokens. This would be

approximately 2,295 words, likely far more than our ESCB answer should require, so more than long enough. A last point to consider regarding length is that our ESCB design intended to take advantage of the semantic context of a chat session, so any subsequent GPT API exchanges included the cumulative chat history (user query + API response) of every previous exchange.

The cost to interact with either *text-davinci-002* or *text-davinci-003* is the same: \$0.02/1,000 tokens (Deprecations - OpenAI API, n.d.). So, assuming the same sending size from the previous paragraph (1,038 tokens) and a GPT response of 95 tokens, the cost for this single, simple notional interaction (1,133 tokens) would be \$0.02266 (2.266 cents). A follow-up second query/response would increase in cost to 1248 tokens or 2.496 cents - 20 token second query + 18 token formatting + 1,000 token CLPD text chunks + 115 token chat history + 95 token second response. Note that each interaction incurs a cost based on token count. While the size limit of 4,097 tokens may not be a concern, the cost of a highly active system might begin to burn through a non-trivial amount of money.

After observing the second and third sets of user evaluations of the ESCB and receiving helpful feedback, we reflected that a third iteration of the action research cycle would be worthwhile to explore changing the text chunk size. We had chosen the 100-token size heuristically based on articles read and example code snippets examined. Since there was little concern about exceeding the 4,097 token size limit and our ESCB was only experimentally used as we permitted, thus throttling costs, we decided to double the text chunk size to 200 tokens. This meant that the example first API call would be 2,038 tokens long and permit replies up to 2,059 tokens (~1,545 words), while the cost for the example first call-response would increase to 4.266 cents. Moreover, as with the previous example, subsequent calls in the same session would include the history, so an example second call would cost 4.496 cents. The result, in our estimation, was that the increase in size of the 10 CLPD text chunks sent with each user query did influence the quality of the GPT responses. The difference is difficult to quantify, but in many cases, we subjectively assessed it to be better. Appendix A provides examples using a mix of language models and text chunk sizes, while Figure 6 provides a single example.

User Community ESCB Interaction

One rather conspicuous aspect of user-ESCB interactions emerged quickly and is related to

the relationship between the quality of the user query and the accuracy of the ESCB response. Independent of the ESCB technical aspects of language model and text chunk size, the more detailed the user query, the better. After observing just a few user-ESCB interactions, we perceived three general user query quality categories emerging and called them vague, direct, and detailed. For example:

- **vague**: Can I bring my own device?
- **direct**: As a professor, can I bring my own device?
- **detailed**: As a professor, can I bring my own device to teach my class from and use in presentations?

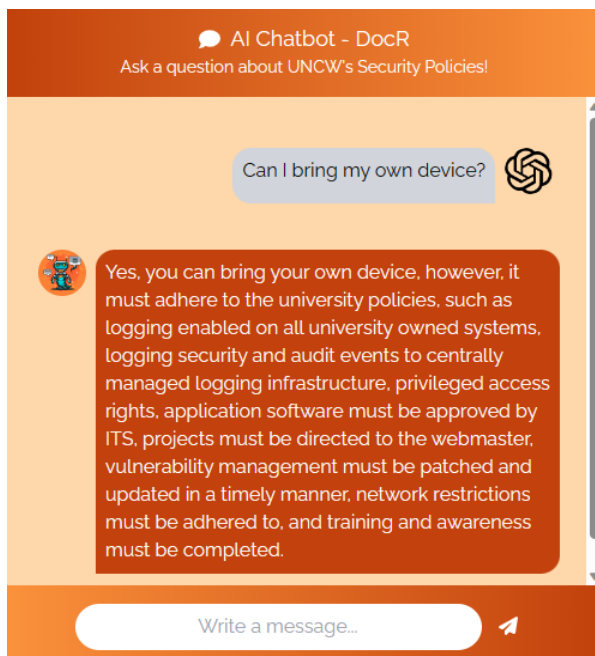


Figure 6: A look and feel of the ESCB; depicted is an answer to the vague phrasing of the example query using text-davinci-003 and 200-token CLPD text chunks.

It was not possible to strictly define the categories and precisely quantify the results, but it appeared that queries we judged as vague returned satisfactory answers regularly (maybe two-thirds of the time), with follow-up queries periodically required. Queries we labeled as direct received satisfactory answers often – about four out of five times on the first try and detailed queries got helpful answers almost every time – better than ~95% of the time.

These observations hold for queries relevant to the CLPD topics and related to our local organization's policies. General

information/knowledge queries were excluded from this analysis.

Another aspect of user-ESCB interactions we qualitatively considered relates to the GPT response quality after changing the language model (*text-davinci-002* to *text-davinci-003*) and text chunk size (100 to 200 tokens). Perhaps unsurprisingly, we noted that changing the language model resulted in improved answers from GPT. There was also an improvement, though arguably less noticeable, in the answer quality when CLPD chunks of 200 tokens were provided to GPT with the user query.

To provide one example, here is a vague query and response using *text-davinci-002* with text chunk token size 100 compared with a detailed query and the response using *text-davinci-003* with text chunk token size 200:

Query: *Can I bring my own device?*

Response: *The university requires that all data storage and processing devices be examined prior to disposal or transfer to surplus. Additionally, privileged access should only be granted to individuals after an evaluation of need and an official approval from the appropriate authority. Furthermore, all university servers must log security and audit events to centrally managed logging infrastructure.*

Query: *As a professor, can I bring my own device to teach my class from and use in presentations?*

Response: *Yes, you can bring your own device to teach your class from and use in presentations as long as it is approved by ITS and all software installed on the device is also approved by ITS. You must also ensure that all data storage and processing devices are examined prior to disposal or transfer to surplus and any institutional or protected data, proprietary software, or software not licensed to be transferred with the computer is removed using ITS Security approved data destruction process or by physically destroying the media. Additionally, all university computers must be kept up-to-date with patches, security, and managed with best practices, and all university data storage or processing devices must be logged and monitored for inconsistencies or unauthorized access.*

Supplemental AI-Impact Observations

In addition to the previous discussion of the technical considerations of language model and

text chunk size/number, we note that this applied research illuminated other points meriting brief mention. These include implications on the gap among technology users of varying familiarity with AI, the challenge of making AI more “human,” and AI affordability.

The development of the technique described to create a usable ESCB demonstrates that one’s understanding of LLM training, NLP, and how semantic search is computationally possible are not prerequisites. This shows that a high level of expertise in AI-related technologies is not required to begin leveraging them to meet an organization’s operational needs. The gap between what highly technical AI practitioners could do a few years ago compared to semi/quasi-technical practitioners has closed remarkably. This has important implications for the democratization of data access, data use, and the broader adoption of AI technologies across diverse sectors and contexts.

One of the challenges of implementing AI in customer service or other interactive scenarios where a human is being augmented or replaced is the need to make the interaction feel “human.” While this is a complex and multifaceted challenge, our chatbot’s performance suggests that a key factor in achieving “human” interaction is the ability to understand and respond accurately to diverse and sometimes vaguely phrased user queries. By evolving the model and improving its interpretative capabilities, we were able to make our chatbot interaction more “human.” Given the parameterized manner in which we accomplished this, we observed that our chatbot’s responses became more articulate and grammatically sound. Although this does not necessarily translate to a more human-like interaction, it signifies a substantial improvement in ESCB capacity to understand and respond to diverse user queries coherently and comprehensively.

Our ESCB development technique also demonstrates how utilizing local data with established chatbot models can enable organizations to leverage LLMs and AI technologies without spending substantial amounts of money on proprietary data training. This opens the possibility for even small businesses and organizations with limited resources to leverage the power of AI in their operations. The per-use cost of OpenAI’s GPT API has already fallen significantly since the conclusion of our applied research project with *gpt-3.5-turbo-instruct*, the currently

recommended replacement for *text-davinci-003* (Deprecations - OpenAI API, n.d.), advertised to cost ~90% less (Pricing - OpenAI, n.d.) at \$0.0015/1000 tokens input and \$0.002/1000 tokens output. The total expenditure for API calls during this research totaled about \$10.50. We estimate that similar work today would cost less than \$2.00 for API calls.

Potential Improvements

There are several areas for improvement to our technique for ESCB development and the resultant chatbot. First, to enhance user interaction, we suggest the introduction of a loading animation during the process of an API request. This addition provides explicit feedback to the user, informing them that a prompt has been sent and a response is expected. Second, we suggest the implementation of a filter that verifies whether the user’s query is within the scope of the database. The current chatbot design sends accompanying information from the CLPD, semantically like the user’s question, with every request, regardless of its relevance.

The goal of an ESCB is to answer a narrow range of questions – in our case, those related to our local organization’s written operational policies. It is not expected to do well with general knowledge queries. Lastly, we propose introducing a feature that enables toggling between a database call and a query about the retrieved information. This feature would enrich user-data interaction and limit queries sent to the API to only contain the necessary information, resulting in cost reduction.

6. CONCLUSION

This paper presented applied research into the development of a technique for building an enterprise-specific chatbot (ESCB) using readily available AI tools, basic programming competence, and user community feedback. We recognize it may be difficult to generalize this action research result because some unnoticed qualities of our environment led to success. However, we believe this technique is sufficiently straightforward as presented that members of other interested organizations can replicate the process and successfully develop their own ESCB for local use.

This paper underscores the significant potential of AI-powered chatbots for data interaction of organizations of all sizes. By presenting an affordable and versatile model, we hope to show how organizations with limited resources can deploy AI in their local operations. Additionally,

this work exhibits the remarkable potential of modern LLMs, making complex subjects accessible to a broader range of individuals and fostering a more inclusive overall information environment.

Due to the substantial cost of training LLMs, not all companies can afford to train a model with their proprietary data. Consequently, organizations wishing to utilize LLMs in their operations require alternative solutions. Our ESCB illustrates a straightforward approach to this problem. By using word embeddings and semantic searches, users can send only the relevant information needed to answer their queries within an API request.

Integrating OpenAI's GPT LLM into our chatbot allows for understanding vague or abstract language phrasing. As a result, users from many backgrounds, not merely those with technical expertise, can interact more comfortably and derive insights from complex topics. This development enhances the democratization of information and underscores the role AI can play in bridging the gap between complex data and its wider audience.

7. REFERENCES

- Ayanouz, S., Abdelhakim, B. A., & Benhmed, M. (2020, March). A smart chatbot architecture based NLP and machine learning for health care assistance. In Proceedings of the 3rd international conference on networking, information systems & security (pp. 1-6). <https://dl.acm.org/doi/10.1145/3386723.3387897>
- Bartleby. (2023). Familiar quotations. Compiled by John Bartlett. <https://www.bartleby.com/lit-hub/familiar-quotations/1970-franisco-1561-1626-john-bartlett/>
- Bavarian, M., Jiang, A., Jun, H. & Pondé, H. (2022, March 15). New GPT-3 capabilities: Edit & insert. OpenAI. <https://openai.com/blog/gpt-3-edit-insert>
- Cleverbot. (2023). <https://www.cleverbot.com/>
- Davison, R., Martinsons, M. G., & Kock, N. (2004). Principles of canonical action research. Information systems journal, 14(1), 65-86. <https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1365-2575.2004.00162.x>
- Deprecations - OpenAI API (n.d.). OpenAI Documentation. <https://platform.openai.com/docs/deprecations/>
- Generative pre-trained transformer. (2023, July 8). In Wikipedia. https://en.wikipedia.org/wiki/Generative_pre-trained_transformer#Foundational_models
- GPT-3. (2023). Models. <https://platform.openai.com/docs/models/gpt-3>
- Hi, I'm Kuki. (2023). @kuki_ai. <https://www.kuki.ai/>
- IBM Watson. (2023). <https://www.ibm.com/watson>
- Kalla, D., & Smith, N. (2023). Study and Analysis of Chat GPT and its Impact on Different Fields of Study. International Journal of Innovative Science and Research Technology, 8(3). https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4402499
- Laike, J. [@janLeike]. (2022, November 28). Check out OpenAI's new text-davinci-003! [Tweet]. Twitter. <https://twitter.com/janleike/status/1597355354433916928>
- Landsteiner, N. (2005). Eliza. <https://www.masswerk.at/elizabot/>
- Leedy, P. D., & Ormrod, J. E. (2010). Practical research, planning and design, 9th edn, New Jersey: Pearson. <https://josemartimast.net/wp-content/uploads/2021/07/AP-Capstone-Research-Planning-and-Designing-E-Book.pdf>
- Mauldin, M.L. (1994, August). Chatterbots, tinymuds, and the turing test: entering the Loebner prize competition. AAAI (Vol. 94, pp 16-21). <https://cdn.aaai.org/AAAI/1994/AAAI94-003.pdf>
- Mikolov, T., Yih, W. T., & Zweig, G. (2013, June). Linguistic regularities in continuous space word representations. In Proceedings of the 2013 conference of the North American chapter of the association for computational linguistics: Human language technologies (pp. 746-751). <https://aclanthology.org/N13-1090.pdf>
- Models - OpenAI API (n.d.). OpenAI Documentation. <https://platform.openai.com/docs/models/gpt-3-5>
- Nath, S., Marie, A., Ellershaw, S., Korot, E., & Keane, P. A. (2022). New meaning for NLP: the trials and tribulations of natural language processing with GPT-3 in ophthalmology. British Journal of Ophthalmology, 106(7), 889-892. <https://doi.org/10.1136/bjophthal>

mol-2022-321141

- Nuruzzaman, M., & Hussain, O. K. (2018, October). A survey on chatbot implementation in customer service industry through deep neural networks. In 2018 IEEE 15th International Conference on e-Business Engineering (ICEBE) (pp. 54-61). IEEE. <https://doi.org/10.1109/ICEBE.2018.00019>
- Partalidou, E., Spyromitros-Xioufis, E., Doropoulos, S., Vologianidis, S., & Diamantaras, K. (2019, October). Design and implementation of an open source Greek POS Tagger and Entity Recognizer using spaCy. In IEEE/WIC/ACM International Conference on Web Intelligence (pp. 337-341). <https://doi.org/10.48550/arXiv.1912.10162>
- Patel, N., & Trivedi, S. (2020). Leveraging Predictive Modeling, Machine Learning Personalization, NLP Customer Support, and AI Chatbots to Increase Customer Loyalty. *Empirical Quests for Management Essences*, 3(3), 1-24. <https://researchberg.com/index.php/eqme/article/view/46>
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. *the Journal of machine Learning research*, 12, 2825-2830. <https://doi.org/10.48550/arXiv.1201.0490>
- Pricing - OpenAI. (n.d.). OpenAI. <https://openai.com/pricing>
- Raf. (n.d.a). How do text-davinci-002 and text-davinci-003 differ? Openai help center. <https://help.openai.com/en/articles/6779149-how-do-text-davinci-002-and-text-davinci-003-differ>
- Raf. (n.d.b). What are tokens and how to count them? Openai help center. <https://help.openai.com/en/articles/4936856-what-are-tokens-and-how-to-count-them>
- Susman, G. & Evered, R. (1978). An Assessment of The Scientific Merits of Action Research. *Administrative Science Quarterly*, (23) 4, 582-603. <https://doi.org/10.2307/2392581>
- Wei, W., Barnaghi, P. M., & Bargiela, A. (2008). Search with meanings: an overview of semantic search systems. *International journal of Communications of SIWN*, 3, 76-82. <http://www.bargiela.com/papers/a35.pdf>
- Weizenbaum, J. (1966). ELIZA – a computer program for the study of natural language communication between man and machine. *Communications of the ACM*, 9(1), 36-45. <https://dl.acm.org/doi/pdf/10.1145/365153.365168>

APPENDIX A

Examples of ESCB Query-Response

Below are query-response examples for a single topic user question using three different question framings, two different language models, and two different text chunk sizes.

The table below refers to the following user query framing examples:

Vague (V): *Can I bring my own device?*

Direct (Di): *As a professor, can I bring my own device?*

Detailed (De): *As a professor, can I bring my own device to teach my class from and use in presentations?*

The Tokens/Cost column reflects the number of tokens in the user query (7, 11, or 20) + the 18 formatting tokens + the number of tokens in the 10 CLPD text chunks (1000 or 2000) + the number of tokens in the GPT response (varies from 52 – 154).

Query	Response	Tokens/ Cost
text-davinci-002 with 100-token CLPD text chunks		
V	The university requires that all data storage and processing devices be examined prior to disposal or transfer to surplus. Additionally, privileged access should only be granted to individuals after an evaluation of need and an official approval from the appropriate authority. Furthermore, all university servers must log security and audit events to centrally managed logging infrastructure.	1086 \$0.02172
Di	The university requires that all individuals who have access to electronic information stored on its information technology resources complete information security training as deemed appropriate by their supervisor, department, or other appropriate authority. In addition, security training may be required based on the sensitivity of the role or compliance requirement.	1084 \$0.02168
De	The university's information technology resources are powerful tools that are shared among all members of the campus community. These resources are intended to be used in useful and productive ways. However, the university cannot protect individuals against the existence or receipt of materials that may be offensive or annoying to them. Additionally, the university requires authorized employees who have access to university information technology resources to take appropriate action when necessary to ensure the orderly administration and function of these resources.	1126 \$0.02252
text-davinci-003 with 100-token CLPD text chunks		
V	Yes, you can bring your own device, provided that you adhere to the relevant data and policies outlined in the conversation history. This includes not posting or disseminating personal or sensitive information, controlling entry points, providing identification when transferring assets off-site, granting privileged access rights only after an evaluation of need and approval, keeping systems current with patches and security, consulting the webmaster before committing to new web-based technology, enabling logging on all university-owned systems, logging security and audit events to a	1125 \$0.02250
Di	Yes, you can bring your own device as a professor, however you must adhere to the policies outlines in the university's Responsible Use, Server Management and Standards, Privacy and Confidentiality, Information Security, and Campus Web Resources policies. This includes not posting or disseminating personal or sensitive information, notifying your department chair, dean, director, or administrator if you have concerns about the conduct of a member of the university community, keeping systems current with patches, security, and managed with best practices	1129 \$0.02258
De	Yes, you can bring your own device to teach your class from and use in presentations as long as you comply with the university's policies and	1136

	procedures regarding the use of electronic information and data processing devices. You should also be aware of the university's policy on server management and standards, responsible use, campus web resources, and information security. Additionally, you should ensure that all systems are kept current with patches, security, and managed with best practices, and that you have completed the necessary information security	\$0.02272
text-davinci-002 with 200-token CLPD text chunks		
V	No, users may not post or disseminate personal or sensitive information about an individual or individuals. Such information includes, but is not limited to, academic records, medical information, social security numbers, or similar information of a personal and confidential nature that, if disseminated, could have legal or otherwise damaging implications either for the targeted person or the institution.	2095 \$0.04190
Di	The university policies regarding user access and registration state that users are responsible for all activity performed with their personal IDs. When a user's relationship with the university is terminated, their access to university computing resources will be denied. IDs cannot be reassigned after being issued.	2081 \$0.04162
De	The university's policy on electronic information states that authorized employees may have access to electronic information stored on university information technology resources in order to ensure the orderly administration and function of these resources. However, the university cannot protect individuals against the existence or receipt of materials that may be offensive or annoying to them. Additionally, the university's policy on server management and standards requires that all software to be installed on university enterprise systems be approved by ITS, and that application owners consult the Webmaster before committing to new web-based technology.	2141 \$0.04282
text-davinci-003 with 200-token CLPD text chunks		
V	Yes, you can bring your own device, however, it must adhere to the university policies, such as logging enabled on all university owned systems, logging security and audit events to centrally managed logging infrastructure, privileged access rights, application software must be approved by ITS, projects must be directed to the webmaster, vulnerability management must be patched and updated in a timely manner, network restrictions must be adhered to, and training and awareness must be completed.	2114 \$0.04228
Di	Yes, as a professor at UNCW, you can bring your own device, provided that you adhere to the university's policies and procedures for responsible use. This includes not posting or disseminating personal or sensitive information about individuals, reporting any concerns you have about the conduct of a member of the university community, or the propriety of a given situation or activity, and keeping your device up-to-date with patches, security, and managed with best practices. Additionally, you must ensure that any software you install on the university's enterprise systems is approved by ITS, and that any privileged access rights you are granted are approved by the appropriate authority and documented. Finally, you must complete any information security training deemed appropriate by your supervisor, department, or other appropriate authority.	2182 \$0.04364
De	Yes, you can bring you own device to teach your class from and use in presentations as long as it is approved by ITS and all software installed on the device is also approved by ITS. You must also ensure that all data storage and processing devices are examined prior to disposal or transfer to surplus and any institutional or protected data, proprietary software, or software not licensed to be transferred with the computer is removed using ITS Security approved data destruction process or by physically destroying the media. Additionally, all university computers must be kept up-to-date with patches, security, and managed with best practices, and all university data storage or processing devices must be logged and monitored for inconsistencies or unauthorized access.	2175 \$0.04350

