

Joseph Annand

DSE5004

Position Paper #1

Introduction

The *data-information-knowledge* continuum outlines how raw data is transformed into meaningful text and visuals so that audiences can perceive patterns and relationships and make meaning of the data (Cairo, 2011). There are several challenges associated with moving along this continuum so that data scientists can help businesses make data-based decisions. Initially, organizational challenges arise when businesses do not have functional, efficient infrastructure that makes data analytics feasible and timely. Even when this infrastructure does exist, though, the quality of data will greatly impact the data scientist's ability to find meaningful insights from it. Lastly, the information obtained from the data must be summarized and delivered to the audience effectively so that a higher knowledge that may be obtained and smartly inform business decisions.

Organizational Challenges

The key organizational challenges in translating data into information from which knowledge can be extracted are centered around data availability through access and sharing. A data scientist must question during any project about what data pipelines and data governance policies are in place so that they have sufficient resources to complete the project in mind. In order to translate data into information that will become meaningful and worthwhile to business decisions, high-quality data must be readily available and easily attainable. One issue that may arise in data availability is the storage of cross-functional data. For example, data from different departments, like finance and sales, may be stored in separate locations, making a difficult for analysts to locate and retrieve data necessary for a particular assignment. Additionally, how this data is catalogued is another factor in data availability. If, for example, the company does not have standardized file naming conventions, it may be difficult to locate specific files or data sets (*7 Best Practices for Successful Data Management*).

Data security is essential for any company, as data breaches are a significant financial risk; however, poor governance policies that inhibit access to data may have a negative impact on analytics and, thus, business efficiency. If data security protocols are too strict, data scientists and analysts will face challenges in retrieving data, which will waste time and money (*What is Data Management? Importance & Challenges*). Companies need to strike the right balance between convenience and security by allowing appropriate individuals access to data while maintaining security through de-identification efforts and protective data storage and transfer methods, such as encryption.

Data Quality

Data quality presents a major issue in the *data-knowledge* continuum. Poor-quality data makes it more difficult for analysts to develop information that will provide valuable insights for business decisions. To begin, an understanding between analysts and those creating data entries needs to be established. Whether it is external data that researchers found and added to the database or it is other company employees entering internal data, data scientists and analysts need to communicate with whomever it is appropriate the needs and expectations they have in data collection and storage so that they can save time and more effectively complete analytical tasks (Redman, 2020). If not, the analysts may find themselves correcting mistakes, like formatting errors, or needing to request additional information from the individual who originally entered the data because crucial features are missing.

Additionally, there is a wide variety data quality features that must be accounted for, such as accuracy, completeness, timeliness, consistency, and data granularity and relevance (Jones, 2023). Without taking these features into account, it will be harder for data scientists to translate data into information, and it puts the company at risk of making poorly informed decisions. Data validation rules and methods should be in place to ensure accuracy; data sets should be complete by handling missing values and merging with other relevant data sets; and data must be kept up to date so that business decisions are informed by relevant information. When data

is too complex, or has excessive granularity, the transfer of information to knowledge may be placed in jeopardy: data that contains unnecessary detail will make it difficult to draw meaningful insights that translate to business actions (Jones, 2023). The same is true if the data is not detailed enough, so finding a balance is crucial to the *data-knowledge* continuum.

How Data is Used

Once data scientists have access to high-quality data, the next major challenge arises in pulling useful information from that data and making important decisions based on some new insights. To achieve this, the data scientist first needs to understand the data and how it will help inform business decisions. This requires a collaborative effort between the analysts and the business users, in which business users, stakeholders, and other consultants relay their needs so data scientists know what to be looking for in the data, what models need to be developed, and if the current data on-hand will be sufficient to achieve the desired results. Once the data scientist has performed their analysis and developed any models necessary, they must relay their findings to business decision-makers. Their success in doing so will be contingent on their understanding of business needs, their skills in analyzing the data, and their ability to summarize and visualize their findings into engaging and understandable content. Poor visualization of their findings can lead to an inaccurate understanding of the data by the audience, which will result in poor business decisions (Stobierski, 2021). This can be avoided by first understanding what is important for business decisions and what decision-makers are looking for. Having that knowledge allows the data scientist to cut down on the detail in their visualizations and draw attention to what is most important to the task at hand. Other important strategies to prevent misleading the audience with visualizations is choosing the right chart or graph to convey your insights and avoiding weird scaling that forces misunderstanding of relationships in the data (Leo, 2019).

Conclusion

Translating data into valuable knowledge to inform business decisions comes with several challenges that can be solved with careful consideration of data availability, data quality, and visualization techniques. Organizational hurdles, such as data availability and security, are significant barriers in accessing and utilizing data. Establishing robust data governance policies, promoting cross-functional collaboration, and striking a balance between data accessibility and security are essential steps in meeting analytical needs. Additionally, the quality of data emerges as a critical factor influencing the effectiveness of data analysis efforts. Collaborative efforts between data scientists and data creators, coupled with the implementation of data validation rules and methods, are imperative to ensure data accuracy, completeness, and relevance. Finally, effectively translating data into actionable insights requires not only technical expertise but also a solid understanding of business needs. Data scientists must carefully summarize and visualize their findings by employing best practices to enhance comprehension that will inform decision-making.

References

- Cairo, A. (2013). *The Functional Art: An Introduction to Information Graphics and Visualization*. New Riders.
- Jones, E. (2023, May 30). *6 Pillars of Data Quality and How to Improve Your Data*. IBM. Retrieved March 24, 2024, from <https://www.ibm.com/blog/6-pillars-of-data-quality-and-how-to-improve-your-data/>
- Leo, S. (2019, March 27). *Mistakes, we've drawn a few. Learning from our errors in data... | by Sarah Leo*. The Economist. Retrieved March 24, 2024, from <https://medium.economist.com/mistakes-weve-drawn-a-few-8cdd8a42d368>
- Redman, T. C. (2020, February 10). *To Improve Data Quality, Start at the Source*. Harvard Business Review. Retrieved March 24, 2024, from <https://hbr.org/2020/02/to-improve-data-quality-start-at-the-source>
- 7 Best Practices for Successful Data Management*. (n.d.). Tableau. Retrieved March 24, 2024, from <https://www.tableau.com/learn/articles/data-management-best-practices>

Stobierski, T. (2021, January 28). *Bad Data Visualization: 5 Examples of Misleading Data*. HBS Online. Retrieved March 24, 2024, from <https://online.hbs.edu/blog/post/bad-data-visualization>

What Is Data Management? Importance & Challenges. (n.d.). Tableau. Retrieved March 24, 2024, from <https://www.tableau.com/learn/articles/what-is-data-management>