



IBM Data Analyst capstone project

Odile Nzambazamariya

March 28/2024

OUTLINE



- Executive Summary
- Introduction
- Methodology
- Results
 - Visualization – Charts
 - Dashboard
- Discussion
 - Findings & Implications
- Conclusion
- Appendix

EXECUTIVE SUMMARY



- Work with data in different formats.
- Collect data from various sources.
- Identify duplicate values in the dataset.
- Remove duplicate values from the dataset.
- Identify missing values in the dataset.
- Determine the missing values in the dataset.
- Normalize data in the dataset.

INTRODUCTION



Managing and analyzing data requires traveling through a variety of forms and sources. This entails deftly managing duties like gathering information from multiple sources, spotting duplicates, and organizing missing values in databases. Normalizing data also improves analytical accuracy and guarantees consistency. Now let's explore these important facets of optimization and data manipulation.

METHODOLOGY



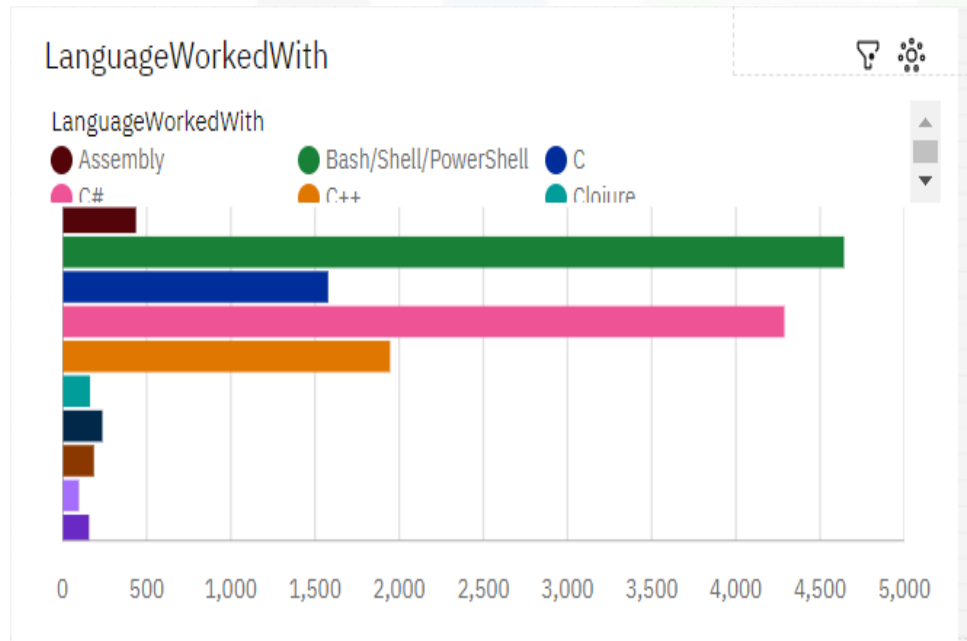
- Data Collection.
- Duplicate Identification.
- Duplicate.
- Missing Value Detection.
- Missing Value Imputation.
- Data Normalization.

RESULTS

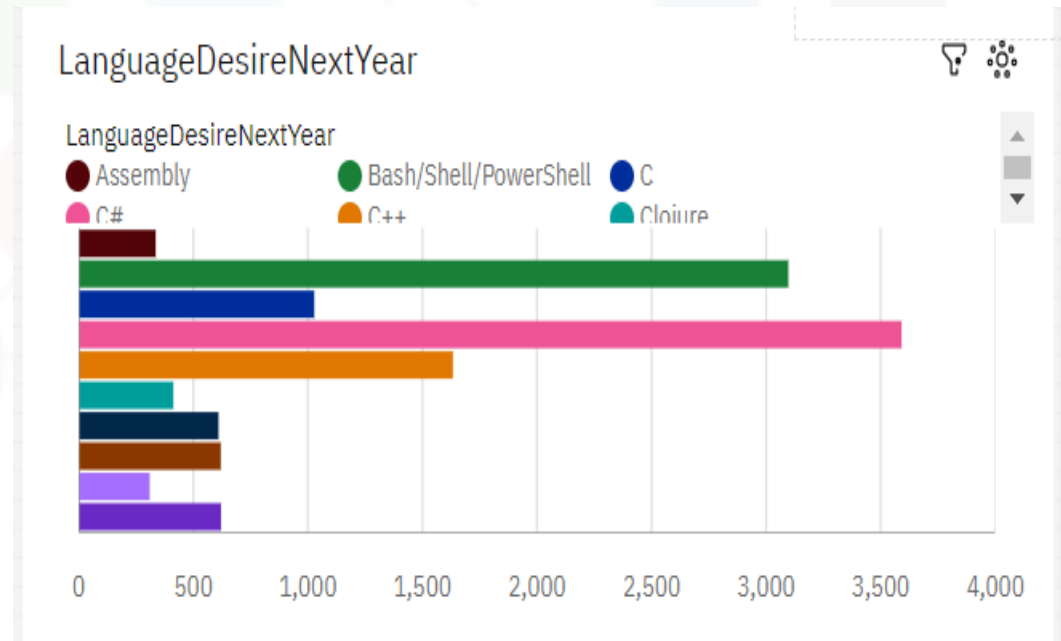


PROGRAMMING LANGUAGE TRENDS

Current Year

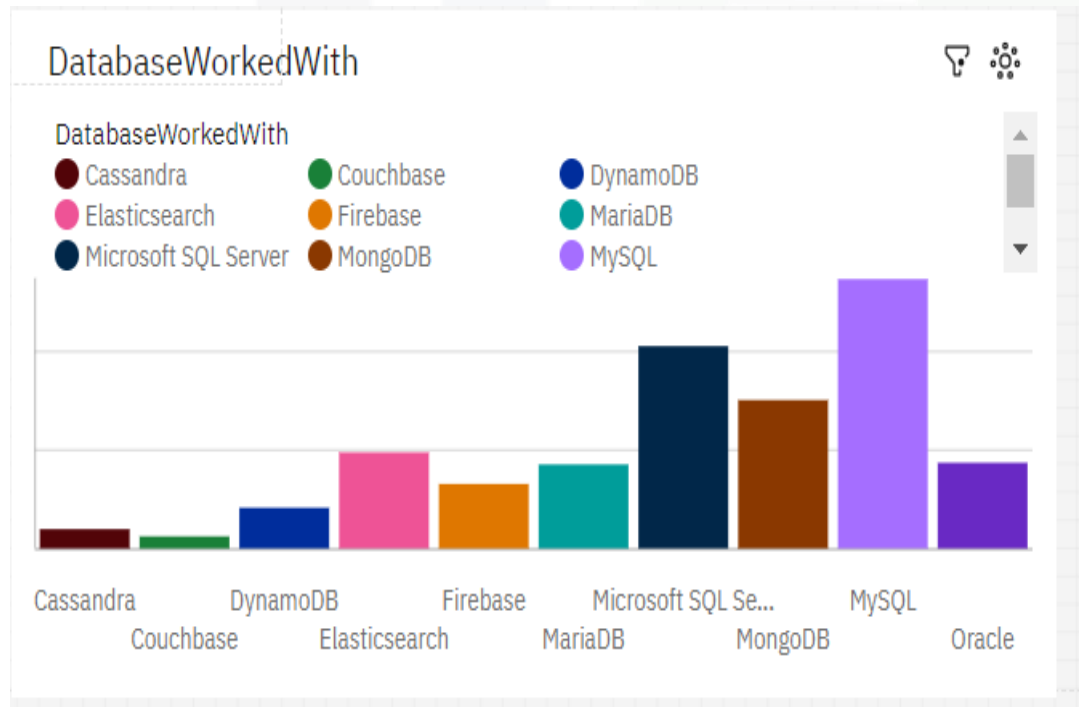


Next Year

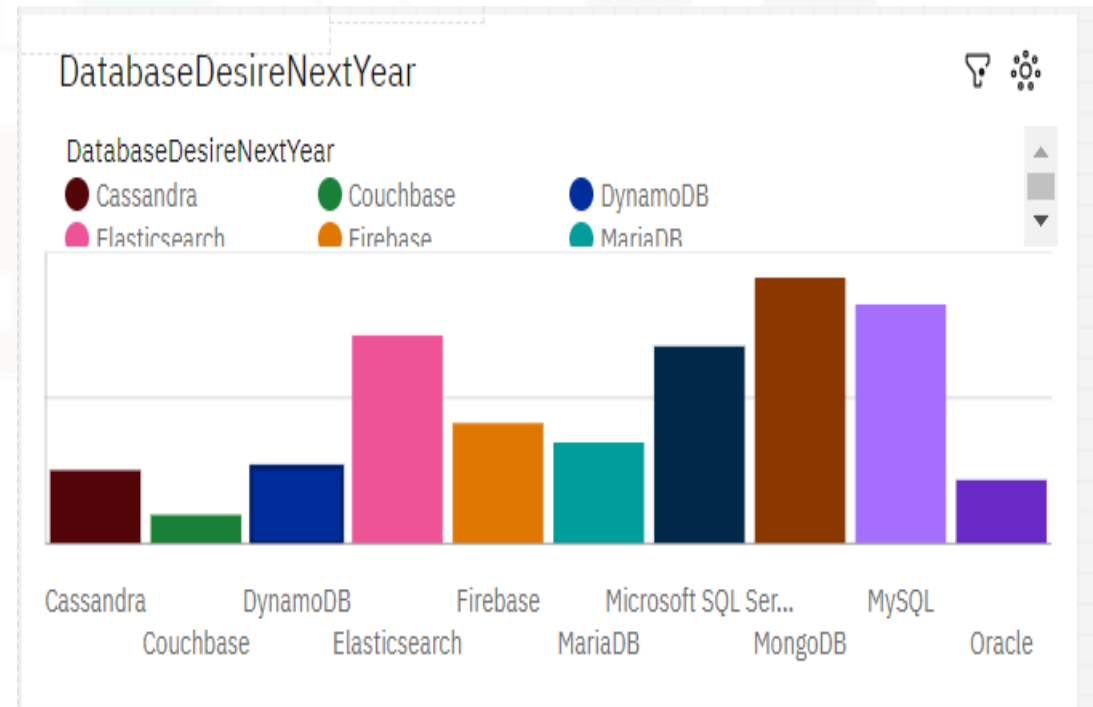


DATABASE TRENDS

Current Year



Next Year



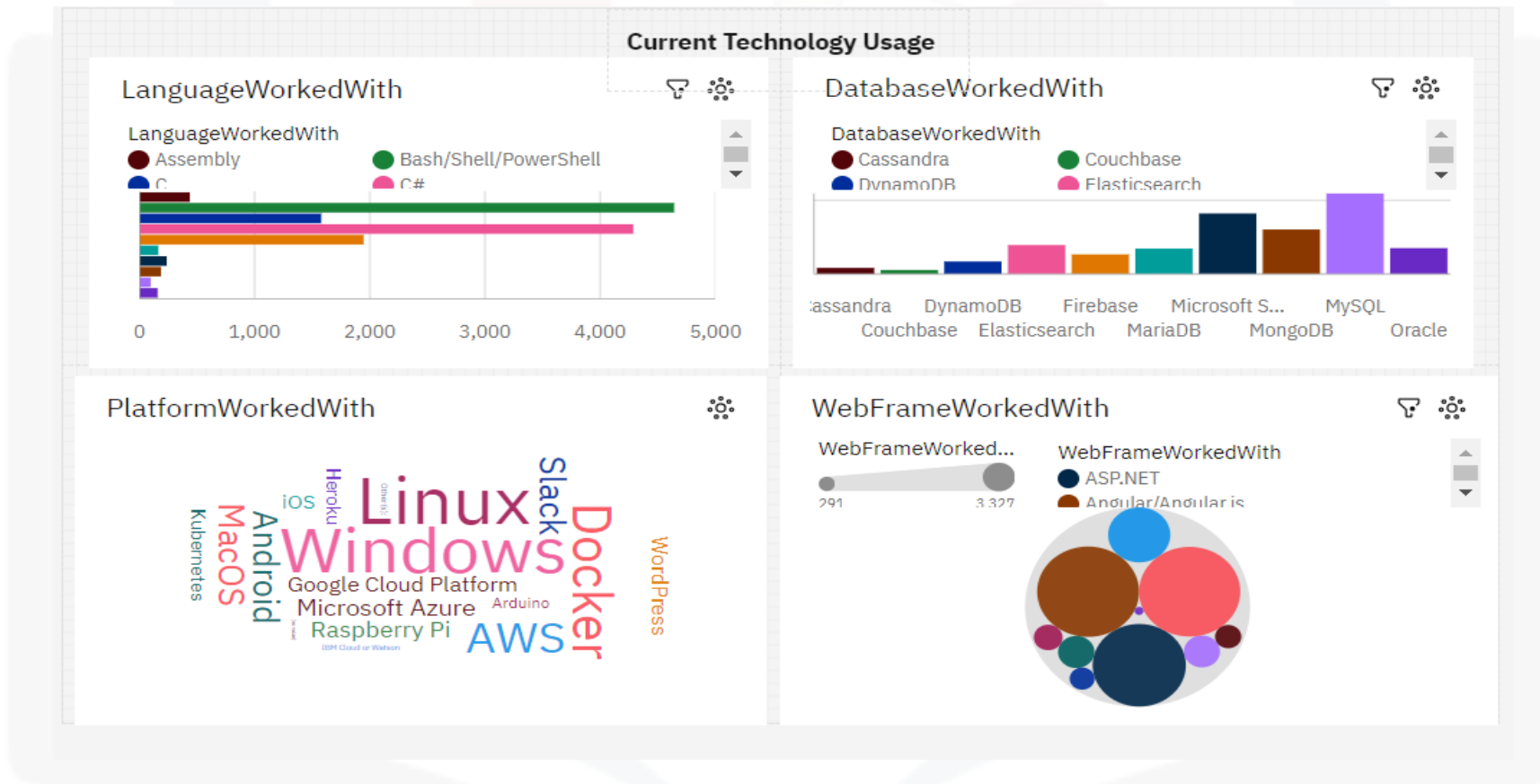
DASHBOARD



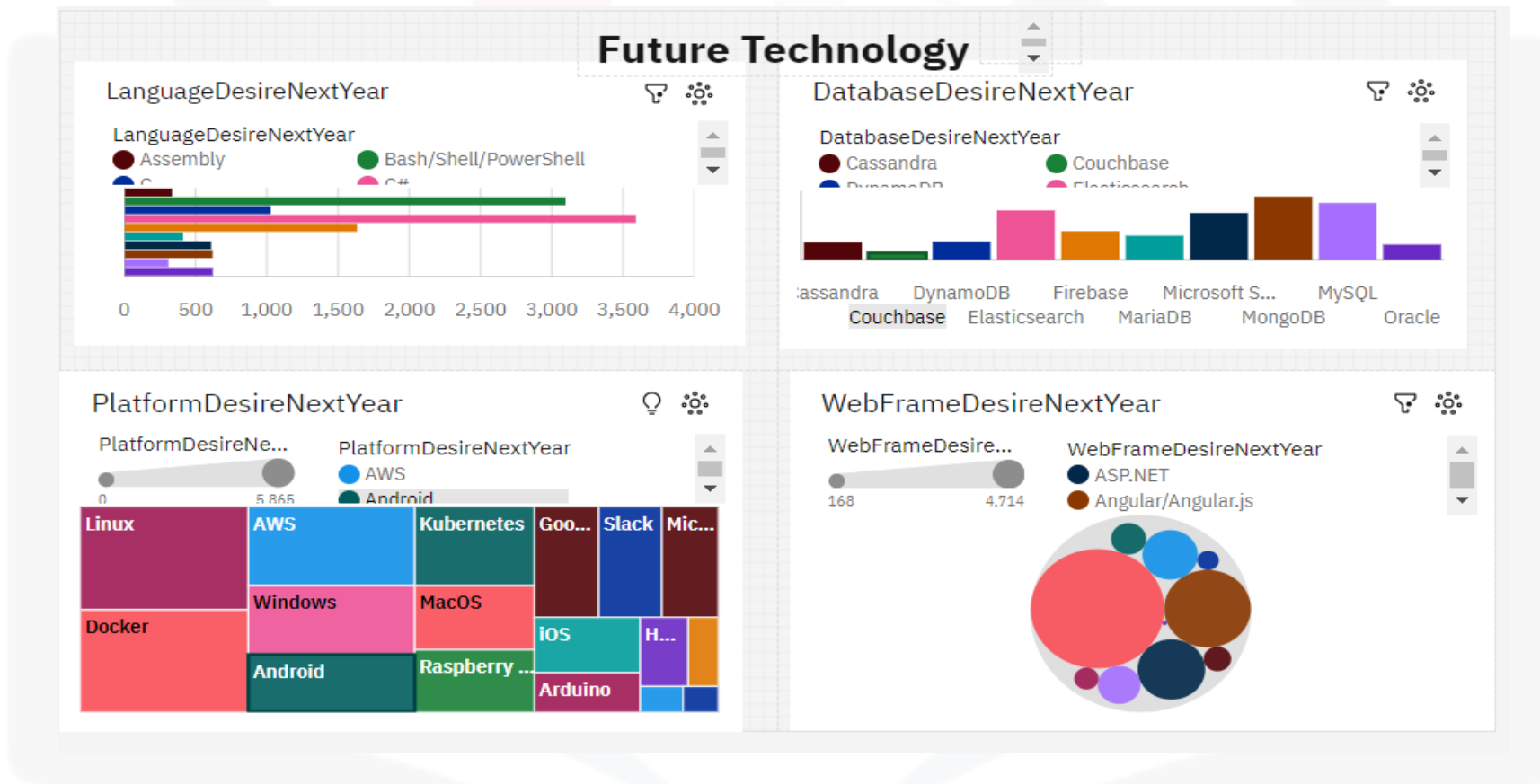
The GitHub link

https://github.com/Odile-nza/IBM_DA_Capstone/blob/main/lab5.pdf

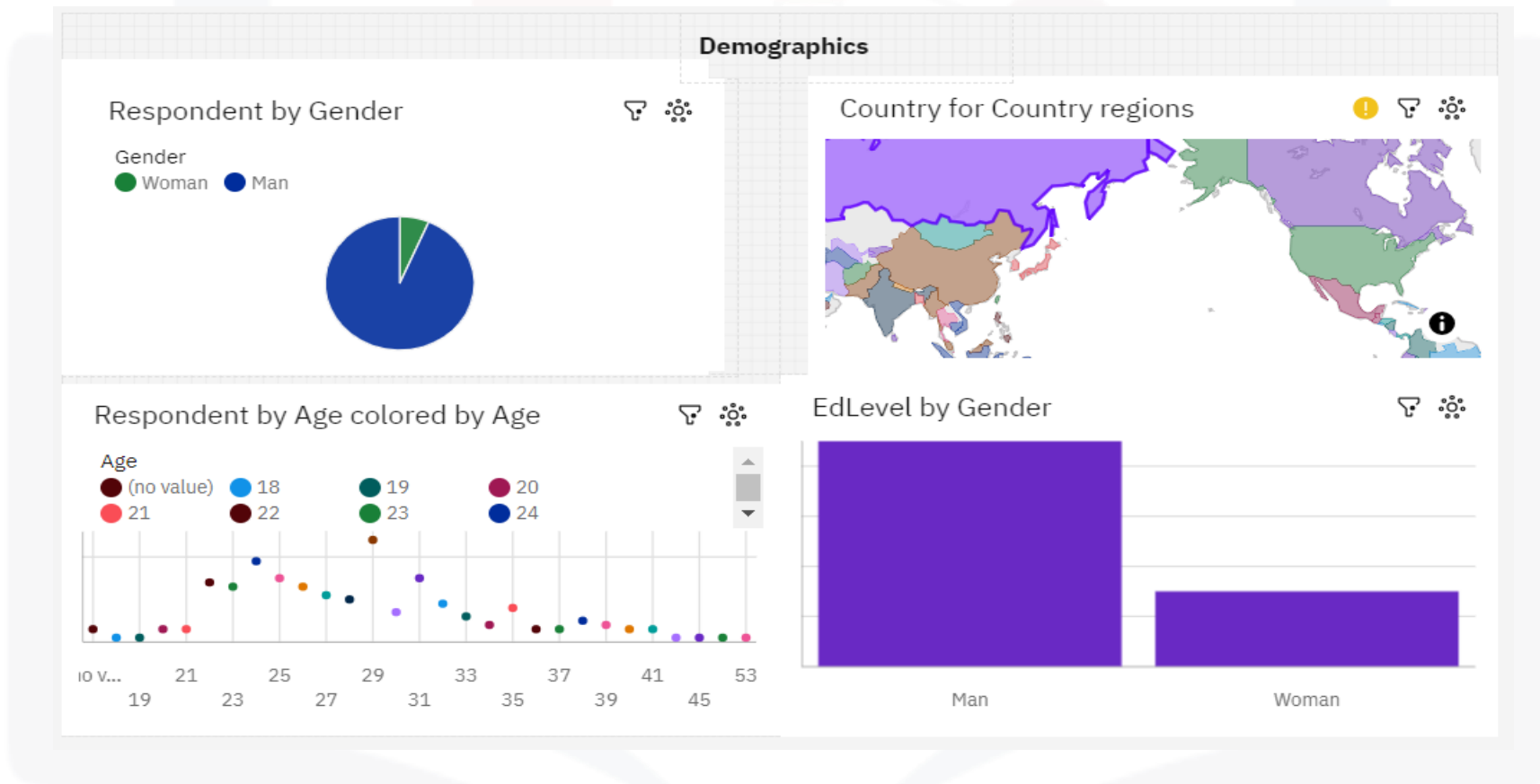
DASHBOARD TAB 1



DASHBOARD TAB 2



DASHBOARD TAB 3



DISCUSSION



OVERALL FINDINGS & IMPLICATIONS

Findings

- Enhanced Data Quality
- Improved Decision Making
- Efficiency Gains

Implications

- Mitigated Risks
- Facilitated Innovation
- Long-Term Impact

CONCLUSION

Effective data management practices, including identifying duplicates, addressing missing values, and normalizing datasets.

Implementing robust data management methodologies not only improves analytical accuracy but also yields benefits

Investing in comprehensive data management strategies is essential for organizations to capitalize on their data assets and maintain competitiveness in the data-driven landscape.

