Capstone Project Summary

Analysis of Data Requirements Timelines and Product Defects in Data Releases

MSc Business Analytics

Muhammad Arbash Malik

June 2024

Contents	
1. Overview of the Project	2
2. Methodology	2
2.1 Data Collection and Cleaning	2
2.2 Analysis Techniques:	
3. Findings and Implications	
3.1 Key Observations:	3
3.2 Implications for Future Releases:	4
4. Learnings and Outcomes	4

<u>1. Overview of the Project</u>

This project explores the relationship between data requirement timelines and the incidence of product defects in data releases at a Financial Institution. The institution conducts four data interface releases annually: two major releases in January and July, focusing on the expansion of data interfaces, and two quarterly releases in April and October, which address central transformations and data collection.

The primary goal is to minimize conditional defects that are unresolved at the time of a rollout but deemed non-critical in future releases. This study seeks to identify contributing factors to these defects and determine correlations to aid in reducing their occurrence.

2. Methodology

2.1 Data Collection and Cleaning:

Data for this analysis was sourced from the institution's release management team and self-queried using the company's project management tool platform. Challenges in data consistency arose due to a migration to a newer version of the platform, requiring significant data cleaning efforts to align datasets from different versions of the platform for coherent analysis.

2.2 Analysis Techniques:

- Feature Engineering: Creation of calculated columns to align data with analysis needs, such as 'Release Date', 'Issue_created_difference', 'Created_difference', and 'Resolved_difference'.
- **Exploratory Analysis:** Distribution checks for key variables such as 'Release Type', 'Issue Type', 'Severity', and 'Priority' to understand data spread and outlier impacts.
- Main Analysis: A comprehensive examination using data slices to observe values across segments and their impact on project defects. Key findings were derived from the relationships among scoped requirement timing, defect detection timing, and the urgency of fixes.

<u>3. Findings and Implications</u>

3.1 Key Observations:

- **Timing and Defects:** A consistent finding across the data was the negative correlation between the timeliness of scoping requirements and the number of defects. Requirements scoped closer to the release date often led to rushed implementations and a higher likelihood of defects.
- Severity and Impact: Most defects in major releases were of high severity, necessitating immediate attention to avoid significant impacts on subsequent data consumption layers.
- **Defect Resolution:** Analysis revealed that defects identified later in the testing cycle had a longer resolution time, emphasizing the importance of early testing and defect identification.

3.2 Implications for Future Releases:

The analysis suggests that improving the scoping timeline and enhancing early testing protocols could significantly reduce the incidence of critical defects. Additionally, a more robust alignment with group and local stakeholders during the data requirement specification phase could further minimize issues during rollouts.

This project underscores the importance of strategic planning and rigorous testing in managing data releases effectively. The insights derived from this analysis are expected to guide future release strategies, aiming for fewer defects and smoother rollouts at the institution.

4. Learnings and Outcomes

During my capstone project, I gained invaluable insights into the operational dynamics of large corporations, particularly in how cross-departmental collaboration is orchestrated within sprint workflows. Through my interactions with various teams at the company, I learned the importance of synchronized efforts across different departments to meet tight deadlines and manage complex data needs effectively. This experience not only enhanced my understanding of corporate structure and project management but also significantly deepened my knowledge of data architecture. I explored various aspects of data management, from collection and cleaning to analysis, which improved my technical skills and provided a practical understanding of how robust data architectures support business objectives and decision-making processes in a large-scale enterprise environment.