Literature Review

# Introduction

The literature review covers the key aspects of two research papers focused on the Data Understanding phase in the Data Science Lifecycle. Below are the summarized points.

# Existing Literature and Theoretical Framework

- Both research papers build on existing literature in Data Science and Business Intelligence.

- Theoretical frameworks used include descriptive statistics, data visualization, and correlation analysis.

- Emphasis on the importance of initial data exploration, pattern identification, and data quality assessment.

# Methodology

- Systematic approach to Data Understanding.

- Data collection from various sources.

- Initial data exploration using descriptive statistics and data visualization.

- Rigorous data quality assessment to ensure relevance and reliability.

- Use of correlation matrices and scatter plots to understand relationships between variables.

# Findings and Results

- Initial data exploration reveals insights into data distribution and trends.

- Identification of missing values and outliers as potential issues.

- Data quality assessment highlights the importance of addressing inconsistencies.

- Thorough initial analysis is crucial for guiding data preparation and modeling.

# Discussion and Conclusion

- The Data Understanding phase is crucial for the success of data-driven projects.

- Systematic data collection, exploration, and quality assessment mitigate potential issues early.

- Proactive approach enhances the accuracy and reliability of the project's outcomes.

- Future research should focus on improving methods for initial data exploration and quality assessment.

# Synthesis and Integration

- The Data Understanding phase is foundational in the Data Science Lifecycle.

- Structured methodologies and rigorous assessments provide a template for future research.

- Highlights the critical role of initial data exploration and quality assessment in achieving reliable results.