[**Note**: This is a template to write a structured report for a data mining project. The template was developed following the [CRISP-DM process framework](https://en.wikipedia.org/wiki/Cross-industry_standard_process_for_data_mining). Complete and remove the sections in [] before submitting your report.]

**[Project Title]**

[Optional: Add an image/logo]

[Your Name/Team]

**Executive Summary:** [¼ page description of the project highlights. The summary needs to answer the following questions: What is the problem you are working on? Why is it important? What are the key results and how are they useful to the reader?]

Table of Contents

[1 Problem Description 3](#__RefHeading___Toc488_3265188426)

[2 Data Collection and Data Quality 3](#__RefHeading___Toc490_3265188426)

[3 Data Exploration 3](#__RefHeading___Toc492_3265188426)

[4 Modeling and Evaluation 4](#__RefHeading___Toc494_3265188426)

[5 Recommendations 4](#__RefHeading___Toc356_1301979095)

[6 Conclusion 4](#__RefHeading___Toc496_3265188426)

[7 List of References 4](#__RefHeading___Toc498_3265188426)

[8 Appendix 4](#__RefHeading___Toc500_3265188426)

[8.1 Student Contributions 5](#__RefHeading___Toc358_1301979095)

# Problem Description

[**Business Understanding:** Frame the Problem. Add a short description of the problem area. Who is your stakeholder and what does the **stakeholder** want? What are the **questions** you need to answer? Why are they important? What kind of data do you need? Do some research and add references to the Reference section.]

# Data Collection and Data Quality

[**Data Understanding part 1**: Describe data source, expected data quality and reliability. If you have several sources, how can you combine the data? At this point, you want to clean the variable names so they are better readable later on.]

# Data Exploration

[**Data Understanding part 2:** Inspect **data quality** and clean data. Present descriptive **statistics** (e.g., in the table shown below) including how much data you have before and after cleaning. Use appropriate visualization (**histograms, bar charts, scatter plots**, etc.) and methods like **correlations**. Discuss what we can learn from exploring the data and what **recommendations** you can give based on your finding.

Do not use screenshots. Use copy&paste to copy tables and charts from your R Notebook rendered as HTML or a Word Document into this document and format them in Word appropriately. Add captions in Word. You will need to adjust the axis label font sizes to make them readable. Labels in figures should be roughly the same size as regular text in the document.]

|  |  |  |
| --- | --- | --- |
| **Variable name** | **Short description** | **Statistics** |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

Table 1: Variable description and Statistics

# Modeling and Evaluation

[**Modeling and Evaluation**: What type of model do we apply to the data?

Describe why you chose the particular model, model assumption and limitations, what variable you use for the model, and how well the model works. ]

# Recommendations

[Deployment: Describe how to interpret the model and what **recommendations** you can make based on the findings. How would the stakeholder use the findings and why is the recommendation useful to the stakeholder.]

# Conclusion

[Does the project answer the initial questions? Repeat the key findings and why they are important.]

# List of References

[Cite sources of information in your document and but complete references here. You may use any citation style as long as you are consistent. You can find the basics about how to properly cite references using APA style [here](https://owl.purdue.edu/owl/research_and_citation/apa_style/apa_formatting_and_style_guide/in_text_citations_the_basics.html). ]

# Appendix

[Put code and long tables that you do not need inside your report here. ]

## Student Contributions

Add a list with who contributed to what part of this report.