[**Note**: This is a template to write a structured report for a data science project. The template was developed by for the course [DS 1300](https://michael.hahsler.net/SMU/DS1300) and follows the [Data Science Process](https://www.kdnuggets.com/2016/03/data-science-process.html). Complete and remove the sections in [] before submitting your report.

This report is the result of step: Communicate the results in the Data Science Process. We will use it to document each step in the process.]

**[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 2](#_Toc57109743)

[2. Data Collection and Data Quality 2](#_Toc57109744)

[3. Data Exploration 2](#_Toc57109745)

[4. Modeling 2](#_Toc57109746)

[5. Conclusion 3](#_Toc57109747)

[6. List of References 3](#_Toc57109748)

[7. Appendix 3](#_Toc57109749)

# Problem Description

[Step: Frame the Problem

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

[Describe data source, expected data quality and reliability. If you have several sources, how can you combine the data?]

# Data Exploration

[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 and pivot tables**. 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 into this document and format them in Word appropriately. Add captions in Word. Charts in Rapidminer can be exported as png (see top right corner of the chart in Rapidminer) and inserted as a picture 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 : Variable description and Statistics

# Modeling

[What type of model can we apply to the data? Example model types are:

* **Trend lines**: Show the fitted trend line. Discuss why you use the selected type (e.g., linear or exponential). Can you use the model for prediction/forecasting?
* **Clustering**: Look at the cluster centers and some other visualization (e.g., a scatterplot with colors indicating the cluster). How do the clusters group the data? How can you use the grouping for recommendations?
* **Decision trees**: Show the tree and discuss how it splits the data. How can you use the model and predictions made by the model for recommendations?

Describe why you chose the particular model, model assumption and limitations, what variable you use for the model, and how well the model works. Then describe how to interpret the model and what **recommendations** you can make based on the findings]

# 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. ]