Data Preprocessing Report

Prepared by

[Team Name/Team Members]

Date: 25/11/24

Table of Contents

[1. Introduction 3](#_Toc182386182)

[1.1. Background 3](#_Toc182386183)

[1.2. Objective 3](#_Toc182386184)

[2. Data Loading 3](#_Toc182386185)

[2.1. Purpose 3](#_Toc182386186)

[2.2. Justification 3](#_Toc182386187)

[3. Data Validation 3](#_Toc182386188)

[3.1. Purpose 3](#_Toc182386189)

[3.2. Justification 3](#_Toc182386190)

[4. Handling Missing Values with Predictive Modeling 3](#_Toc182386191)

[4.1. Purpose 3](#_Toc182386192)

[4.2. Justification 3](#_Toc182386193)

[5. Handling Inconsistencies and Duplicates 3](#_Toc182386194)

[5.1. Purpose 3](#_Toc182386195)

[5.2. Justification 3](#_Toc182386196)

[6. Handling Outliers 3](#_Toc182386197)

[6.1. Purpose 3](#_Toc182386198)

[6.2. Justification 3](#_Toc182386199)

[7. Encoding Categorical Variables 4](#_Toc182386200)

[7.1. Purpose 4](#_Toc182386201)

[7.2. Justification 4](#_Toc182386202)

[8. Feature Scaling 4](#_Toc182386203)

[8.1. Purpose 4](#_Toc182386204)

[8.2. Justification 4](#_Toc182386205)

[9. Feature Engineering with Interaction Metrics 4](#_Toc182386206)

[9.1. Purpose 4](#_Toc182386207)

[9.2. Justification 4](#_Toc182386208)

[10. Data Anonymization 4](#_Toc182386209)

[10.1. Purpose 4](#_Toc182386210)

[10.2. Justification 4](#_Toc182386211)

[11. Saving the Cleaned Dataset 4](#_Toc182386212)

[11.1. Purpose 4](#_Toc182386213)

[11.2. Justification 4](#_Toc182386214)

[12. Conclusion 4](#_Toc182386215)

[12.1. Summary of Steps 4](#_Toc182386216)

[12.2. Impact on Data Quality 4](#_Toc182386217)

[12.3. Next Steps 4](#_Toc182386218)

[References 5](#_Toc182386219)

[Appendices 5](#_Toc182386220)

[A. Code Snippets 5](#_Toc182386221)

[B. Additional Tables or Figures 5](#_Toc182386222)

## 2.2. Implementation

# 1. **Introduction**

## 1.1. Background

## 1.2. Objective

# **2. Data Loading**

## 2.1. Purpose

## 2.2. Implementation

## 2.3. Justification

# **3. Data Validation**

## 3.1. Purpose

## 3.2. Implementation

## 3.3. Justification

# **4. Handling Missing Values with Predictive Modeling**

## 4.1. Purpose

## 4.2. Implementation

## 4.3. Justification

# **5. Handling Inconsistencies and Duplicates**

## 5.1. Purpose

## 5.2. Implementation

## 5.3. Justification

# **6. Handling Outliers**

## 6.1. Purpose

## 6.2. Implementation

## 6.3. Justification

# **7. Encoding Categorical Variables**

## 7.1. Purpose

## 7.2. Implementation

## 7.3. Justification

# 8. Feature Scaling

## 8.1. Purpose

## 8.2. Implementation

## 8.3. Justification

# 9. Feature Engineering with Interaction Metrics

## 9.1. Purpose

## 9.2. Implementation

## 9.3. Justification

# 10. Data Anonymization

## 10.1. Purpose

## 10.2. Implementation

## 10.3. Justification

# 11. Saving the Cleaned Dataset

## 11.1. Purpose

## 11.2. Implementation

## 11.3. Justification

# 12. Conclusion

## 12.1. Summary of Steps

## 12.2. Impact on Data Quality

## 12.3. Next Steps

# References

# Appendices

## A. Code Snippets

## B. Additional Tables or Figures