Data Analysis and Preprocessing Documentation

1. Introduction

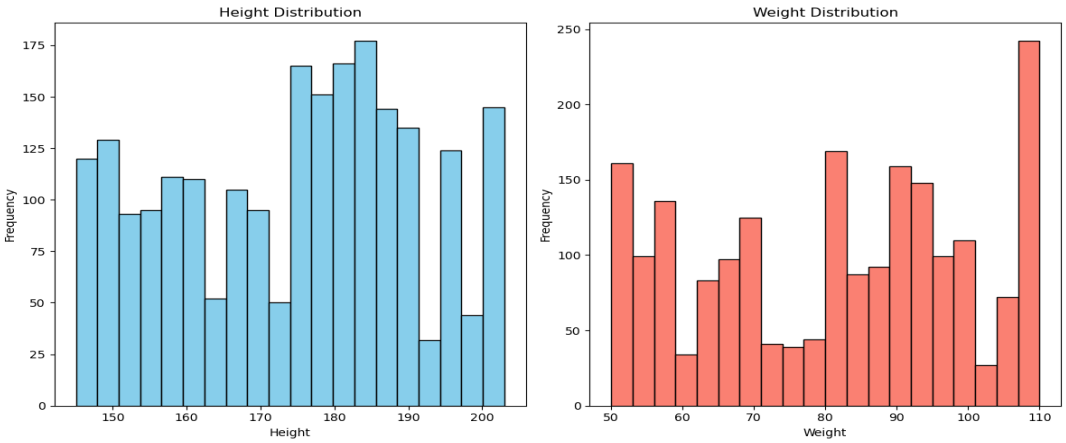
This document summarizes the findings from the Exploratory Data Analysis (EDA) and outlines the steps taken during the data preprocessing phase. The goal is to prepare the dataset for further analysis and modeling.

2. Data Loading and Exploration

- **Description**: The dataset was loaded into the analysis environment.
- **Findings**: An initial examination revealed various attributes, including user demographics, medication details, and reported side effects.

3. Data Visualization

- **Height and Weight Distribution**: Visualizations were created to analyze the distributions of height and weight among the participants.
- **Gender Distribution**: The gender distribution was also visualized to understand the sample composition.

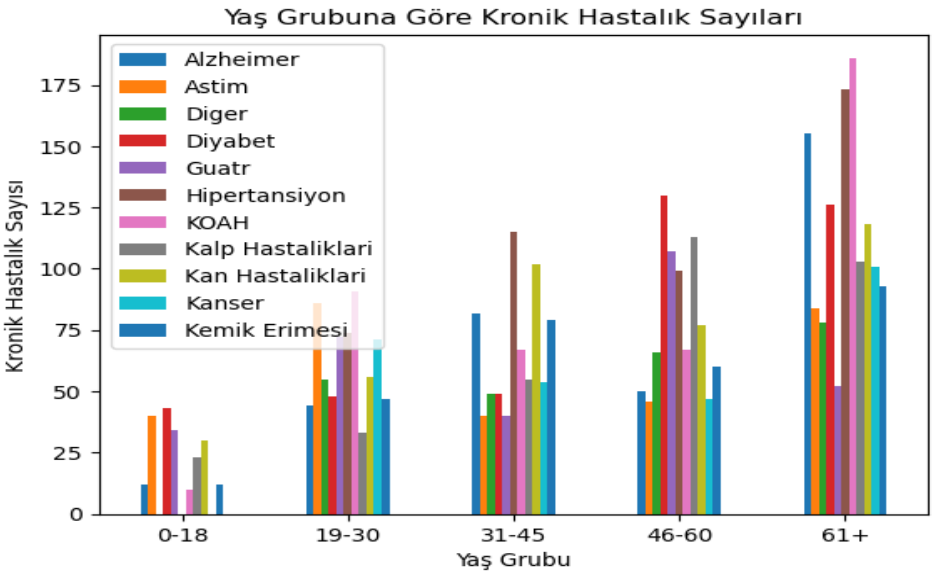


4. Allergy Frequencies

- **Description**: A frequency analysis was performed on reported allergies.
- **Findings**: Common allergies were identified, providing insight into the health concerns of the participants.

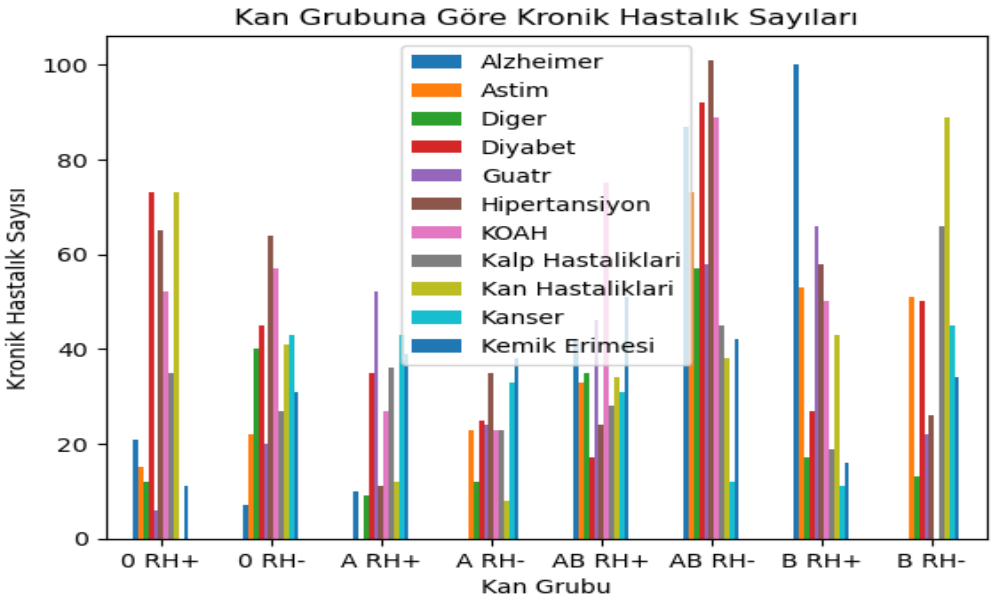
5. Chronic Diseases by Age Group

- **Description**: The relationship between age groups and chronic diseases was examined.
- **Findings**: Patterns in chronic disease prevalence were observed across different age demographics.



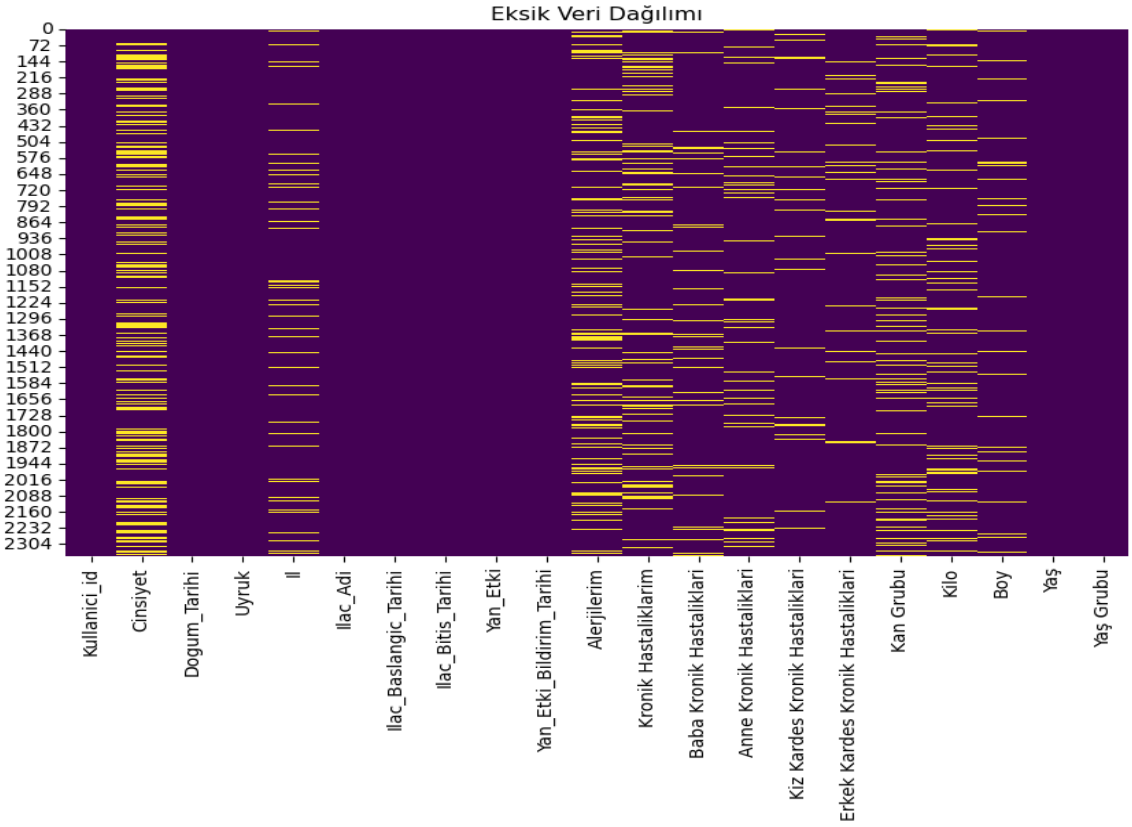
6. Chronic Diseases by Blood Group

- **Description**: An analysis of chronic diseases based on blood groups was conducted.
- **Findings**: This analysis provided insights into how chronic conditions vary by blood type.



7. Missing Data Analysis

- **Description**: An examination of missing data was conducted using a heatmap.
- **Findings**: Patterns of missingness were identified, highlighting variables with significant gaps.



8. Missing Data Imputation

- **Description**: Steps were taken to fill in missing values.
- **Method**: Imputation strategies were applied based on the nature of the data (e.g., mean for numerical data, most frequent for categorical data).

9. Encoding

- **Description**: Categorical variables were encoded to prepare the dataset for modeling.
- **Method**: One-hot encoding was applied to categorical features, ensuring they can be used effectively in machine learning algorithms.

10. Conclusion

The data preprocessing steps have prepared the dataset for subsequent modeling and analysis. This includes handling missing values and converting categorical variables into a suitable format for machine learning.