Heart Disease Analysis

hGROUP 13

INTRODUCTION

this project aims to analyze a dataset containing information about heart disease, smoking habits, and biking activities. The primary objective is to explore the relationship between smoking, biking, and the occurrence of heart disease. The insights gained from this analysis can have significant implications for public health and well-being, providing evidence-based conclusions to support decision-making processes.

DATA UNDERSTANDING

The first step of the analysis involves understanding the dataset. The code will load the data, display its dimensions, and provide summary statistics to gain insights into the data structure. Additionally, data types of the columns will be displayed to identify any potential data quality issues.

DATA PREPARATION

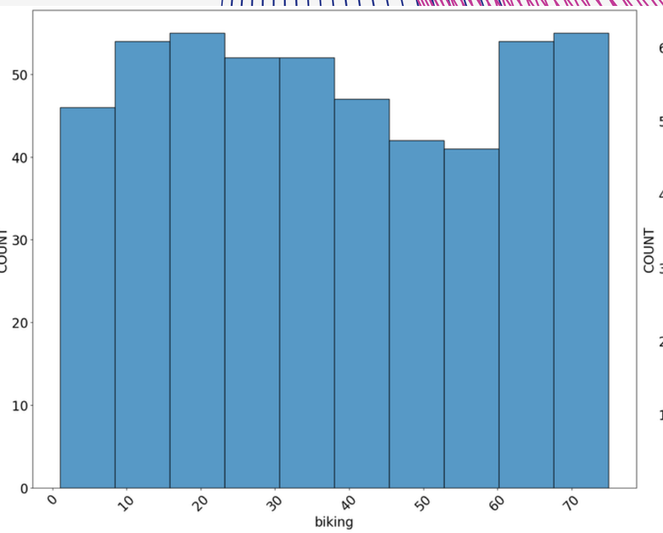
Data preparation is crucial for a clean and reliable analysis. This step involves:

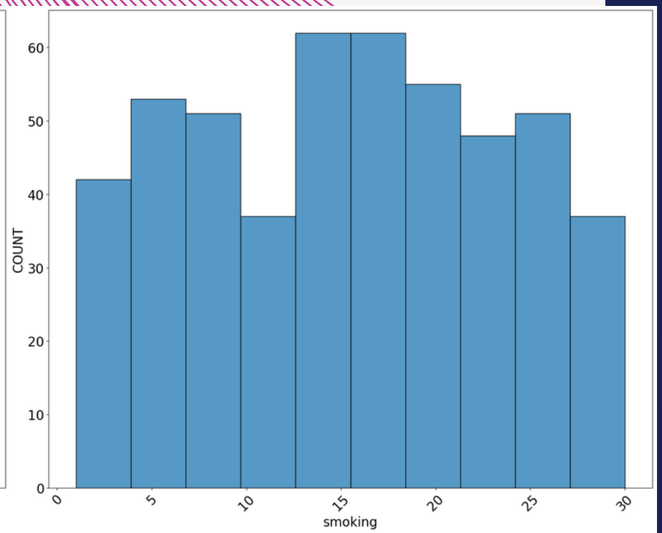
1. Renaming the column heart. disease to heart_disease for consistency.
2. Rounding off the values in the dataset to the nearest integer to remove decimal points.

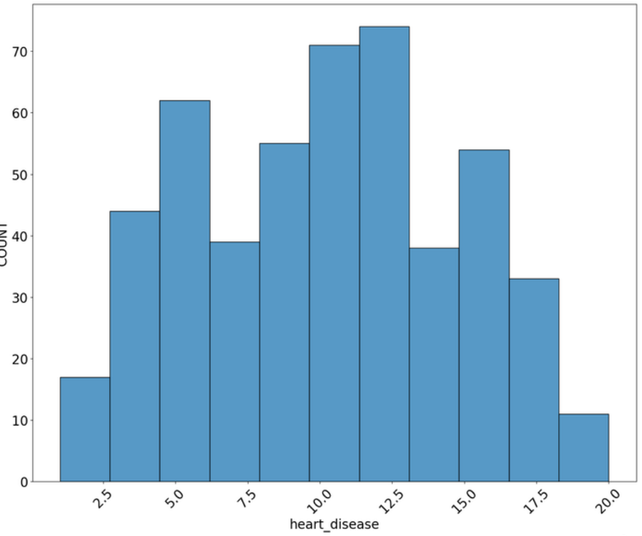
DATA ANALYSIS

Data preparation is crucial for a clean and reliable analysis. This step involves:

1. Renaming the column heart. disease to heart_disease for consistency.
2. Rounding off the values in the dataset to the nearest integer to remove decimal points.







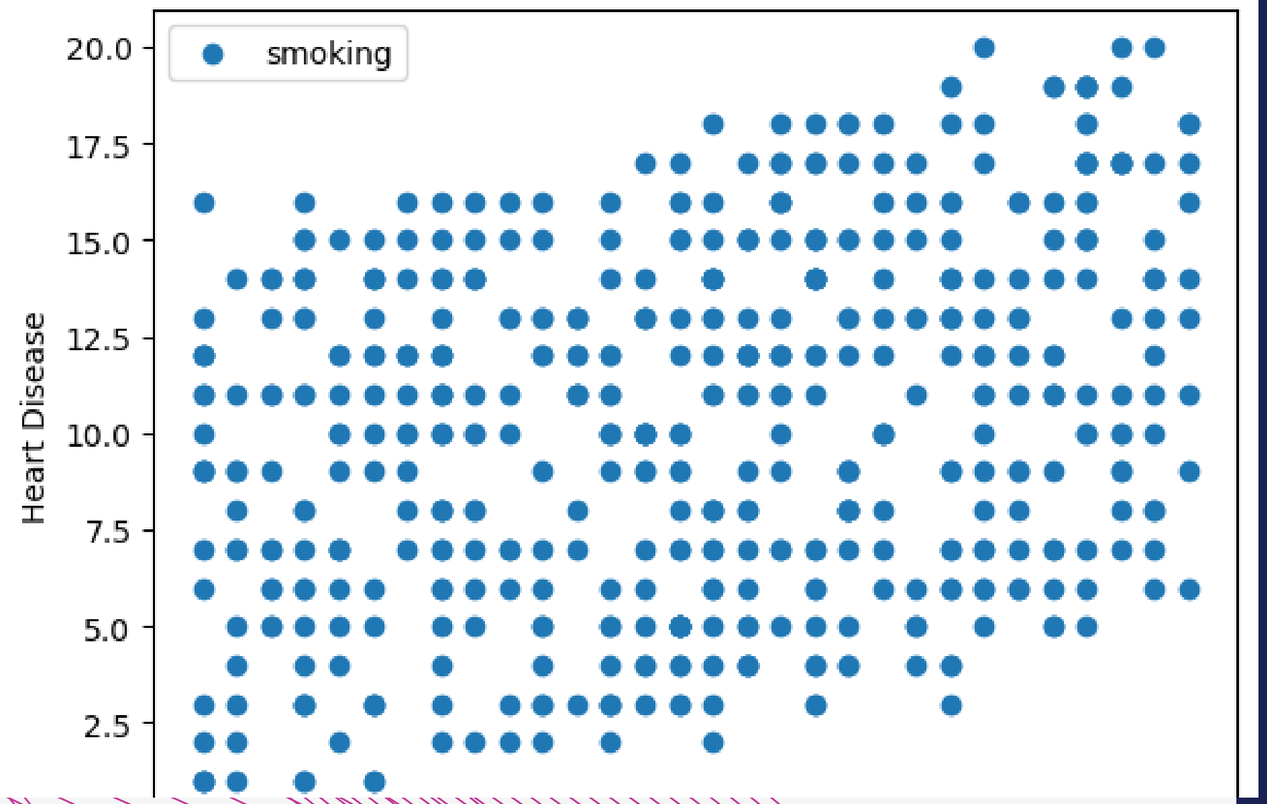
DATA ANALYSIS

1.)UNIVARIATE ANALYSIS

It involves examining individual variables in isolation to understand their distribution and summary statistics

The smoking variable does not seem to have a linear relationship with the heart disease variable.

Biking variable seems to have a negative linear relationship with the heart disease variable.

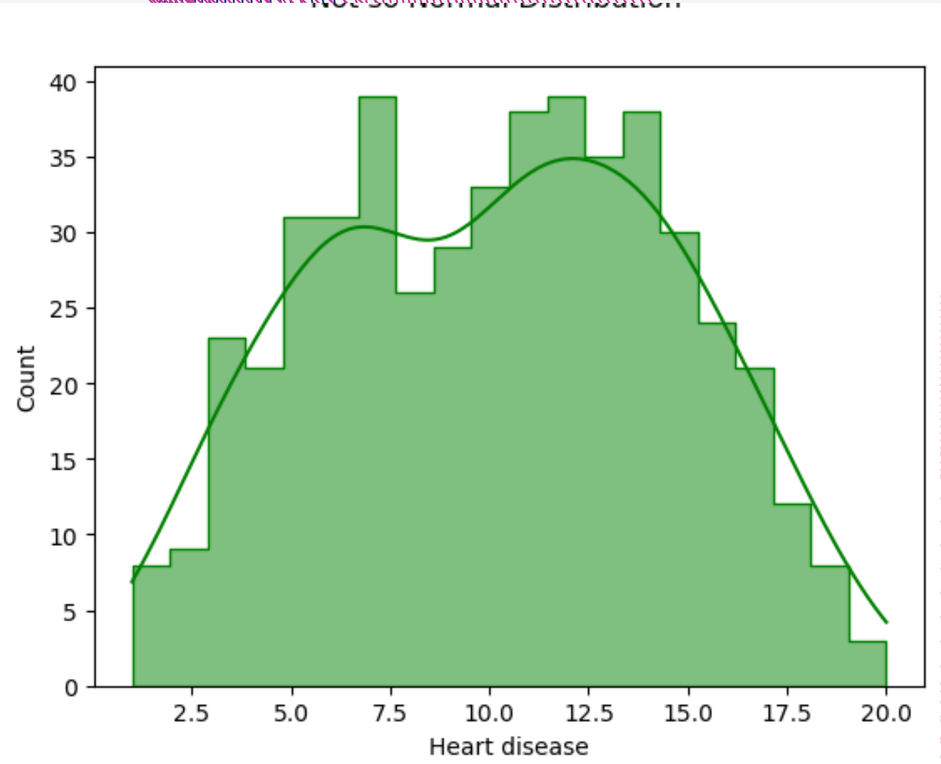


DATA ANALYSIS

2.)BIVARIATE ANALYSIS

Explores the relationships between two variables in the dataset. It helps to identify how changes in one variable are associated with changes in another

The target variable heart disease seems to have a somewhat normal distribution.



FINDINGS

The target variable heart disease seems to have a somewhat normal distribution.

CONCLUSION

In conclusion, this code provides a comprehensive analysis of the heart disease dataset, including data preparation, exploratory data analysis, and building a linear regression model. The insights gained from this analysis can help better understand the relationship between smoking, biking, and heart disease, providing valuable information for further research or decision-making processes.