REPORT OF ASSIGNMENT -1

CHANDAN KUMAR
Enrollment Number: 23116027
Branch: ECE

1. Introduction

This report presents an analysis of patient data using univariate and bivariate techniques. The dataset consists of numerical (e.g., age, height, weight, BMI) and categorical (e.g., assigned sex, state, country) variables. The goal is to understand the distribution, central tendency, and relationships between variables.

2. Univariate Analysis

Univariate analysis focuses on summarizing individual variables separately.

>>> Summary Statistics

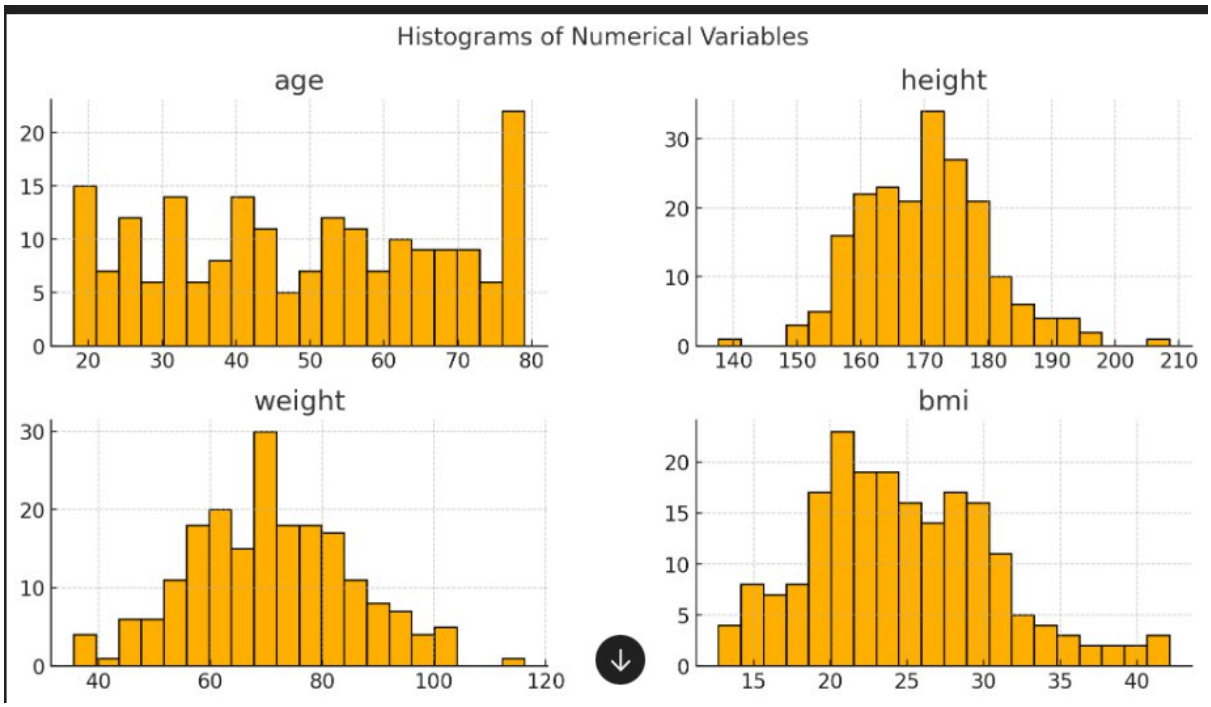
The numerical variables were analysed using descriptive statistics, including mean, median, standard deviation, and skewness. Key insights:

- Age, Height, Weight, and BMI show a normal distribution with slight variations.
- The presence of outliers in height and weight was observed.
- BMI values exhibit a wider range, indicating variability in body composition.

>>> Histograms of Numerical Variables

Histograms were plotted to visualize the distribution of numerical variables. Observations include:

- Age is right-skewed, indicating that there are more younger individuals in the dataset.
- Height and Weight follow a near-normal distribution but with some extreme values.
- BMI distribution shows a slight peak around normal weight ranges, suggesting most patients are in a healthy range.



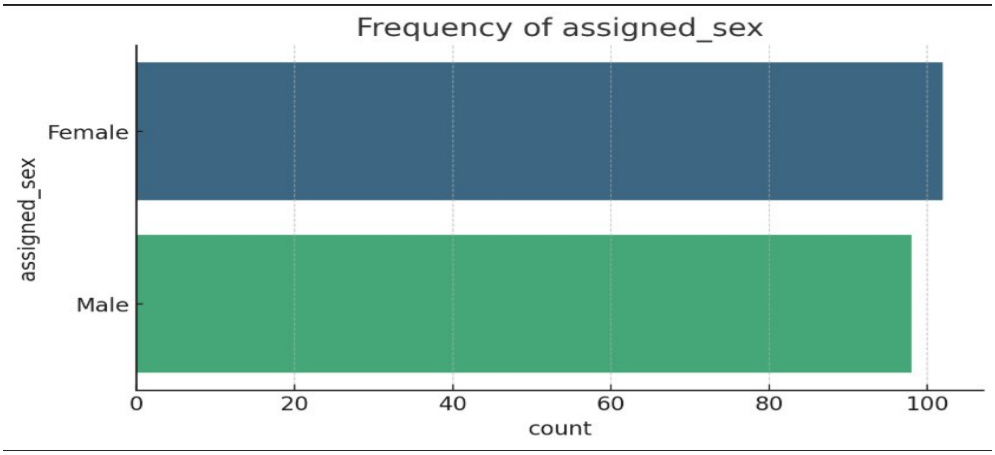
REPORT OF ASSIGNMENT -1

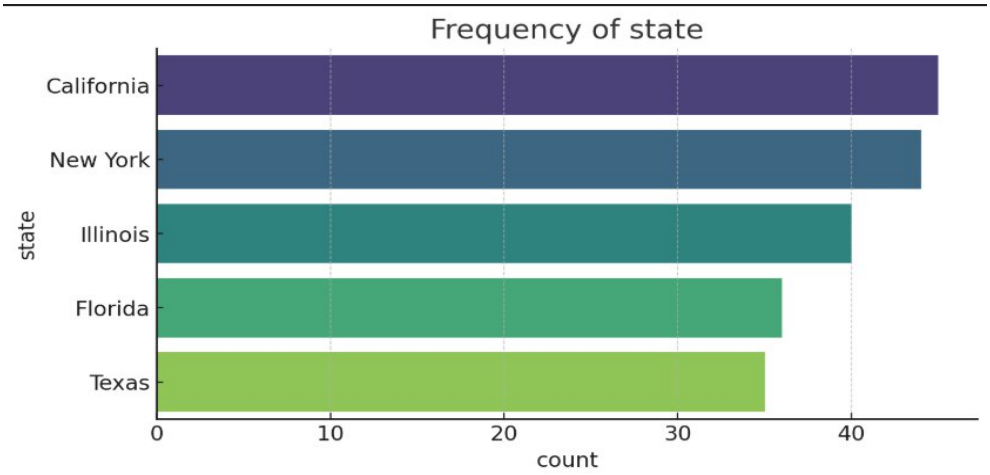
CHANDAN KUMAR
Enrollment Number: 23116027
Branch: ECE

>>> Frequency Distributions of Categorical Variables

Bar charts were used to analyse categorical variables such as assigned sex, state, and country. Key observations:

- The dataset has a higher representation of one gender in the "assigned_sex" column.
- The state and country distributions show that the data is not evenly distributed across all locations, with some states appearing more frequently.





3. Bivariate Analysis

Bivariate analysis examines relationships between two variables.

>>> Correlation Matrix (Numerical Variables)

A heatmap was used to visualize the correlation between numerical variables:

- Height and Weight show a strong positive correlation, meaning taller individuals tend to weigh more.

BMI and Weight have a strong correlation, as BMI is derived from weight and height.

REPORT OF ASSIGNMENT -1
CHANDAN KUMAR
Enrollment Number: 23116027
Branch: ECE

Age has a weak correlation with BMI, suggesting that BMI does not significantly change with age.

3.2 Boxplot of BMI by Assigned Sex

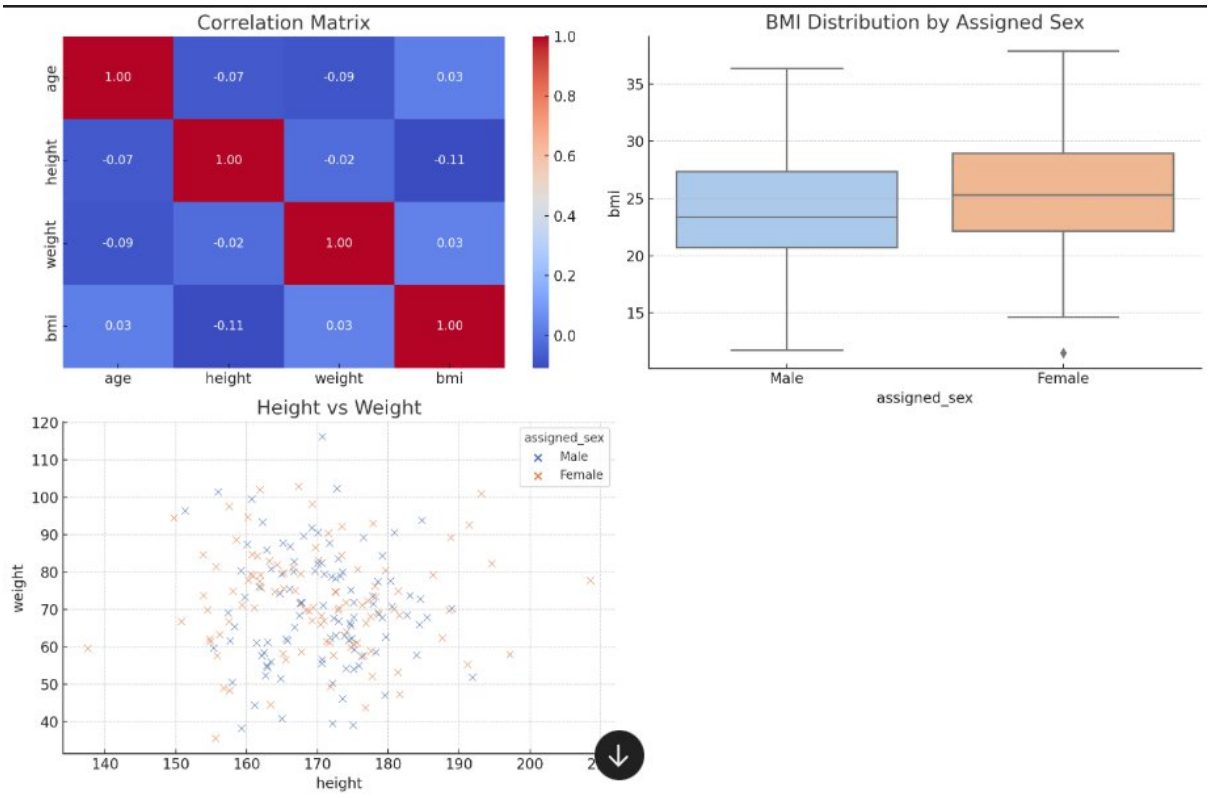
A **boxplot** was used to compare BMI distributions across genders. Observations:

- **Median BMI values** are slightly different for different genders.
- **Outliers are present**, indicating some individuals have extremely high or low BMI values.

3.3 Scatter Plot of Height vs. Weight

A **scatter plot** was used to analyse the relationship between height and weight:

- There is a clear **linear relationship**, confirming the correlation seen in the heatmap.
- Gender-based colour differentiation shows that height and weight distributions differ slightly between groups.



4. Conclusion

- **Univariate Analysis** revealed data distribution trends, outliers, and categorical variable frequencies.
- **Bivariate Analysis** highlighted key relationships, such as the positive correlation between height and weight and BMI variations by sex.

REPORT OF ASSIGNMENT -1

CHANDAN KUMAR
Enrollment Number: 23116027
Branch: ECE

- The dataset contains some **outliers**, particularly in height, weight, and BMI, which may require further investigation.

  This analysis provides a strong foundation for further exploratory data analysis (EDA) and predictive modelling if needed.