DATA PRE-PROCESSING

DATA VISUALIZATION

Date	19 November 2022
Team ID	PNT2022TMID04221
Project Name	Project – Statistical Machine Learning Approaches to Liver Disease Prediction

- Data visualization is where a given data set is presented in a graphical format. It helps the detection of patterns, trends and correlations that might go undetected in text-based data.
- Understanding your data and the relationship present within it is just as important as any algorithm used to train your machine learning model. In fact, even the most sophisticated machine learning models will perform poorly on data that wasn't visualized and understood properly.
- To visualize the dataset we need libraries called Matplotlib and Seaborn.
- The Matplotlib library is a Python 2D plotting library which allows you to generate plots, scatter plots, histograms, bar charts etc.

Univariate Analysis

Univariate analysis is the simplest form of data analysis where the data being analyzed contains only one variable.

Bivariate Analysis

It involves the analysis of two variables (often denoted as X, Y), for the purpose of determining the empirical relationship between them.

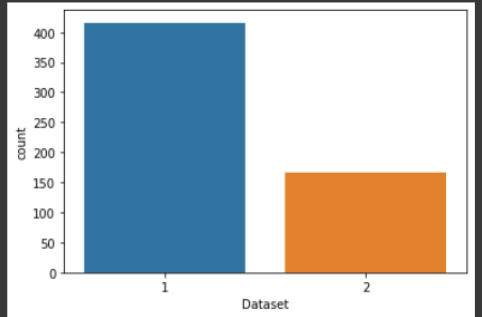
Let's visualize our data using Matplotlib and searborn library.

1. Bar plot between Gender and Count

```
sns.countplot(data=df, x = 'Dataset', label='Count')

LD, NLD = df['Dataset'].value_counts()
print('Number of patients diagnosed with liver disease: ',LD)
print('Number of patients not diagnosed with liver disease: ',NLD)
```

```
Number of patients diagnosed with liver disease:  416
Number of patients not diagnosed with liver disease:  167
```



We can observe from above bar plot is, count of liver disease is observed in male is higher compare to females.

2. Bar Plot which describe about the total number of disease.

```
sns.countplot(data=df, x = 'Gender', label='Count')

M, F = df['Gender'].value_counts()
print('Number of patients that are male: ',M)
print('Number of patients that are female: ',F)
```

```
Number of patients that are male:  441
Number of patients that are female:  142
```