Exploratory Data Analysis (EDA)

The process of analyzing the data, discovering the patterns, spotting anomalies, testing hypotheses, and checking the assumption. We use summary statistics and graphical representations for the EDA.

Project Objective

Study the data set of white wine quality. The objective of the study is to conduct EDA on several parameters of the data set.

Exploratory Data Analysis

1. Understand the data using the pandas library
    1. Print the first five rows of the dataset
    2. Print the last five rows of the dataset
    3. Find out the total number of rows and columns of the data set
    4. Find out the columns, data types, and presence of null values or missing values in the data set
2. Univariate Analysis
    1. Find out missing values graphically
    2. Find out the number of rows, mean, std deviation, min, Q1, Q2, Q3, max values for each variable. Document your observations to check if the outliers are present in the data set
    3. Draw the histogram, kernel density estimate (kde) to check the distribution and skewness for each variable. Document your observation
    4. Create a frequency distribution table and bar chart for the output variable
    5. Draw box plot for each variable and identify the IQR and outliers
3. Multivariate Analysis
    1. Draw the correlation matrix and identify the variables that are correlated to each other.