**Exploratory data analysis Of Wine Quality Check:-**

* Importing of CSV file.
* One of the common tasks in data analysis is to check for duplicates and drop them if necessary. Duplicates are rows that have identical values in all or some of the columns. They can cause errors or bias in the results and should be removed before further processing. Checking of null values.
* Use `df.info()` to check for NAN values and dtypes in a dataframe.
* To see how many missing values and what data types are in each column, run `df.info()`
* df.info() is a handy method to get an overview of NAN values and dtypes in your dataFilling of NAN’s in all columns with mode.
* Use the unique() function to extract the unique values of a column.
* The project aims to develop a nonvisual interface for accessing and manipulating data.
* Statistical analysis of the project, which includes the minimum, maximum, standard deviation, mean, and median values of the data.
* The visual analysis of the project is a crucial step to evaluate its progress and outcomes. It involves creating and interpreting graphs, charts, diagrams, and other visual representations of the data collected and processed during the project. The visual analysis can help identify trends, patterns, outliers, correlations, and causal relationships among the variables.
* Outlier treatment is the process of identifying and handling the extreme values in a data set that deviate significantly from the rest of the observations.
* Outliers can affect the accuracy and validity of statistical analysis, as they can skew the distribution, inflate the variance, and bias the estimates of parameters.
* There are different methods for outlier treatment, such as:
* Deleting or dropping the outliers from the data set, if they are due to errors or anomalies.
* Replacing or imputing the outliers with mean, median, mode, or a value based on domain knowledge or business logic.
* To check for outliers by box plot, follow these steps:
* Draw a box plot using the five-number summary of the data: minimum, first quartile, median, third quartile, and maximum.
* The boxplot() function takes an array of data as an argument and returns a graphical representation of the distribution of the data.
* Identify any data points that lie beyond the whiskers of the box plot. These are potential outliers.
* Apply a criterion to determine if the potential outliers are true outliers. One common criterion is the 1.5\*IQR rule, which states that a data point is an outlier if it is more than 1.5 times the interquartile range (IQR) away from either quartile.
* Report the outliers and their values, or remove them from the data set if appropriate.
* Transforming or scaling the data to reduce the impact of outliers, such as using IQR, log, square root, or z-score transformations.
* By plotting a boxplot, we can recheck the outliers and see if they are valid observations or errors in the data collection or processing.