**Statistics and Trends**

Name:-

Student Number:-

Github Link:-

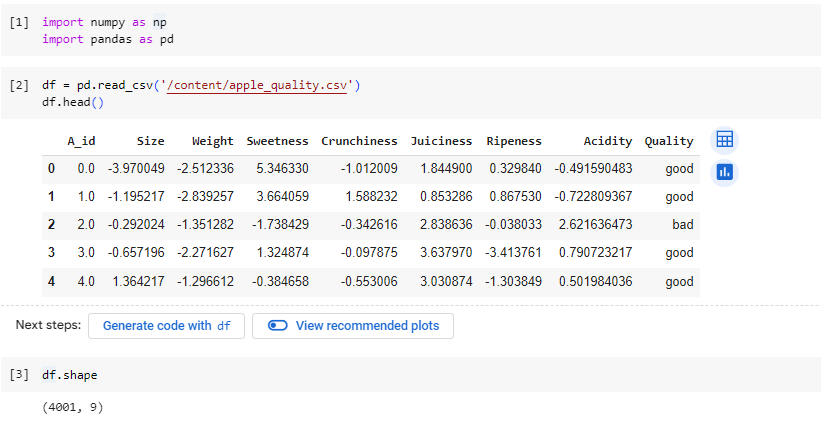
**Introduction**:-

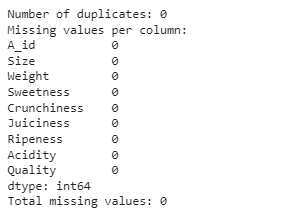
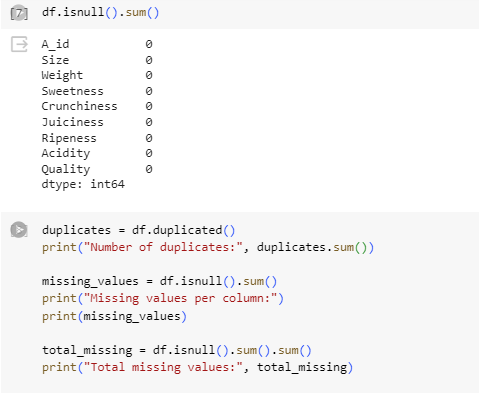
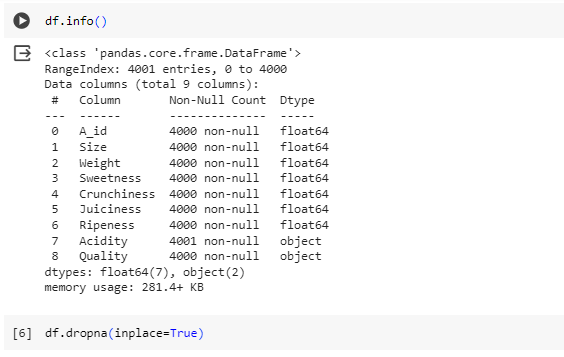
The analysis aims to explore and understand the patterns within the provided dataset, focusing on various attributes related to apple quality. The dataset comprises information on Apple ID, Size, Weight, Sweetness, Crunchiness, Juiciness, Ripeness, Acidity, and Quality. The analysis involves data loading, preprocessing, and visualization to draw insights into the distribution and relationships within the data.

**Data Description**

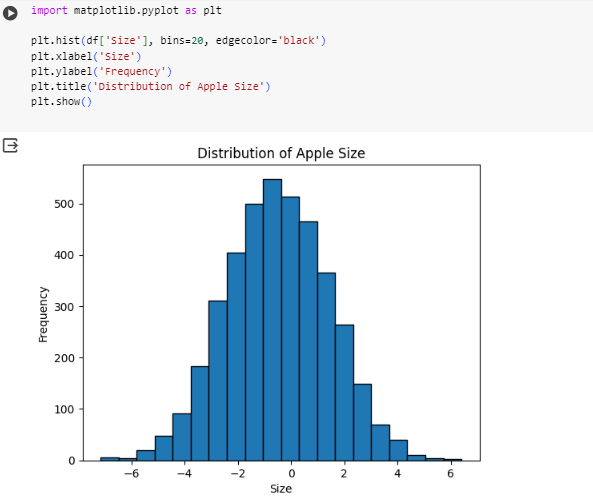
Dataset Overview:-

The dataset consists of 4001 entries with nine columns, including both numerical and categorical features. The attributes range from Apple ID to Quality, with some missing values in the 'Acidity' column. The dataset was preprocessed by dropping missing values, ensuring a clean dataset for analysis.

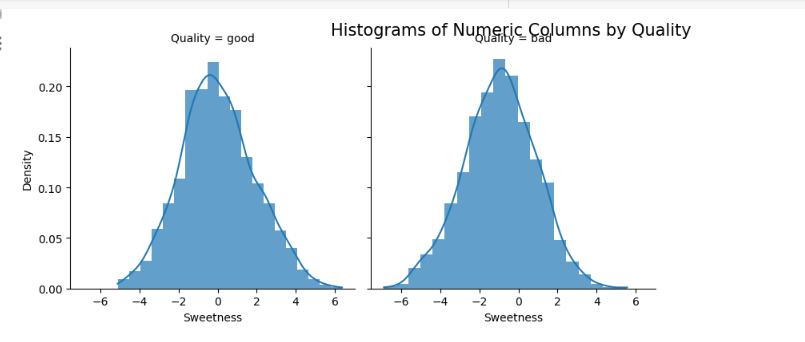




**Visualization:-**

**Histogram**:-****

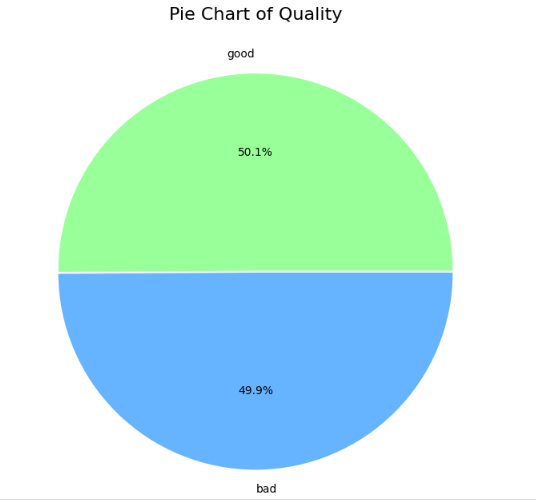
In the histogram, we visualized the distribution of apple sizes in the dataset. The histogram displays the frequency of different size ranges, divided into 20 bins. The x-axis represents the size of apples, while the y-axis shows the corresponding frequency of each size range. The uniform distribution observed suggests that apple sizes are evenly spread across the dataset, indicating a diverse representation of apple sizes in the given data.



Numeric Columns by Quality

A facet grid of histograms further explores numeric attributes concerning the 'Quality' category. The histograms are color-coded for 'good' and 'bad' qualities.

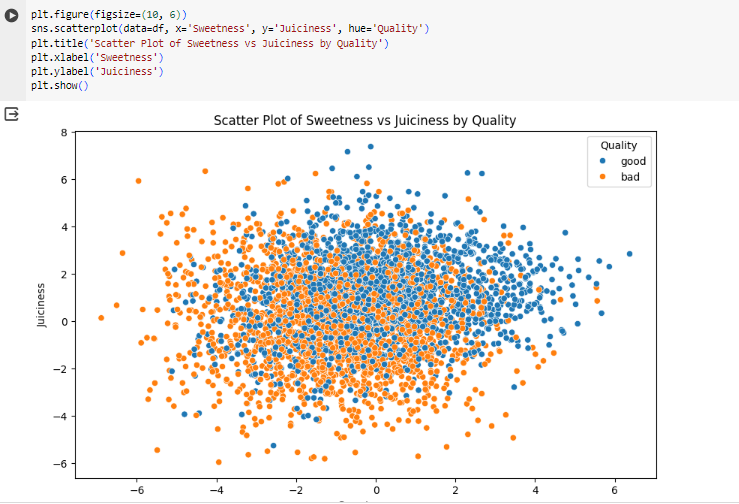
**Pie chart**:-



Pie Chart of Quality

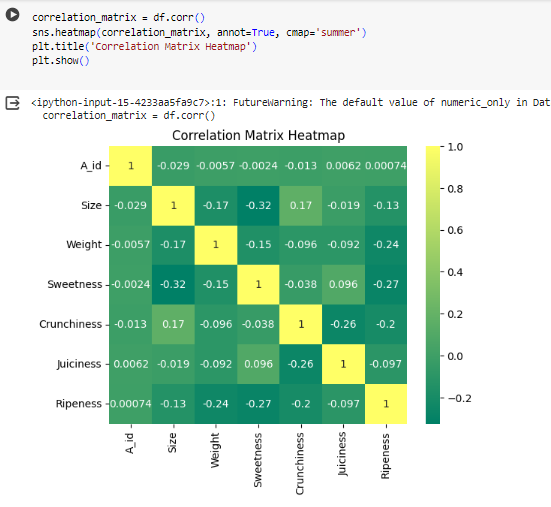
A pie chart illustrates the distribution of 'Quality' values in the dataset, indicating the proportion of 'good' and 'bad' apples.

**Scatter Plot:-**



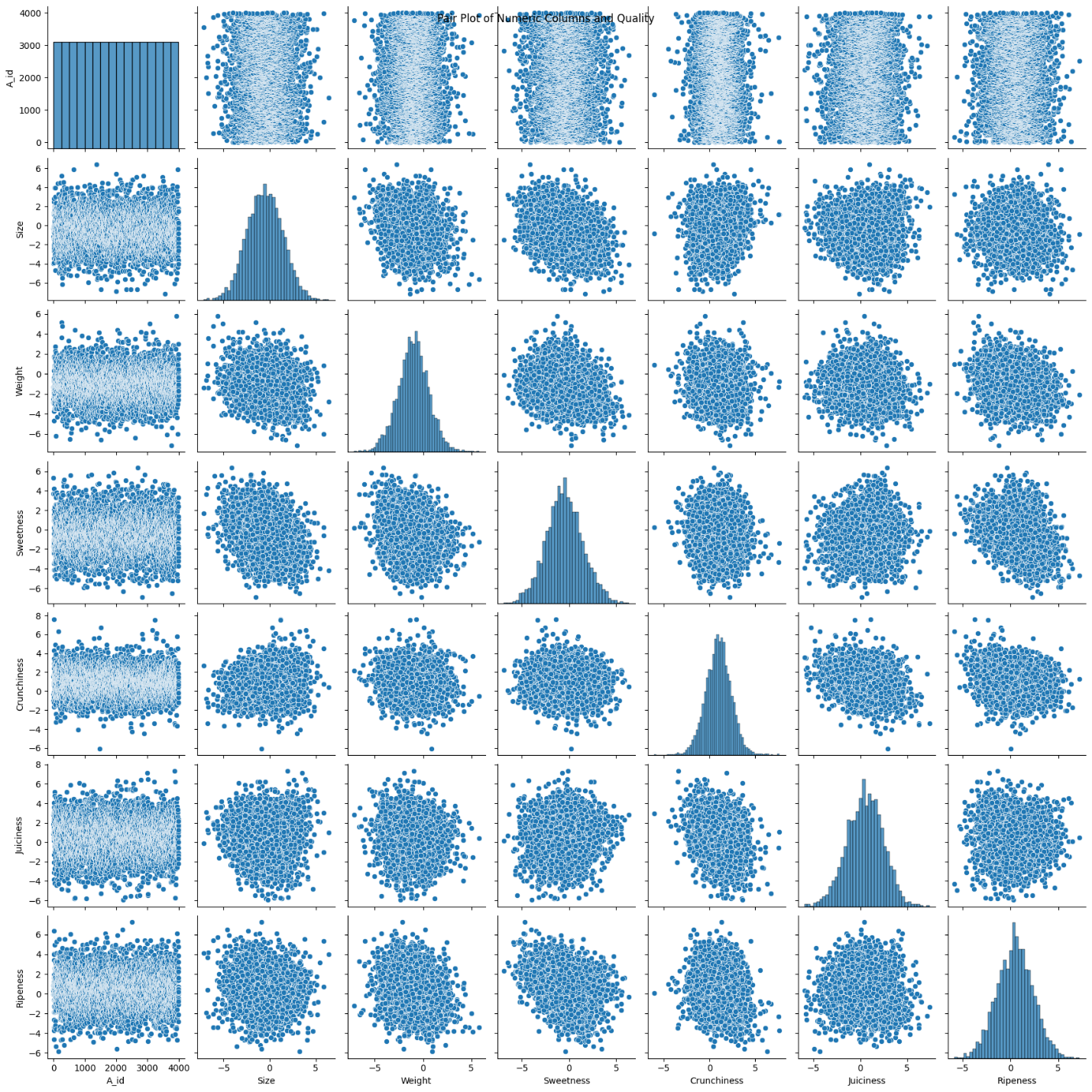
Scatter Plot of Sweetness vs Juiciness by Quality

A scatter plot visualizes the relationship between Sweetness and Juiciness, differentiating 'good' and 'bad' qualities.

**Correlation Matrix Heatmap:-** ****

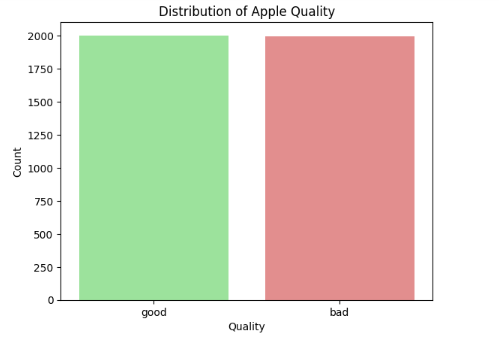
A heatmap provides a graphical representation of the correlation matrix, making it easy to identify strong correlations.

**Pair Plot:-**

****

A pair plot visualizes pairwise relationships among numeric columns and quality, offering insights into potential patterns

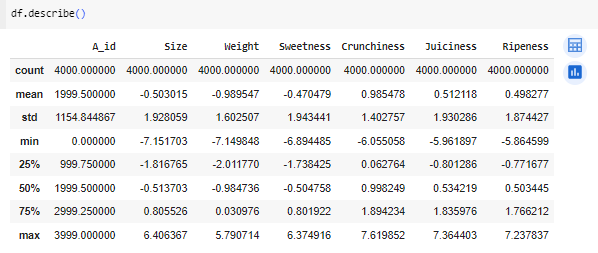
**Count Plot**:-



Count plot displays the distribution of apple quality, using light colors to distinguish between 'good' and 'bad' qualities.

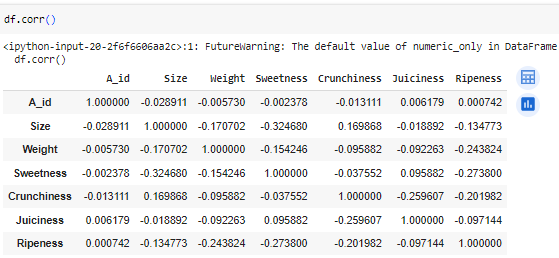
**Statistics**:-

**Descriptive Statistics:-**



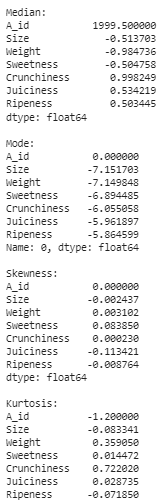
The descriptive statistics summary provides a comprehensive overview of the dataset's central tendencies and variability. It includes key metrics such as mean, standard deviation, and quartiles for each numerical attribute, offering insights into the distribution and range of values within the dataset.

**Correlation Matrix:-**



The correlation matrix illustrates the relationships between numeric attributes. Positive and negative correlations help identify patterns, indicating how attributes change concerning each other.

**Statistical Measures:-**



Computed median, mode, skewness, and kurtosis to capture central tendency, peak, asymmetry, and tail characteristics, providing deeper insights into the distributional properties of the numeric features.

**Conclusion: -** This analysis aligns with the objectives outlined in the question paper, covering essential components such as descriptive statistics and correlation matrices. The visualizations, including histograms and scatter plots, fulfill the requirement of presenting information in a graphical format.