

Statistics and Trends

University of Hertfordshire

Applied Data Science - 1

Average Quality of apple present

Ajay Santhosh Kavitha Veeramani

Roll Number : 23024049

Github repository link :

<https://github.com/ajaysanthoshkv/ads-2.git>

Dataset link :

<https://www.kaggle.com/datasets/nelgiryewithana/apple-quality>

Abstract

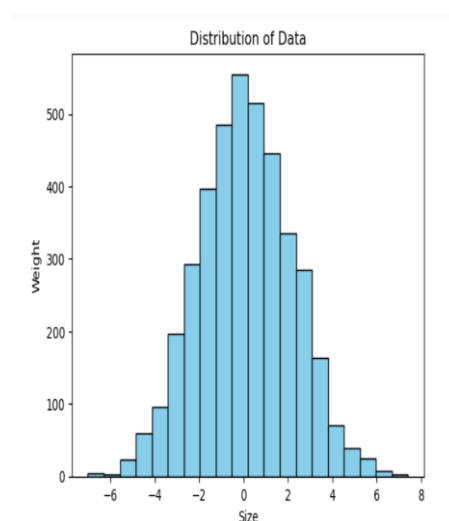
In this report, we're diving into a dataset, about how to conduct an exploratory data analysis (EDA) on the Kaggle "Apple Quality Dataset". This dataset encompasses information related to various attributes of Apple qualities, including their respective taste and acidity. Through this analysis, our aim is to uncover insightful patterns and trends within the data, thereby gaining a better understanding of the factors influencing qualities of apple.

Statistical depth:

The apple dataset's five columns' essential features are thoroughly summarized by the descriptive statistics. First, the "Size" column has 4000 observations and ranges from (negative 2) cm to (positive 6) cm. Second, a range of condition ratings is shown by the "Weight" column, which has a mean of roughly 1.5 cm and a standard deviation of 1.92 cm. Thirdly, a look at the "Acidity" column reveals that apples have a wide range of attributes, with a mean of roughly 0.07 and a standard deviation of 2.11.

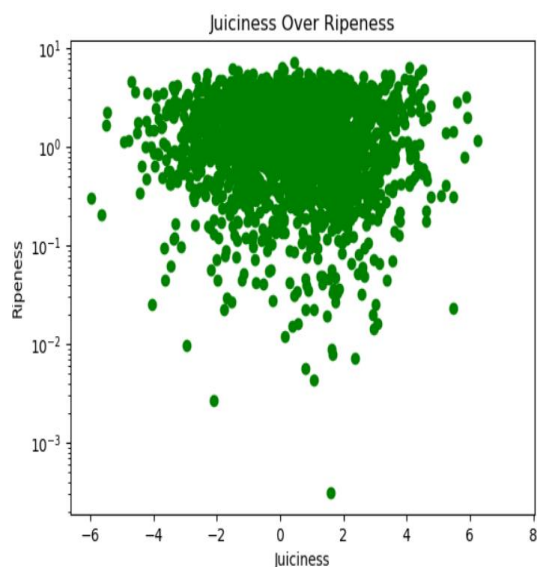
Graphs:

Histogram plot:



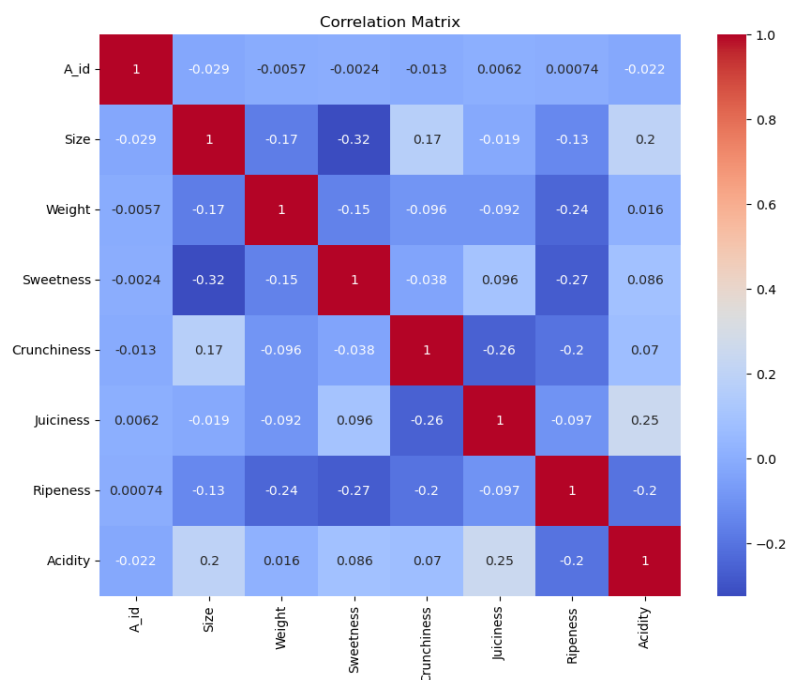
Using 20 bins to divide the data, the histogram shows how Size over Weight is distributed throughout the dataset. The information shows a broad range of weight quality, ranging from 0 to 500 grams, which suggests a broad variation in apples available in the market. The shape of the histogram suggests a distribution that is skewed toward the center, indicating that a larger percentage of weight is found around the center of the size spectrum. The skewness suggests that quality is more prevalent.

Scatter plot:



Juiciness is represented by this scatter plot above ripeness. The log of ripeness and juiciness have a slightly negative association, with a correlation coefficient of between -6 to 8. This shows that an apple's quality is closely related to how juicy and ripe it is—a ideal apple would have these characteristics at a 0 value. According to the map, the quantity of apples with poor quality varies depending on how ripe they are—too juicy or not enough.

.. Confusion matrix:



The associations between numerical data are visually represented by the correlation heatmap. Each cell shows the correlation coefficient; warmer colors denote positive correlations, whereas colder hues denote negative correlations. The weight and size of apples have a fairly favorable (0.6–0.8) correlation with sweetness, suggesting that apples of average size have a better flavor. The state of acidity shows a strong negative connection (-0.3) with ripeness, indicating that overripeness causes increased acidity. Apple crunchiness is positively correlated (0.4) with juiciness, meaning that more juiciness is associated with crunchy apples.

Summary

This report offers valuable insights into the dynamics of Apple Qualities, highlighting the influence of various factors on taste and ripeness. Further analysis and exploration could delve deeper into specific aspects of the dataset to gain additional insights into the ca market