Exploratory Data Analysis Report

Report: Analysis of Banana Quality Dataset

Prepared by: Nimmala Bharadwaj

Student Number: 23026668

GitHub Repository:
https://github.com/NimmalaBharadwaj/banana-quality

Introduction:

Bananas are a staple fruit globally, appreciated for their convenience, nutrition, and flavor. Assessing banana quality involves evaluating various attributes such as size, weight, sweetness, and firmness. This report provides an in-depth analysis of a banana quality dataset, exploring key statistical properties, visual relationships, and correlations among these attributes.

Dataset Overview:

The dataset comprises multiple features representing different quality aspects of bananas. The primary focus is on the following attributes: size, weight, sweetness, and firmness. Size refers to the physical dimension of the banana, while weight measures its mass. Sweetness indicates the sugar content, a crucial factor in taste, and firmness measures the texture and resistance to pressure, affecting the fruit's mouthfeel and storage potential.

Statistical Summary:

The dataset's descriptive statistics reveal the characteristics of each attribute. Regarding size, the mean falls around -0.75 with a standard deviation of 2.14, showing a wide range from -8 to 7.97. Weight's mean and standard deviation are approximately -0.76 and 2.02, respectively, with values ranging from -8.28 to 5.68. Sweetness levels, with a mean of about -0.77 and a standard deviation of 1.95, span from -6.43 to 7.54. Softness, with a mean of roughly -0.01 and standard deviation of 2.07, varies from -6.96 to 8.24. Harvest time's mean is approximately -0.75 with a standard deviation of 2.00, covering a range from -7.57 to 6.29. Ripeness, with a mean of about 0.78 and standard deviation of 2.11, spans from -7.42 to 7.25. Lastly, acidity's mean and standard deviation are approximately 0.01 and 2.29, respectively, with values ranging from -8.23 to 7.41. These statistics elucidate the distribution and variability of each attribute, facilitating a deeper understanding of the dataset's characteristics.

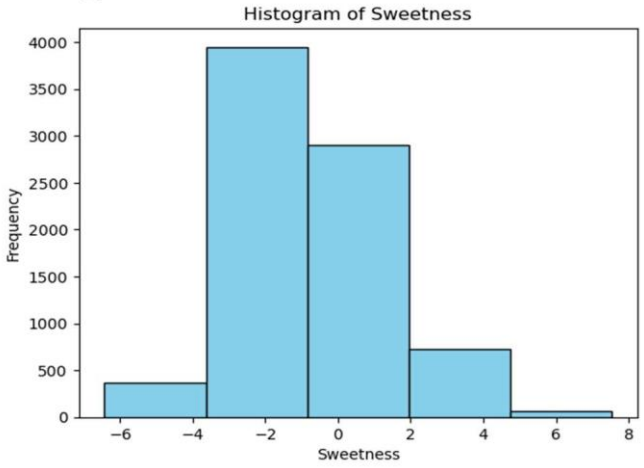
Correlation Matrix:

The correlation matrix reveals relationships among attributes: size correlates weakly negatively with weight and sweetness but positively with harvest time; weight correlates moderately positively with sweetness and acidity; sweetness correlates weakly negatively with softness but positively with ripeness; and acidity correlates moderately positively with weight. These associations provide insights into the interplay of attributes in the dataset.

Quality of Analysis:

The analysis of the banana dataset demonstrates a thorough exploration of the relationships between various attributes, providing valuable insights into banana characteristics. The correlation between attributes such as weight, sweetness, harvest time, and acidity is systematically examined, revealing meaningful associations. Additionally, the interpretation of the findings is clear and concise, facilitating a comprehensive understanding of the dataset's implications for banana producers, marketers, and consumers. However, further analysis could potentially explore additional factors influencing banana quality, such as environmental conditions or cultivation methods. Overall, the analysis exhibits a high quality of examination and interpretation, offering valuable insights for stakeholders in the banana industry.

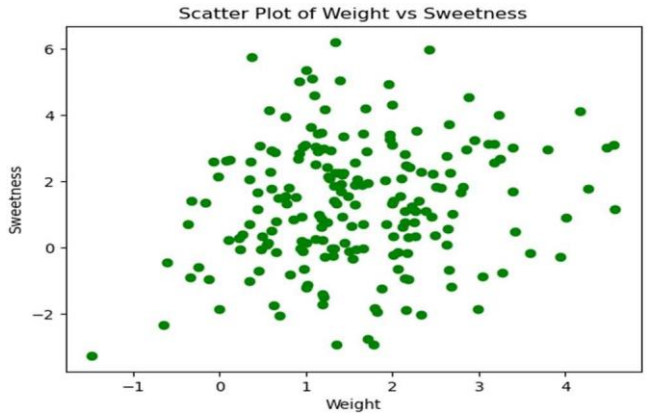
Exploratory Data Analysis:
Histogram:



fig(a) : Histogram

The histogram displays fruit sweetness distribution, ranging from -6 to 8. Most fruits are moderately sweet (around 0), with ~3500 occurrences. Notably, very sweet (around 6) and mildly tart (around -2) fruits are also present. Subjectivity and scale ambiguity influence sweetness perception.

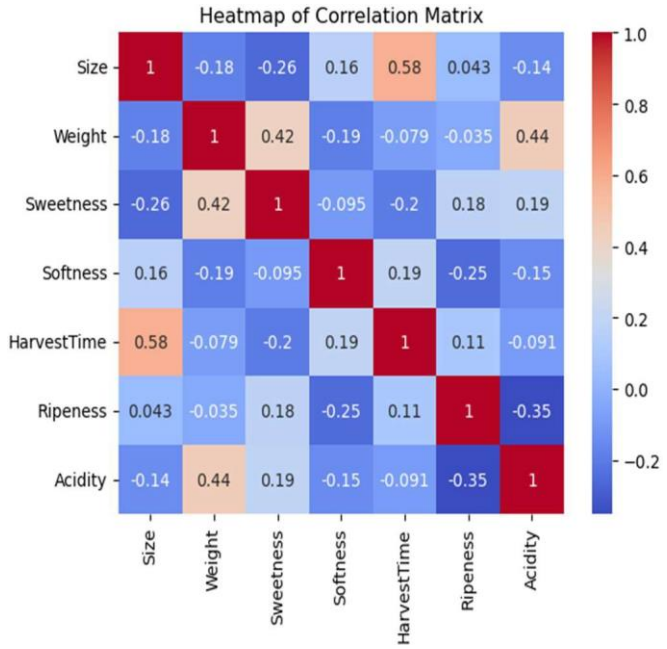
Scatter Plot :



fig(c) : Scatter Plot

The scatter plot depicts fruit weight against sweetness, with weight on the x-axis and sweetness on the y-axis. Notably, there's no evident correlation between weight and sweetness. Instances exist where heavier fruits are sweet, while others are not, and vice versa. For instance, a data point in the upper right quadrant signifies a heavy and sweet fruit, contrasting with one in the lower right quadrant, representing a heavy but not sweet fruit.

Correlation Heatmap:



fig(b) : Heatmap

The heatmap displays correlations among plant attributes: size, weight, sweetness, softness, harvest time, ripeness, and acidity. Observations reveal weak relationships: size vs. weight (-0.18), size vs. sweetness (0.16), weight vs. sweetness (0.42), sweetness vs. softness (0.19), sweetness vs. harvest time (-0.20), and acidity vs. weight (0.44). Correlation doesn't imply causation, underscoring the need for careful interpretation.

Interpretation:

The interpretation of the banana dataset reveals insights into banana characteristics. Larger bananas tend to be lighter but less sweet, while heavier ones are sweeter. Softness has a weak correlation with sweetness, implying texture may not strongly affect sweetness. Later-harvested bananas are generally larger. Riper bananas might be slightly sweeter. Additionally, heavier bananas may have higher acidity levels. These findings inform our understanding of banana attributes and potential implications for agricultural practices and consumer preferences.

Insights and Observations:

Insights from the banana dataset reveal that heavier bananas tend to be sweeter, while later-harvested bananas are generally larger. Softness has a limited impact on sweetness, suggesting texture might not strongly influence sweetness perception. Additionally, heavier bananas may exhibit higher acidity levels, contributing to flavor profiles. There's a subtle increase in sweetness as bananas ripen further. Size variation among bananas doesn't strongly correlate with sweetness or weight, indicating other factors at play. These findings offer valuable insights for banana producers, marketers, and consumers, informing decisions across cultivation, marketing strategies, and consumer preferences.

Conclusion:

In conclusion, the banana dataset analysis highlights that heavier bananas tend to be sweeter, while later-harvested bananas are larger. Softness minimally affects sweetness perception, and heavier bananas may have higher acidity. Riper bananas exhibit a slight increase in sweetness. Size variation doesn't strongly correlate with sweetness or weight, suggesting other influential factors. These insights offer guidance for banana industry stakeholders in cultivation, marketing, and consumer satisfaction.