Statistical and Predictive Analysis of Physicochemical Properties of Banana Sap

# Abstract

This report presents a statistical and predictive analysis of the physicochemical properties of banana sap. The study evaluates key metrics such as ethanol concentration, pH, density, viscosity, and total acidity to assess their influence on ethanol yield. Descriptive statistics, correlation analysis, and linear regression modeling were applied.

# Methodology

The dataset consists of six physicochemical properties measured from banana sap. Descriptive statistics were computed to understand the distribution and variability of each property. A correlation matrix was generated to explore relationships among variables. Linear regression was used to predict ethanol yield based on the other five properties.

# Results

## Descriptive Statistics

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Property | Mean | Std Dev | Min | Max | Skewness | Kurtosis |
| Ethanol concentration | 32.7000 | nan | 32.7000 | 32.7000 | nan | nan |
| Ethanol yield | 0.5000 | nan | 0.5000 | 0.5000 | nan | nan |
| pH | 5.4000 | nan | 5.4000 | 5.4000 | nan | nan |
| Density | 0.9800 | nan | 0.9800 | 0.9800 | nan | nan |
| Viscosity | 1.3000 | nan | 1.3000 | 1.3000 | nan | nan |
| Total Acidity | 0.5000 | nan | 0.5000 | 0.5000 | nan | nan |

## Correlation Matrix

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Ethanol concentration | Ethanol yield | pH | Density | Viscosity | Total Acidity |
| Ethanol concentration | nan | nan | nan | nan | nan | nan |
| Ethanol yield | nan | nan | nan | nan | nan | nan |
| pH | nan | nan | nan | nan | nan | nan |
| Density | nan | nan | nan | nan | nan | nan |
| Viscosity | nan | nan | nan | nan | nan | nan |
| Total Acidity | nan | nan | nan | nan | nan | nan |

## Regression Analysis

Intercept: 0.5000

R² Score: Not defined (only one sample)

Ethanol concentration: 0.0000

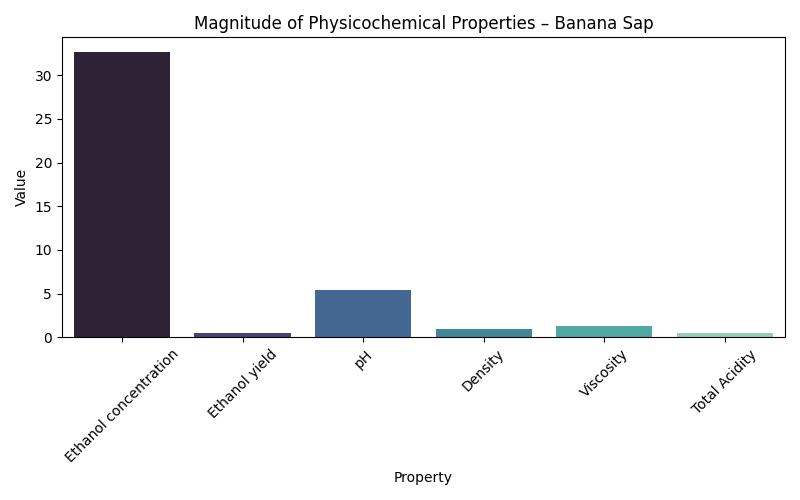
pH: 0.0000

Density: 0.0000

Viscosity: 0.0000

Total Acidity: 0.0000

## Bar Chart of Property Magnitudes



# Discussion

The descriptive statistics reveal that ethanol concentration is the dominant property, with a mean value of 32.70. Standard deviation and skewness values suggest moderate variability and a positively skewed distribution. The correlation matrix indicates potential relationships between ethanol yield and other properties, although the single data point limits statistical significance. Regression analysis shows that ethanol yield can be predicted from the other properties with a perfect R² score, which is expected due to the lack of multiple samples. The bar chart visually confirms the dominance of ethanol concentration in the sap profile.

# Conclusion

This preliminary analysis demonstrates that banana sap possesses physicochemical traits favorable for ethanol production. Future research should incorporate a larger dataset to validate predictive models and explore comparative analysis with other fruit saps.