M4 Discussion

**1. Variable Description**

**1.1. Malic Acid – Variable Description**

Malic and tartaric acids are the most prominent organic acids in wine and play a crucial role in the winemaking process, including the organoleptic quality and the physical, biochemical and microbial stability of wine. Their respective levels found in wine can vary greatly but in general one would expect to see 1,000 to 4,000 mg/L tartaric acid and 0 to 8,000 mg/L malic acid. Deacidification of grape must and wine is often required to produce well-balanced wines. Malolactic fermentation induced by the addition of malolactic starter cultures, regarded as the preferred method for naturally reducing wine acidity, efficiently decreases the acidic taste of wine, improves the microbial stability and modifies to some extent the organic character of wine (Volschenk et al., 2006).

Besides the importance of flavor compounds, acidity in wine directly or indirectly affects several different levels of the winemaking process and ultimately determines wine quality in terms of the perceived organoleptic and aesthetic character. Wine acidity also influences the ageing potential or the shelf-life of wine, as it determines the physical, biochemical and microbial stability of wine. In addition, wine acidity and pH affect the timely succession of cellar events and the effectiveness of several techniques applied by winemakers.

**1.2. Measures Wine producers Use to Measure Wine Quality**

Wine’s quality is assessed by physicochemical and sensory tests. Physicochemical tests include features such as density, alcohol, and pH values. The production of quality wines requires a judicious balance between the sugar, acid and flavour components of wine.

Sensory tests are performed by human taste experts which is a very complex and expensive process. In our literature review, we found eleven consistent physicochemical properties common to almost all data sets designed to evaluate wines’ quality (Cortez et al., 2009). The eleven physicochemical properties are: fixed acidity, volatile acidity, citric acid, residual sugar, chlorides, free sulfur dioxide, total sulfur dioxide, density, pH, sulphates, and alcohol (Cortez et al., 2009)

**2. Table: Descriptive Statistics of Malic Acid for Each Wine Type**

Table

Description automatically generated

**3. Boxplot: Malic Acid for Each Wine Type**

**Chart, box and whisker chart

Description automatically generated**

**4. Observation, Results and Conclusions**

**4.1. Observation from Table and Graph:**

1. After checking the distribution of wine type with a histogram on the side, the data clearly shows that the distribution for Aurora and Rainbow wine for Malic Acid is not symmetric because mean and median is very far from each other. On the other hand, Sunset Wine has symmetric distribution.
2. Also, we can observe from Boxplot Graph that Aurora has more outliers than Rainbow and Sunset. And Sunset has no outliers which is good sign for Wine Quality
3. Sunset has no skewness which is good sign, but Aurora and Rainbow are right (positive) skewed

**4.2.Used Box Plot Graph for my data analysis of wine quality because of the following reasons:**

1. It allows to compare many categorical variables with one continuous numerical variable very nicely.
2. It helps to assess the spread of the points to determine how much your sample varies. The greater the variation in the sample, the more the points will be spread out from the center of the data.
3. Often, skewness is easiest to detect with a histogram or boxplot. When data are skewed, most of the data are located on the high or low side of the graph.
4. To check for Outliers, which are data values that are far away from other data values, can strongly affect the results of your analysis. Often, outliers are easiest to identify on a boxplot.

Conclusions

1. **Challenges and questions**: Explain any challenges you encountered preparing your work and ask your classmates any question or advice needed.

Challenges

1. I was really confused with what type of graph should I use for data analysis, then I explored online what type of graph is best for comparing numerous categorical variables with one continuous variable. I found that boxplot is best for this analysis.
2. While reading the research papers about the importance of Malic Acid in Wine Making Process, I had to encounter many tough scientific terms during my reading for e.g., Fermentation, Vinificationetc. Then, I watched some videos about them to get familiar, so it will increase my understanding of text.

Advice, After the rounding off digits in my matrix, some digits couldn’t which have zero at end aren’t rounding off.

References

1. Cortez, P., Cerdeira, A., Almeida, F., Matos, T., & Reis, J. (2009). Modeling wine preferences by data mining from physicochemical properties. *Decision support systems*, *47*(4), 547-553.
2. Volschenk, H., Van Vuuren, H. J., & Viljoen-Bloom, M. (2006). Malic acid in wine: Origin, function and metabolism during vinification.