# 1. Introduction

## 1:1 Problem statement and research motivation

The ever-growing wine industry has seen a decline in both consumption and production recently. The growth of the industry relies on the satisfaction of consumers. Thus, the relationship between alcohol content and wine quality is a critical topic in the wine industry, as prior research suggests that alcohol content can significantly impact the sensory characteristics of a wine (Teng-Zhen Ma , et al., 2021). Ensuring the wine industry relishes its glory days again with the help of our research provided us the motivation to figure out whether wine with higher alcohol content has a significantly higher quality rating than then the wines with lower alcohol content.

## 1:2. The data set

Our dataset “wine quality-white” was derived from Kaggle containing a total of 4898 samples of wine along with 12 features next to them. Although our research primarily prioritises on the feature of Alcohol (percentage of Alcohol) and wine quality (rating from 1-10), other features such as density, PH, residual sugar, citric acid etc are also key components of a wine sample which are present in the dataset that can lead many more compelling and insightful research. The dataset is ideal for studying the impact of alcohol content on wine quality.

## 1:3. Research question

To ascertain whether the proportion of alcohol quantity connects to alcohol quality we needed to craft an excellent research question. Therefore, it lead to the research question of –

Is there a difference in the mean quality rating between wines with low alcohol content and wines with high alcohol content?

We plan to answer our research questions through statistical testing revolving around our hypothesis. We want use visualizations such as histogram to test the normality of the data, boxplots to illustrate differences and statistical test such as T-test, Wilcoxon test to evaluate significance. We intend to use the language R through R Studio, utilizing the dataset to explore the effects of alcohol content on wine quality.

## 1:4. Null hypothesis and alternative hypothesis (H0/H1)

Over the course of this assessment, we want to investigate the effect of alcohol content on the quality ratings of white wine. To establish this, we approach it scientifically, establishing two competing hypotheses-

**Null Hypothesis**: There is no difference in the mean quality ratings between the wines with low alcohol content and wines with high alcohol content.

**Alternative Hypothesis**: There is a difference in the mean quality ratings between wines with low alcohol content and wines with high alcohol content.

Though our statistical testing we will be able to evaluate the hypothesis under defined significance level leading us to either reject or accept the null hypothesis.

# 2. Background research

## 2.1. Research papers

Research centring around wine is not a new phenomenon as the earliest production of wine date up to 6000 BC. However, the modern technology provides greater opportunity to research on wine and its elements. So, it is not an unfamiliar event that we found multiple journals that worked with the same dataset as ours.

The research paper titled “Prediction of wine quality using machine learning algorithms (Dahal, et al., 2021)” harnessed our white wine dataset to establish relation between different wine components and the quality using machine learning algorithms. Noteworthy corelation between alcohol content and wine quality was observed. Therefore, it provides us with the basis to study the relation of alcohol with wine quality for different levels.

An article entitled “The influence of Alcohol on the sensory perception of red wines (King, et al., 2013)” established the idea of alcohol content influencing sensory attributes like aroma and taste for the red wines. The article manages to proof the wines containing higher level of alcohols tend to be more intense especially associating with sensory perceptions. Senses are highly corelated with the quality of wine providing us with evidence that alcohol content might influence overall wine quality further solidifying our reason to research on our chosen topic.

The article titled “Impact of dealcoholisation on quality properties in white wine at various alcohol content levels (Liguori, et al., 2019)” emphasizes on the effect of alcohol on white wine’s taste, aroma and volatile compounds through the process of dealcoholisation. The research paper suggests that alcohol exerts a significant influence on the overall texture of a wine.

## 2.2. Why RQ is of interest

Numerous studies have suggested that Alcohol is a key determinant of wine quality as it plays a pivotal role in the sensory attributes such as taste, aroma and texture. Only a handful of research compares the wine quality across different alcohol levels in white wine, establishing a sustainable gap in the research. Our study addresses the gap by comparing means of wine with different alcohol levels statistically and visually demonstrating how wine quality is influenced by alcohol level. Future studies might explore how different wine components other than alcohol and wine quality corelate applying advanced machine learning to improve predictions, contributing further to wine analysis.

# 3. Visualisation

## 3.1. Appropriate plot for the RQ

## A chart with a bar graph Description automatically generated

Figure 1 (Boxplot of Alcohol level with respect to wine quality)

The visualisation of Boxplot is the most appropriate for our research as it effectively demonstrates how the two-alcohol level high and low differs from each other in variability and potential outliers. The upward shift of wine with higher alcohol content hints the possibility of difference in the means of the wo alcohol levels. The visualization aligns with the premise of our research.

## 3.2. Additional information relating to understanding the data

A graph of wine quality

Description automatically generated

Figure 2 (histogram with bell curve on top of dependent variable)

The Histogram of our dependent variable (wine quality) is essential for our research as this will allow us to form the right statistical test. The bell curve and the histogram structure suggest normality of the data suggesting T-test is more suitable for testing our hypothesises. For the exception a Wilcox test would have been a better alternative. Thus, the histogram emerges as an essential tool to for selecting the appropriate statistical test.

## 3.3. Useful information for the data understanding

# 4. Analysis

## 4.1. Statistical test used to test the hypotheses and output

To examine the hypothesis the independent T-test was chossen as the dependent variable Wine quality is mostly normally distributed because it follows the bell curve shape. Normality is the primary condition of the T-test which makes it superior to the Wilcoxon test in our specific scenario. The Wilcoxon Test is non-parametric and doesn’t require normality. T-test will effectively compare the mean of the two different alcohol levels aligning with our research question.

## 4.2. The null hypothesis is rejected /not rejected based on the p-value

According to the result of the T-test the null hypothesis stating there is no difference in the mean quality between the wines with low alcohol content and wines with high alcohol content can be rejected. The p-value (<2.2e-16) is significantly smaller in contrast to the common significance level of 0.05 provides compelling evidence to decisively refuse the null hypothesis. The 95 percent confidence interval further of 0.63966262 and 0.7332534 further give more evidence that the difference in mean lies within this range confirming a difference exists between means of two different alcohol levels supporting to reject the null hypothesis.

# Evaluation – group’s experience at 7COM1079

## 5.1. What went well

In order to guarantee that duties were finished effectively, the group maintained constant contact and cooperation. Weekly gatherings made it easier to coordinate goals and proactively handle obstacles. Code conflicts were reduced, and the project workflow was optimized by using GitHub for version control. By offering their special talents, members improved the calibre of data analysis and visualization. Clear and succinct results were obtained because the investigation was carried out methodically. Overall, the timely and successful completion of project milestones was made possible by our dedication and teamwork.

## 5.2. Points for improvement

## 5.3. Group’s time management

## 5.4. Project’s overall judgement

Our project effectively examined how alcohol content impacts the quality of white wine. By applying statistical tests and visualizations, we were able to assess differences in quality ratings. The findings met our initial expectations, though there is room for further analysis of additional features. Overall, the team was pleased with the outcome.

## 5.5. Comment on the GitHub log output

# Bibliography

Dahal, K. R., J. N. D., H. R. B. & S. G., 2021. Prediction of wine quality using machine learning algorithms.

King, E. S., R. L. D. & H. H., 2013. The influence of alcohol on the sensory perception of red wines.

Liguori, L. et al., 2019. Impact of dealcoholization on quality properties in white wine at various alcohol content levels.

T.-Z. M.et al., 2021. Techniques for Dealcoholization of Wines: Their Impact on Wine Phenolic Composition, Volatile Composition, and Sensory Characteristics.