Dramatic increase in wine consumption has been observed in a global scale. This could be attributed to positive effects of wine consumption to cardiovascular health. Consequently, this has led to augmented sales of wine in the last 5 years [1]. In response to the aforementioned growth in wine consumption and its sales, the industry has been innovating new methodologies in wine production to increase its efficiency and daily output. Additionally, innovation and development of strategic marketing tactics are being performed to optimise its sales volume [2].

There is an extensive variety of red wine that is available in the international market. Despite the differences in wine variety, chemical tests indicate that the majority of red wine has similarity in its chemical components. However, its proportions have varied concentrations in different types of red wine. The difference in chemical quantity is a significant aspect in order to classify each type (Reference). Due to lack of sophisticated technology, there had been a difficulty in performing wine classification in the past. However, data mining techniques such R-language and Python have enable to not only classify wines, but to determine the significance of each chemical component and its correlation to taste, quality and other features. This would allow wineries to perform exclusion or reduction of chemical components that are unnecessary or could potentially affect the wine quality (Reference).

In this report, a public data set of red wine from UCI Repository would be utilised to perform data analysis using R-studio. To assess and predict the wine quality, different methods such as Linear Regression, Decision Trees and Principal Component Analysis would be used (Reference).

World has seen a decent increase in wine consumption lately. This is perhaps due to a recent positive correlation found between wine consumption and healthy heart rate. Consumers now enjoy wine more and more. In response to this increased wine consumption, the industry is now looking for new efficient ways of wine production and marketing tactics to support this growth [1]. There are various kinds of wines in the market and each wine has a different purpose and is produced for a different occasion. Although, most wines contain the same chemicals based on chemical tests, their proportions have varied concentrations in different types of wines. This difference in quantity of chemicals in various wines makes it important to classify each type. In the past, it was difficult to make such classifications due to lack of technology but today with the help of data mining techniques such as R-language and Python, it is possible to not only classify wines but to work out the importance of each chemical in wine in terms of taste, quality and other features. This would allow the wineries to exclude or reduce the chemical components which are unnecessary or have a negative correlation with wine quality.

In this report, we will be using a public dataset of red wine from UCI Repository to perform data analysis using R-studio. For the purpose of analysis, we will use different methods such as Linear Regression, Decision Trees and Principal Component Analysis to assess and predict the quality of red wine.

REFERENCES:

[1] H. Sohaib, A. Bryce, and A. Baranchuk, “Wine and Cardiovascular Health: A Comprehensive Review,” Circulation, 136 (15). DOI: 10.1161/CIRCULATIONAHA.117.030387

[2] P. Cortez, A. Cerderia, F. Almeida, T. Matos, and J. Reis, “Modelling wine preferences by data mining from physicochemical properties,” In Decision Support Systems, Elsevier, 47 (4): 547-553. ISSN: 0167-9236.

[**Wine and Cardiovascular Health: A Comprehensive Review**](https://acu-edu-primo.hosted.exlibrisgroup.com/primo-explore/fulldisplay?docid=TN_ovid00003017-201710100-00009&context=PC&vid=61ACU&lang=en_US&search_scope=61ACU_All&adaptor=primo_central_multiple_fe&tab=61acu_all&query=any,contains,wine%20consumption%20+%20cardiovascular%20health&offset=0)

Haseeb, Sohaib ; Alexander, Bryce ; Baranchuk, Adrian

Circulation, 2017, Vol.136(15), pp.1434-1448 **DOI:**10.1161/CIRCULATIONAHA.117.030387