Executive Summary

Project 4 Group 3 Members: Garrett Kidd, Matthew Bailey, Brenna(Bee) Wallace, Shaunghnessy Robertson

This project aims to use machine learning techniques to predict the quality of the wine based on several features such as alcohol content, pH value, acidity, residual sugar, and more. Examples of visualizations we drew inspiration from to replicate in our project are shown below. This project will use publicly available datasets of red and white wines, each with its features and a quality score. The aim is to build a predictive classification model that can accurately classify wines into two quality categories while also classifying whether the wine is Red or White. Our other goal was to find the region where the best-rated wine originated and where the most wine was produced. We used visualizations primarily through Tableau of bar charts, maps, etc. We chose the datasets from Kaggle called Wine Quality Data Set (Red & White Wine) and Wine Rating & Price.

To answer the question of the region where the most wine is produced, and the highest-rated wine is created, we ran our Wine Rating & Price dataset. Our Tableau visualizations were created showing the region where the most wine is produced, including red and white wine. We also showcased a map where users could see the countries.

Our web application contained several pages, including a landing page, about us, a data table, a machine learning data table, a machine learning wine prediction page, works cited, a dashboard with Tableau visualization, dashboard 2 with another Tableau visualization, and finally a page with an embedded PDF of our executive summary of the project. The landing page contains a brief high-level description of the features and functionality of our website with a navigation bar of dropdowns that the user can use to navigate to other site pages. Our machine-learning wine prediction page contains a button to click to predict both the quality of wine and the type of wine. It also includes a guide for inputting values so that users know the minimum and maximum values to input. If users input values outside of the parameters. The prediction is based on one XGBoost machine learning model and one RandomForest model we created that gave a score of the value of the provided features in our Wine Quality Data Set (Red & White Wine) that we initially combined into one CSV.

We hope that users who use our website can find and know wine quality simply by inputting the relative values of the features and whether or not it is a white or red wine. But even more importantly, we hope that users will understand why the quality rating was given from looking at our data and predictive model. From learning this, users can then, if desired, find the region in which our Tableau visualizations produced the wine.