Vinho Verde Quality Analysis

# ABSTRACT

# Consider a situation where some samples of red wine are to be tested in order to check their quality in a manufacturing factory (say, a brewery). This project aims to check the quality of red wine by testing the several parameters included in the dataset. This involves modeling wine preferences by data mining from physicochemical properties.

The data can be used to test(ordinal) regression or classification (in effect, this is a multi-class task, where the classes are ordered) methods. Other research issues are feature selection and outlier detection. The data includes the following dataset:

 winequality-red.csv - red wine preference samples

Key words: *Data set, wine, quality, regression, research*

### INTRODUCTION

Exploratory data analysis is an approach of analyzing data sets to summarize their main characteristics, often using statistical graphics and other data visualization methods. Using this concept, this project has been brought to life. The dataset was designed keeping in mind the beverage of Vinho Verde, a beverage of North Portugal. The main takeaway was to use the power of **EDA** to establish which **factors** contribute to a great glass of red wine. 1599 wines were tested by at least 3 wine experts and given a quality grade from **0** (terrible) to **10** (excellent). There were **11 factors** measured: fixed acidity, volatile acidity, citric acid, residual sugar, chlorides, free and total sulphur dioxide, density, pH, sulphates and alcohol.

### OBJECTIVES

* To visualize the data set to maximum extent
* To understand the development of an interactive admin panel using the html, css, javascript.
* To apply various visualizing techniques learnt in class
* To understand the various libraries of Python in Jupyter Notebook
* To analyze the attributes and decide the match strategically

# REVIEW OF LITERATURE

## Epidemiological studies have supported that red wine consumption is more CHD-preventative in comparison to the intake of other alcoholic beverages . It is uncertain whether the apparent beneficial properties for health attributed to the consumption of red wine are due solely to the presence of alcohol or also to the concerted action of alcohol and antioxidant compounds other than alcohol present in red wine . In addition to alcohol, red wine contains a wide range of active compounds—polyphenols—with antioxidant and anti-inflammatory properties that could contribute to protection from atherosclerotic pathologies. (Mizik and Jacobson 2003; Spiteri and Dion 2004; Woodruff 1997).

This has been recognized as the fundamental basis in every marketing activity (Holbrook 1994, 1999), and it has been envisioned as a critical strategic weapon in attracting and retaining heart health(Lee and Overby 2004; Wang, Lo, Chi, and Yang 2004).

## Recognition of the relevance of this concept has generated important research focused on the study of its composition and its relationship with other concepts of interest to marketers such as satisfaction, trust, and loyalty. However, even though there is a significant body of knowledge about the concept of consumer value, this research is rather fragmented.

## The extent and heterogeneity of the various studies have created a dispersed, sometimes confusing and still-inconclusive base of knowledge about consumer value. As Wang et al. (2004) contend, different points of view about the meaning of value are advocated in the literature, with no widely accepted way of pulling views together.

## Ulaga (2001, p. 318) regards that "the fundamental question of how to conceptualize value still merits further investigation." Moreover, relevant studies have not yet yielded any unambiguous results.

## Red wine contains high concentrations of polyphenolic compounds such as flavonoids (catechin, epicatechin, quercetin, anthocyanins, and procyanidins), resveratrol (3,5,4′-trihydroxystilbene), and polymeric tannins . In general, red wine is rich in polyphenols and may be considered as an important polyphenols source in the diet. The presence of phenolic compounds in red wine seems to be crucial, since scientific studies have reported that these important secondary metabolites are responsible for desirable biological actions, including cardiovascular protection effects.

Although excessive alcohol intake is associated with the development of chronic diseases and other serious problems, a wealth of data from scientific evidence support an inverse relationship between moderate alcohol consumption and the risk of CHD.

Moderate alcohol consumption is defined in the Dietary Guidelines for Americans 2015–2020 as up to one unit of alcohol per day for women and up to two units of alcohol per day for men.

The main objective of this review is to summarize the various red wine components and their effect in influencing the overall quality of the red wine that is to be consumed by the indirect stakeholders, known as consumers.

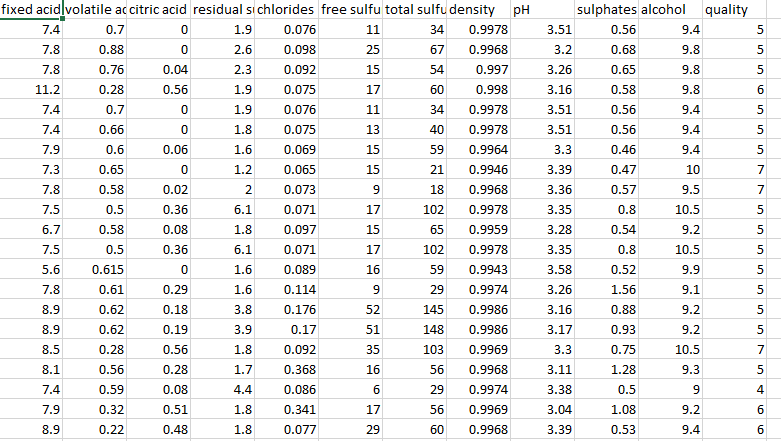
All studies regarding the relationship between red wine consumption and CHD published over the last decade have been taken into consideration.

### METHODOLOGY

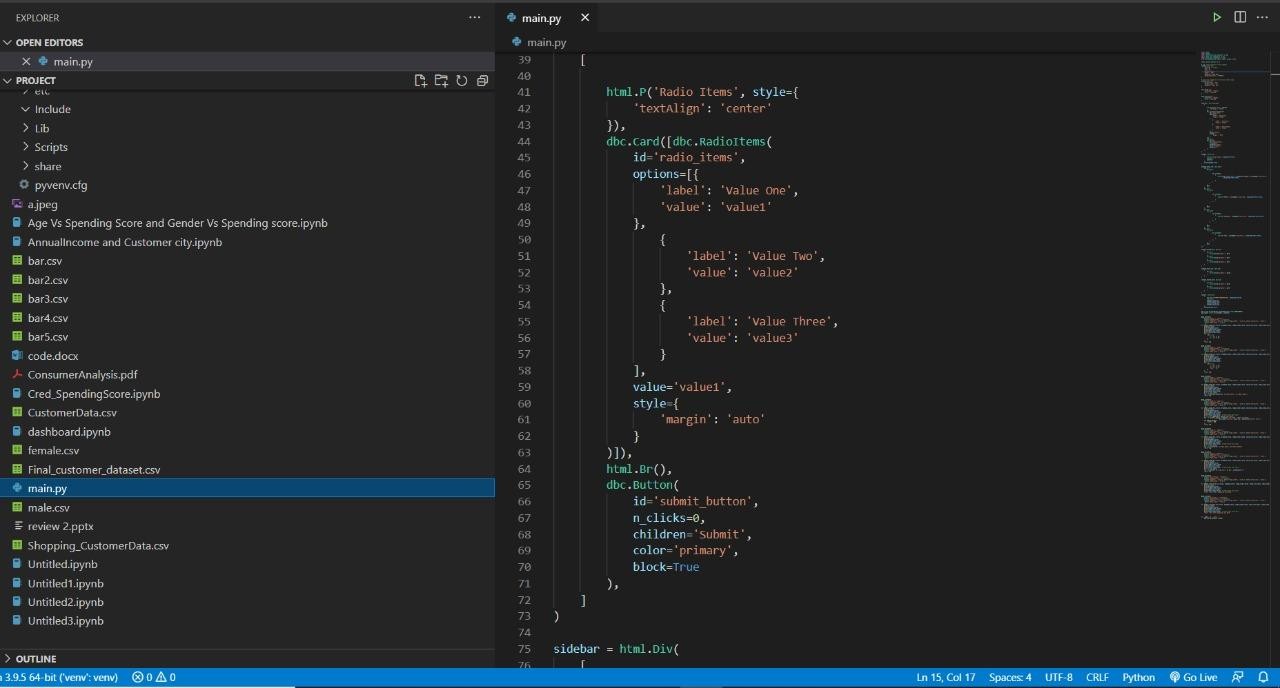
The following dataset have been taken for the project: winequality\_red.csv in notebook



winequality\_red.csv in excel sheet



We have used Jupyter notebook (Python) for the backend and HTML as frontend and set up the necessary environment for the project. The project diary is shown below.



**Libraries used**

Django is a framework that connects the front end with the back end. we have used the Django framework in pycharm as well as Microsoft Visual Studio Code IDE •

The libraries imported for the project are pandas, numpy, matplotlib and seaborn. Matplotlib and seaborn have been imported to visualize the boxplots.

SciPy is a collection of open-source code libraries for maths, science and engineering. NumPy, pandas and Matplotlib are libraries that fall under this umbrella of SciPy.

• Pandas is an open-source Python library providing efficient, easy to use data structure and data analysis tools.

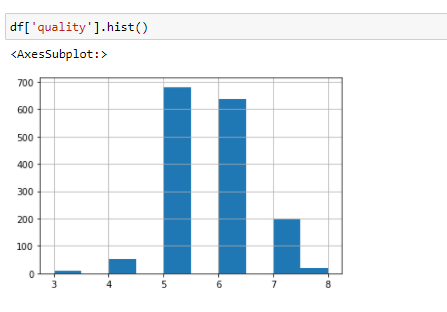
• Matplotlib is a python library that is specifically designers for the development of graphs, charts in order to provide interactive data visualization. It is inspired from MATLAB software and reproduces many of its features.

• The front end has been developed using html, css and JavaScript.

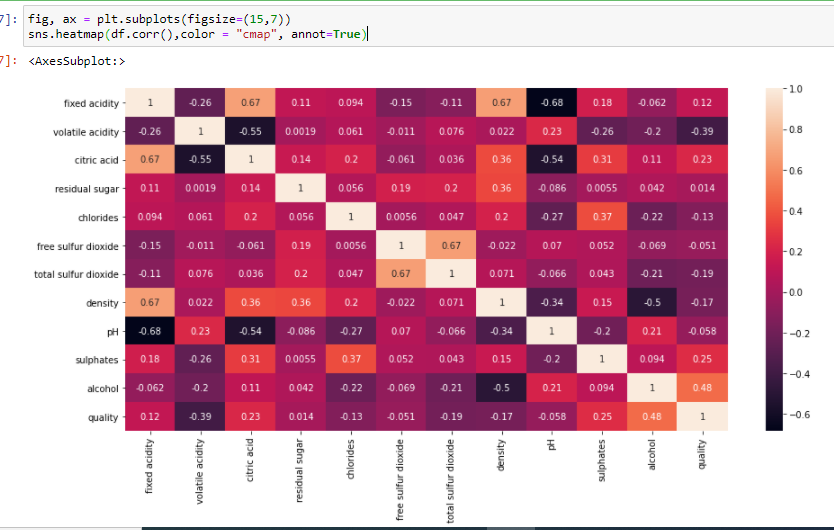
• Google charts has been used on the front end for better visualization.

**Visualizations**

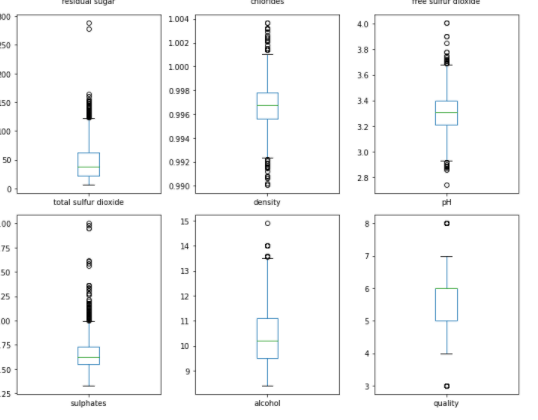
* Histogram of quality vs frequency

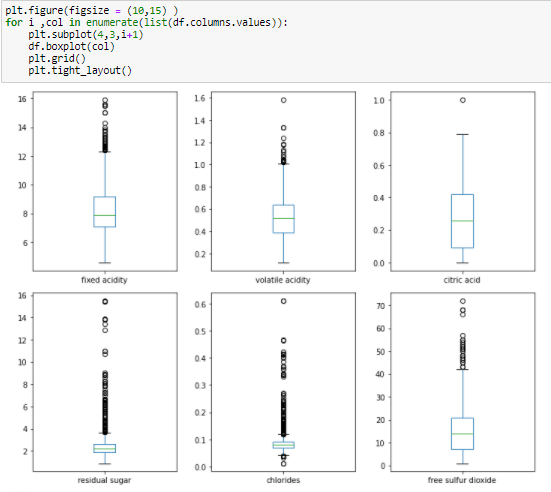


* Plotting heatmap to find out correlation between variables.

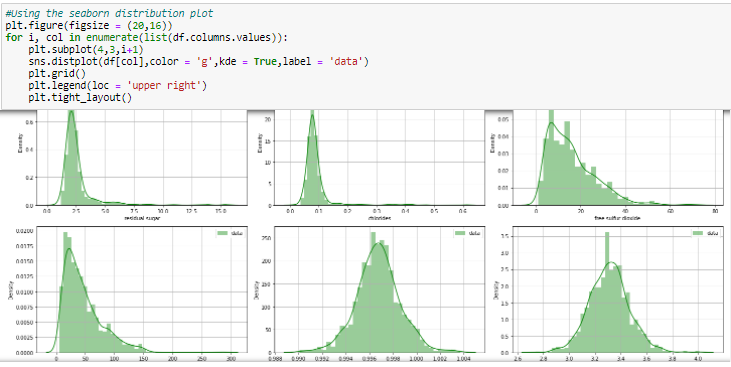


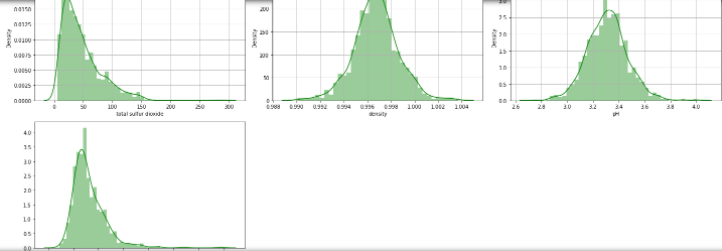
* Box- whisker plots for finding outliers



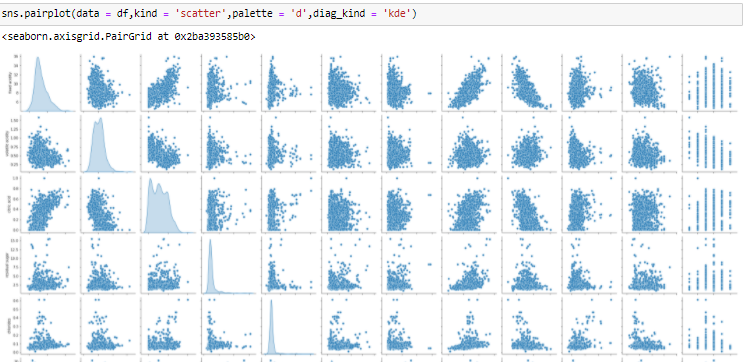


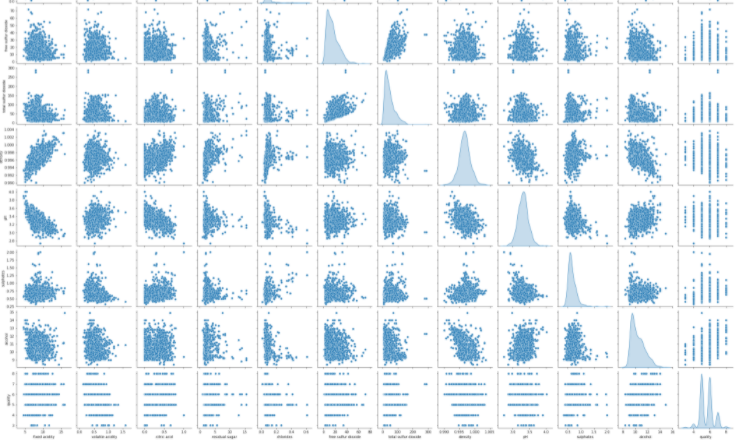
* Probability distribution plot using seaborn





* To understand correlation between variables, we use seaborn to plot graphs.





**OBSERVATIONS AND CONCLUSIONS**

From the above explorations conducted we infer the following.

* The shape of the data is (1599,12) which shows there are 1599 rows and 12 columns.

**For the dataset,**

* Mean value is less than the median value of each column.
* There is a large difference between the 75th% tile and max values of residual sugar, free sulfur dioxide & total sulfur dioxide.
* The variable 'Quality' is very discrete.
* The score ranges from 1 to 10, where 1 is poor and 10 is most acidic (the best).
* **Maximum number of unique values = Density, Minimum number of unique values = Quality**

**For heatmap,**

* Density has a strong positive correlation with residual Sugar. Density has a strong negative correlation with alcohol.
* Free sulphur dioxide and citric acid has almost no correlation with quality.

Help : Dark shades -> Positive Correlation

Light shades -> Negative Correlation

Due to zero correlation, free sulphur dioxide and citric acid can be dropped while chosing variables for linear regression.

**For box whisker plot to find outliers,**

* All the features except for ALCOHOL have outliers

**For probability distribution frequency and pairwise plots,**

* pH feature is approximately normally distributed compared to the others who are skewed

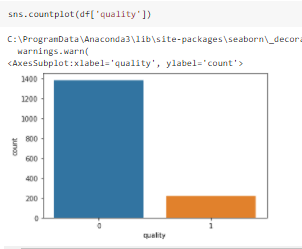
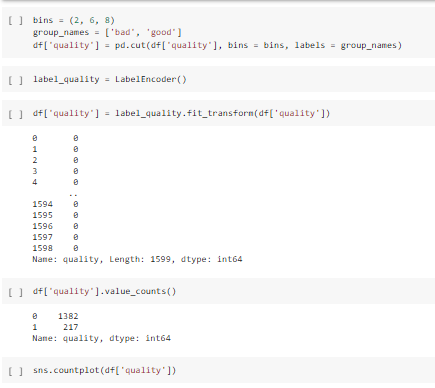
**MACHINE LEARNING ALGORITHM USED**

***Random forest classifier***

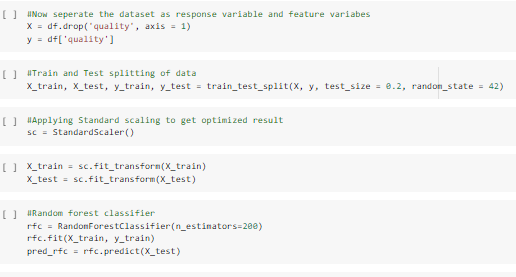
Random forest is a [supervised learning algorithm](https://builtin.com/data-science/supervised-learning-python). The "forest" it builds, is an ensemble of decision trees, usually trained with the “bagging” method. The general idea of the bagging method is that a combination of learning models increases the overall result.

**Put simply: random forest builds multiple decision trees and merges them together to get a more accurate and stable prediction.**

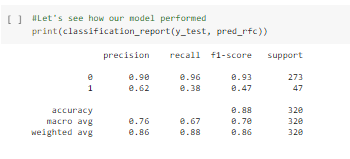
Primary coding before training of data. Data has been put in just two bins: good or bad. There has also been an attempt to count the dataset after plotting the required graphs.



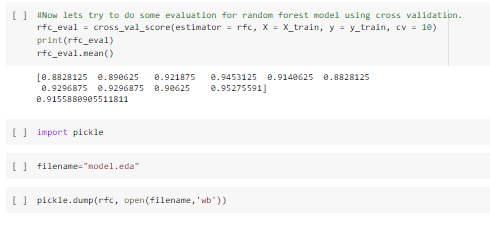
* Training and testing of the dataset



* Evaluating the performance of the model



* And finally we calculate the accuracy of our model



**We observe that the accuracy of our model is 91.55%.**

**Html Code**

<!DOCTYPE html>

<html>

<title>Wine Predictor</title>

<meta charset="UTF-8">

<meta name="viewport" content="width=device-width, initial-scale=1">

<link rel="stylesheet" href="https://www.w3schools.com/w3css/4/w3.css">

<link rel="stylesheet" href="https://fonts.googleapis.com/css?family=Lato">

<link rel="stylesheet" href="https://cdnjs.cloudflare.com/ajax/libs/font-awesome/4.7.0/css/font-awesome.min.css">

<style>

    body,h1,h2,h3,h4,h5,h6 {font-family: "Lato", sans-serif;}

    body, html {

      height: 100%;

      color: #777;

      line-height: 1.8;

    }

    /\* Create a Parallax Effect \*/

    .bgimg-1 {

      background-attachment: fixed;

      background-position: center;

      background-repeat: no-repeat;

      background-size: cover;

    }

    /\* First image (Logo. Full height) \*/

    .bgimg-1 {

      background-image: url('../static/vineyard.jpg');

      min-height: 100%;

    }

    .w3-wide {letter-spacing: 10px;}

    .w3-hover-opacity {cursor: pointer;}

    /\* Turn off parallax scrolling for tablets and phones \*/

    @media only screen and (max-device-width: 1600px) {

      .bgimg-1, .bgimg-2, .bgimg-3 {

        background-attachment: scroll;

        min-height: 600px;

      }

    }

    </style>

   <body>

    <div class="bgimg-1 w3-display-container" id="home">

        <div class="w3-display-middle" style="white-space:nowrap;">

          <span class="w3-center w3-padding-large w3-black w3-xxlarge w3-wide w3-animate-opacity">Wine <span class="w3-hide-small">Quality</span> Predictor</span>

        </div>

      </div>

      <title>Churn Prediction</title>

      <div class="container">

        <div class="row">

          <form action="http://localhost:5000/" method="POST">

            <br>

            <div class="col-sm-9">

             <div class="form-group  purple-border" style="margin:0 16px 8px 16px">

                  <label for="comment">fixed acidity: </label>

                  <input style="margin-left:102px" class="form-control" rows="2" id="query1" name="query1" rows="2" cols="5" style="margin: 10px 16px 2px 16px" autofocus></input>

            </div>

            <div class="form-group  purple-border" style="margin:0 16px 8px 16px">

                  <label for="comment">volatile acidity:</label>

                  <input style="margin-left:85px" class="form-control" rows="2" id="query2" name="query2" rows="2" cols="5" style="margin:0 16px 2px 16px"autofocus></input>

            </div>

            <div class="form-group  purple-border" style="margin:0 16px 8px 16px">

                  <label for="comment">citric acid:</label>

                  <input class="form-control" style="margin-left:118px"  rows="2" id="query3" name="query3" rows="2" cols="5" autofocus></input>

            </div>

            <div class="form-group  purple-border" style="margin:0 16px 8px 16px">

                  <label for="comment">residual sugar:</label>

                  <input class="form-control" style="margin-left:89px"  rows="2" id="query4" name="query4" rows="2" cols="5" autofocus></input>

            </div>

            <div class="form-group  purple-border" style="margin:0 16px 8px 16px">

                  <label for="comment">chlorides:</label>

                  <input class="form-control" style="margin-left:119px"  rows="2" id="query5" name="query5" rows="2" cols="5" autofocus></input>

            </div>

            <div class="form-group  purple-border" style="margin:0 16px 8px 16px">

                  <label for="comment">free sulfur dioxide: </label>

                  <input class="form-control" style="margin-left:60px" rows="2" id="query6" name="query6" rows="2" cols="5" autofocus></input>

            </div>

            <div class="form-group  purple-border" style="margin:0 16px 8px 16px">

                  <label for="comment">total sulfur dioxide:</label>

                  <input class="form-control" style="margin-left:55px"  rows="2" id="query7" name="query7" rows="2" cols="5" autofocus></input>

            </div>

            <div class="form-group  purple-border" style="margin:0 16px 8px 16px">

                  <label for="comment">density:</label>

                  <input class="form-control" style="margin-left:130px"  rows="2" id="query8" name="query8" rows="2" cols="5" autofocus></input>

            </div>

            <div class="form-group  purple-border" style="margin:0 16px 8px 16px">

                  <label for="comment">pH:</label>

                  <input class="form-control" style="margin-left:158px"  rows="2" id="query9" name="query9" rows="2" cols="5"  autofocus></input>

            </div>

            <div class="form-group  purple-border" style="margin:0 16px 8px 16px">

                  <label for="comment">sulphates:</label>

                  <input class="form-control" style="margin-left:115px"  rows="2" id="query10" name="query10" rows="2" cols="5" autofocus></input>

            </div>

            <div class="form-group  purple-border" style="margin:0 16px 8px 16px">

                  <label for="comment">alcohol:</label>

                  <input class="form-control" style="margin-left:130px"  rows="2" type="number" id="query11" name="query11" rows="2" cols="5" autofocus></input>

            </div>

            </div>

            <div class="col-sm-3">

                <button class= "w3-black" type="submit">

                    <i class="fa fa-paper-plane"></i> Submit

                  </button>

            </div>

          </form>

          </div>

        <div class="row">

            <div class="col-sm-9 purple-border" style="margin:0 16px 8px 16px">

                <br>

            <textarea class="form-control " rows="2" id="comment" name="query6" rows="2" cols="50" autofocus>{{output1}}</textarea>

        </button>

            </div>

        </div>

    </div>

   </body>

</html>

PYTHON BACKEND CODE

# coding: utf-8

import pandas as pd

from sklearn.model\_selection import train\_test\_split

from sklearn.ensemble import RandomForestClassifier

from sklearn import metrics

from flask import Flask, request, render\_template

import pickle

app = Flask("\_\_name\_\_")

df\_1=pd.read\_csv("winequality-red (1).csv")

q = ""

@app.route("/")

def loadPage():

    return render\_template('home.html', query="")

@app.route("/", methods=['POST'])

def predict():

    '''

    fixed acidity

    volatile acidity

    citric acid

    residual sugar

    chlorides

    free sulfur dioxide

    total sulfur dioxide

    density

    pH

    sulphates

    alcohol

    '''

    inputQuery1 = request.form['query1']

    inputQuery2 = request.form['query2']

    inputQuery3 = request.form['query3']

    inputQuery4 = request.form['query4']

    inputQuery5 = request.form['query5']

    inputQuery6 = request.form['query6']

    inputQuery7 = request.form['query7']

    inputQuery8 = request.form['query8']

    inputQuery9 = request.form['query9']

    inputQuery10 = request.form['query10']

    inputQuery11 = request.form['query11']

    model = pickle.load(open("model.eda", "rb"))

    data = [[inputQuery1, inputQuery2, inputQuery3, inputQuery4, inputQuery5, inputQuery6, inputQuery7,

             inputQuery8, inputQuery9, inputQuery10, inputQuery11]]

    new\_df = pd.DataFrame(data, columns = ['fixed acidity', 'volatile acidity', 'citric acid', 'residual sugar',

                                           'chlorides', 'free sulfur dioxide', 'total sulfur dioxide', 'density', 'pH',

                                           'sulphates', 'alcohol'])

    single = model.predict(new\_df.tail(1))

    if single==1:

        o1 = "The wine quality is bad"

    else:

        o1 = "The wine quality is good"

 return render\_template('home.html', output1=o1,

                           query1 = request.form['query1'],

                           query2 = request.form['query2'],

                           query3 = request.form['query3'],

                           query4 = request.form['query4'],

                           query5 = request.form['query5'],

                           query6 = request.form['query6'],

                           query7 = request.form['query7'],

                           query8 = request.form['query8'],

                           query9 = request.form['query9'],

                           query10 = request.form['query10'],

                           query11 = request.form['query11'])

app.run()

**Code for Homepage**

<!DOCTYPE html>

<html>

<title>Wine Predictor</title>

<meta charset="UTF-8">

<meta name="viewport" content="width=device-width, initial-scale=1">

<link rel="stylesheet" href="https://www.w3schools.com/w3css/4/w3.css">

<link rel="stylesheet" href="https://fonts.googleapis.com/css?family=Lato">

<link rel="stylesheet" href="https://cdnjs.cloudflare.com/ajax/libs/font-awesome/4.7.0/css/font-awesome.min.css">

<style>

body,h1,h2,h3,h4,h5,h6 {font-family: "Lato", sans-serif;}

body, html {

height: 100%;

color: #777;

line-height: 1.8;

}

/\* Create a Parallax Effect \*/

.bgimg-1 {

background-attachment: fixed;

background-position: center;

background-repeat: no-repeat;

background-size: cover;

}

/\* First image (Logo. Full height) \*/

.bgimg-1 {

background-image: url('../static/vineyard.jpg');

min-height: 100%;

}

.w3-wide {letter-spacing: 10px;}

.w3-hover-opacity {cursor: pointer;}

/\* Turn off parallax scrolling for tablets and phones \*/

@media only screen and (max-device-width: 1600px) {

.bgimg-1, .bgimg-2, .bgimg-3 {

background-attachment: scroll;

min-height: 600px;

}

}

</style>

<body>

<div class="bgimg-1 w3-display-container" id="home">

<div class="w3-display-middle" style="white-space:nowrap;">

<span class="w3-center w3-padding-large w3-black w3-xxlarge w3-wide w3-animate-opacity">Wine <span class="w3-hide-small">Quality</span> Predictor</span>

</div>

</div>

<title>Churn Prediction</title>

<div class="container">

<div class="row">

<form action="http://localhost:5000/" method="POST">

<br>

<div class="col-sm-9">

<div class="form-group purple-border" style="margin:0 16px 8px 16px">

<label for="comment">fixed acidity: </label>

<input style="margin-left:102px" class="form-control" rows="2" id="query1" name="query1" rows="2" cols="5" style="margin: 10px 16px 2px 16px" autofocus></input>

</div>

<div class="form-group purple-border" style="margin:0 16px 8px 16px">

<label for="comment">volatile acidity:</label>

<input style="margin-left:85px" class="form-control" rows="2" id="query2" name="query2" rows="2" cols="5" style="margin:0 16px 2px 16px"autofocus></input>

</div>

<div class="form-group purple-border" style="margin:0 16px 8px 16px">

<label for="comment">citric acid:</label>

<input class="form-control" style="margin-left:118px" rows="2" id="query3" name="query3" rows="2" cols="5" autofocus></input>

</div>

<div class="form-group purple-border" style="margin:0 16px 8px 16px">

<label for="comment">residual sugar:</label>

<input class="form-control" style="margin-left:89px" rows="2" id="query4" name="query4" rows="2" cols="5" autofocus></input>

</div>

<div class="form-group purple-border" style="margin:0 16px 8px 16px">

<label for="comment">chlorides:</label>

<input class="form-control" style="margin-left:119px" rows="2" id="query5" name="query5" rows="2" cols="5" autofocus></input>

</div>

<div class="form-group purple-border" style="margin:0 16px 8px 16px">

<label for="comment">free sulfur dioxide: </label>

<input class="form-control" style="margin-left:60px" rows="2" id="query6" name="query6" rows="2" cols="5" autofocus></input>

</div>

<div class="form-group purple-border" style="margin:0 16px 8px 16px">

<label for="comment">total sulfur dioxide:</label>

<input class="form-control" style="margin-left:55px" rows="2" id="query7" name="query7" rows="2" cols="5" autofocus></input>

</div>

<div class="form-group purple-border" style="margin:0 16px 8px 16px">

<label for="comment">density:</label>

<input class="form-control" style="margin-left:130px" rows="2" id="query8" name="query8" rows="2" cols="5" autofocus></input>

</div>

<div class="form-group purple-border" style="margin:0 16px 8px 16px">

<label for="comment">pH:</label>

<input class="form-control" style="margin-left:158px" rows="2" id="query9" name="query9" rows="2" cols="5" autofocus></input>

</div>

<div class="form-group purple-border" style="margin:0 16px 8px 16px">

<label for="comment">sulphates:</label>

<input class="form-control" style="margin-left:115px" rows="2" id="query10" name="query10" rows="2" cols="5" autofocus></input>

</div>

<div class="form-group purple-border" style="margin:0 16px 8px 16px">

<label for="comment">alcohol:</label>

<input class="form-control" style="margin-left:130px" rows="2" type="number" id="query11" name="query11" rows="2" cols="5" autofocus></input>

</div>

</div>

<div class="col-sm-3">

<button class= "w3-black" type="submit">

<i class="fa fa-paper-plane"></i> Submit

</button>

</div>

</form>

</div>

<div class="row">

<div class="col-sm-9 purple-border" style="margin:0 16px 8px 16px">

<br>

<textarea class="form-control " rows="2" id="comment" name="query6" rows="2" cols="50" autofocus>{{output1}}</textarea>

</button>

</div>

</div>

</div>

</body>

</html>

**Screenshot of Output Homepage**



**CLOSING REMARKS**

The project has been implemented according to the understanding and requirements of all exploratory data analysis methods and techniques using the necessary coding, and software and the output is displayed.

**Team 10**

**Harshini K Aiyyer (19MIA1050)**

**Gaurav Trivedi (19MIA1077)**