1932022 - Mahashruthi K B
1932055 - Vandana A

Coimbatore Institute of Technology

WINE REVIEWS AND RATINGS

PROBLEM DESCRIPTION

The aim of this project is to analyse the data of winery details collected from all over the world containing details of winery, location, variety, price and ratings. Our goal is to provide a platform where wine enthusiasts can share their experiences and rate the wines they have tasted, helping others discover new wines and wineries to try. User can search for wineries and wines by location, price, rating, and other criteria, and read reviews and ratings from other users. Users can also add their reviews and ratings to the portal, which will be further visualised using various charts and filters

QUESTIONS

1. Do wine ratings correlate with price?
2. Compare popular varietals from the top 5 wine-producing countries
3. Top 20 highest-rated wines based on price range and varietal
4. Average rating for the most popular varietal produced by each region:

TECHNOLOGIES USED

- rStudio
- MySQL workbench
- Javascript

7 STEPS VISUALISATION TECHNIQUES

1. Acquire

The dataset is sourced from Kaggle website which has reviews and ratings for various winery with varieties and origin.

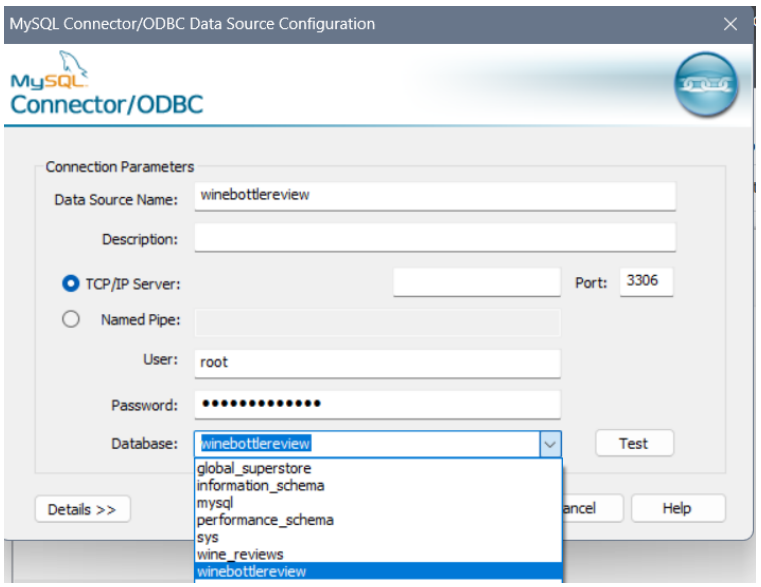
Columns included:

1. Winery_id : Unique key to store the data entered by the user
2. Winery name : Place of wine production
3. Variety: Type of wine produced
4. Designation: The vineyard within the winery where the grapes that made the wine are from
5. Year: Production year
6. Country: The country that the wine is from
7. Province: States and districts of the country where its produced
8. Region_1: The wine growing area in a province or state
9. Region_2: More specific regions specified within a wine growing area
10. Price: Cost for a bottle of the wine
11. Ratings : oOn a scale of 1-5 from excellent to poor based of quality.
12. Points : The number of points WineEnthusiast rated the wine on a scale of 1-100

2. Parse

The parse step is used to provide some structure for the data's meaning, and order it into categories. **MySQL** database connection has been used for this purpose.

The data is loaded into schema and stored in tables and read into RShiny website using **RODBC** connection



3. Filter

In the filter step, only the data of interest has been filtered and the remaining fields can be filtered out. In this dataset, entries with null values were removed. The 'title' column is formed from concatination of winery, variety, different components of location in MySQL Workbench.

```
5 •  ALTER TABLE wine.data MODIFY title varchar(200) DEFAULT ( concat(winery,' ',designation,' ',year,' ',variety,' ','(', region_1,')' ));
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

winery_id	title	winery	variety	designation	year	country	province	region_1
E1	Quinta dos Avidagos 2011 Avidagos Red (Douro)	Quinta dos Avidagos	Portuguese Red	Avidagos	2011	Portugal	Douro	NA
E2	Rainstorm 2013 Pinot Gris (Willamette Valley)	Rainstorm	Pinot Gris	NA	2013	US	Oregon	Willamette Va
E3	St. Julian 2013 Reserve Late Harvest Riesling (L...	St. Julian	Riesling	Reserve Late Harvest	2013	US	Michigan	Lake Michigar
E4	Sweet Cheeks 2012 Vintner's Reserve Wild Chil...	Sweet Cheeks	Pinot Noir	Vintner's Reserve Wild Child Block	2012	US	Oregon	Willamette Va
E5	Tandem 2011 Ars In Vitro Tempranillo-Merlot (N...	Tandem	Tempranillo-Merlot	Ars In Vitro	2011	Spain	Northern Spain	Navarra
E6	Terre di Giurfo 2013 Belsito Frappato (Vittoria)	Terre di Giurfo	Frappato	Belsito	2013	Italy	Sicily & Sardinia	Vittoria

4. Mine

This method is used to apply methods from statistics or data mining as a way to discern patterns or place the data in

mathematical context. Here, **linear regression model** has been used to fit a linear model to the data. A model is built to predict whether ratings correlate with price or not. Additionally, the mean and median values for ratings and price fields have also been calculated for further processes.

CODE :

```
wine_df = wine_df %>%
  mutate(price_range = case_when(price <= 10 ~ 1,
      price <=25 & price > 10 ~ 2, price <=50 & price > 25 ~
3,
     price <=100 & price > 50 ~ 4, price <=500 & price >100 ~
     5, price > 500 ~ 6))
y_int = 78.978654
m = 6.556461
wine_df = wine_df %>%
  mutate(Rating = case_when(ave_score <= 82 ~ "Acceptable",
                ave_score <= 86 & ave_score >= 83 ~ "Good",
                    ave_score <= 89 & ave_score >= 87 ~ "Very
Good",
              ave_score <= 93 & ave_score >= 90 ~ "Excellent",
             ave_score <= 97 & ave_score >= 94 ~ "Superb",
            ave_score >= 98 ~ "Classic")) %>%
  mutate(model_score = y_int + m * log10(price)) %>%
  mutate(Value = case_when(ave_score > (+model_score + 2.5) ~
"Good Value"))
```

5. Represent

Various visual representation techniques are chosen to represent the data and derive meaningful insights and patterns.

CHARTS USED :

- Regression Plot - correlation between price and ratings
- Horizontal Bar chart - varietals of top 5 countries

- HTML maps - average ratings in each region by countries and provinces

CODE :

```
output$barGraph2 <- renderPlot ({
    wbar2 = wine_df2 %>% filter(country %in%
wine_countries[1:5], variety %in% input$pickVar) %>%
      group_by(country,variety) %>% summarise(mean_rating =
mean(ave_score)) ggplot(wbar2, aes(x=variety, y=mean_rating))
+ geom_col(position="dodge", aes(fill=country)) +
        scale_y_continuous(limits=c(80,95), oob =
rescale_none) + coord_flip() +
    labs(title = "Average Varietal Ratings by Country")# }})
```

6. Refine

Refining of the graphs is done to improve the basic representation of the charts and make it clearer and more visually engaging. For example, to show the average ratings around the world, HTML maps are used to be more visually appealing and hovering on maps will display the top variety produced and its rating.

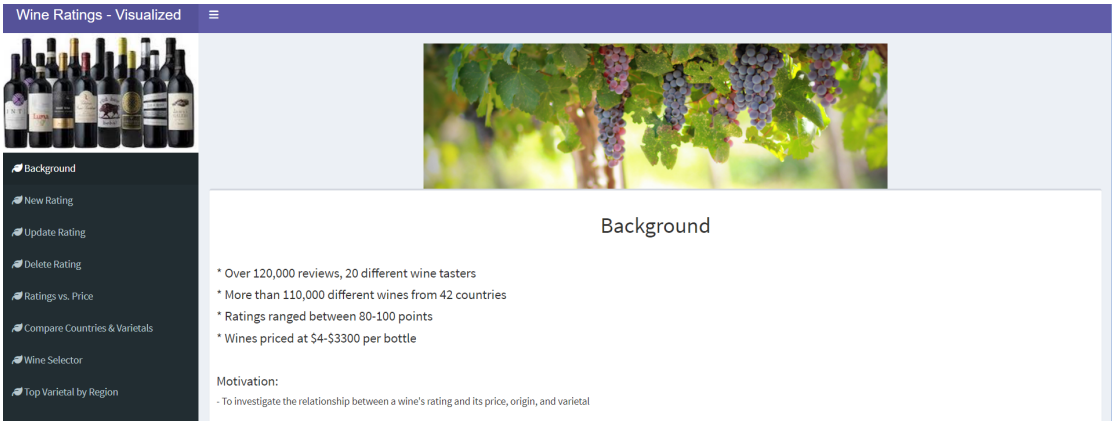
7. Interact

Interact is used to Add methods for manipulating the data or controlling what features are visible. The following function have been implemented to manipulate data in the backend “**RODBC**”

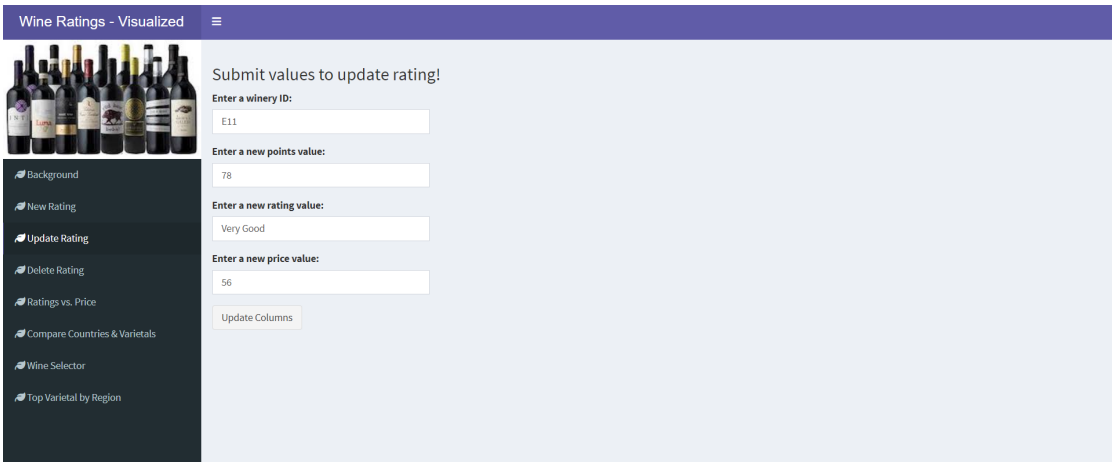
- INSERT
- UPDATE
- DELETE
- FILTER
- SEARCH

OUTPUT

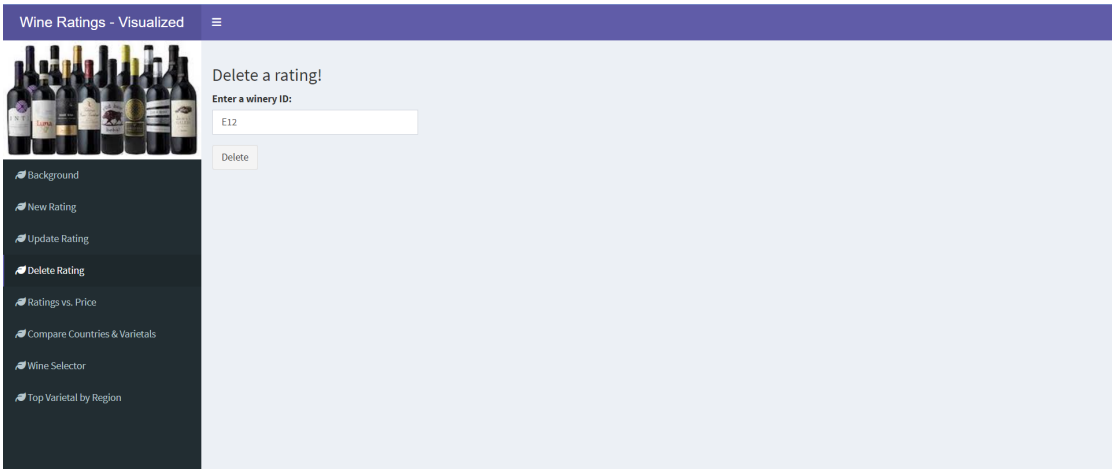
Home page



Update ratings and price with winery_id

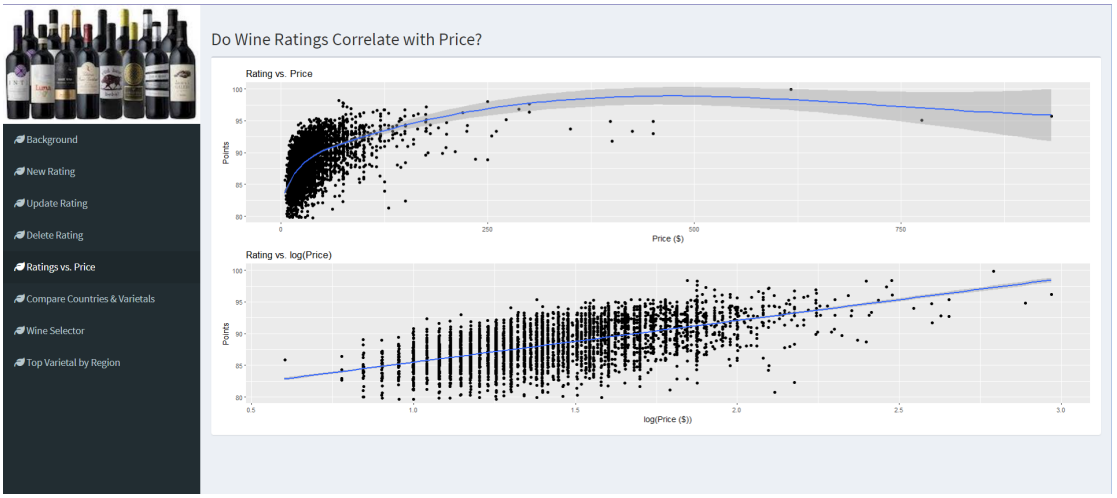


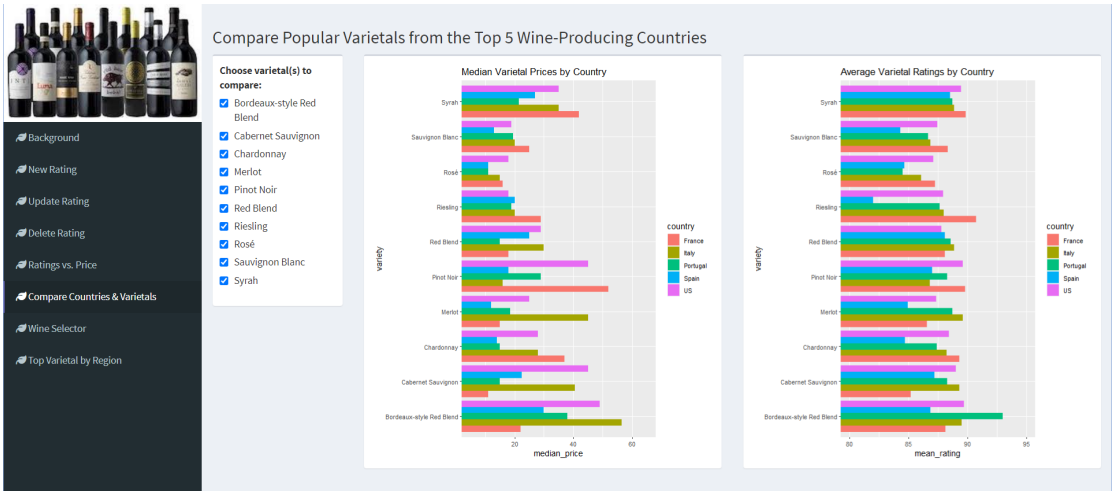
Delete entire row based on winery_id

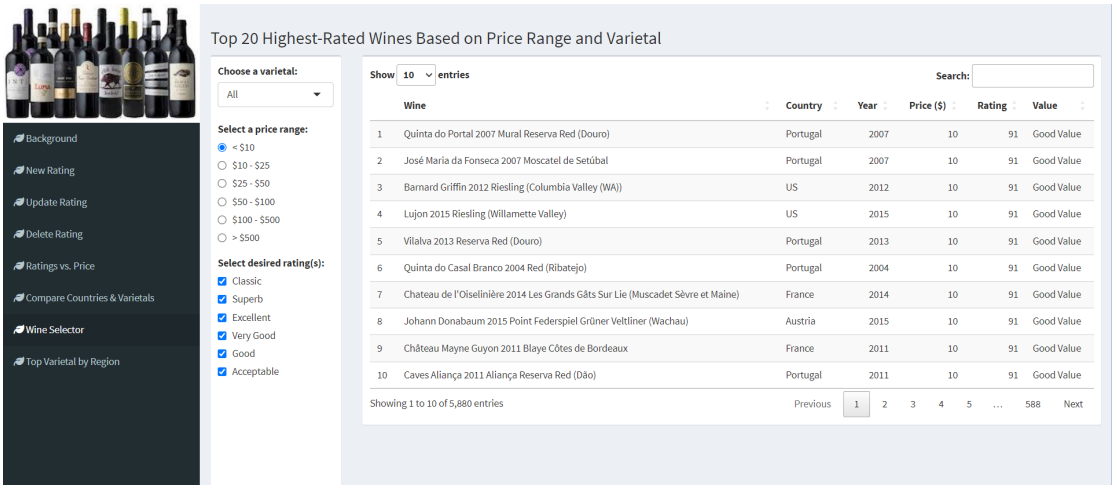


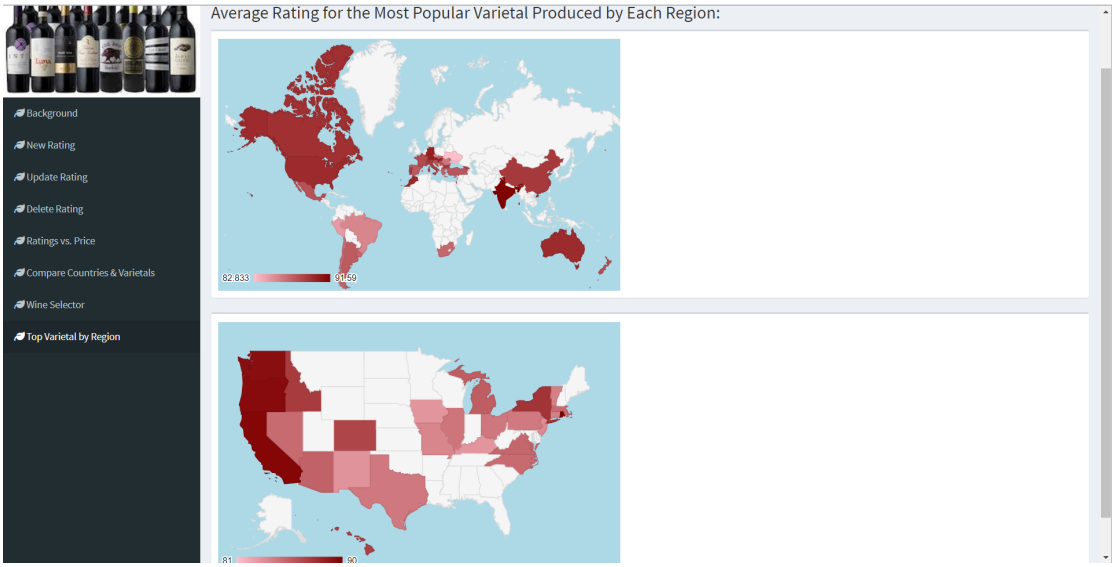
Visualisations

1. Do wine ratings correlate with price
2. Comparing popular varietals from top 5 wine producing countries
3. Top highest rated wines based on price range and variety
4. Average rating based on variety in countries and province









INFERENCE

From the project, we were able to find the relation between ratings and prices of wine based on points inputted. Top 5 countries were shortlisted to find the sales of wine varieties. Top 20 wines were filtered out through filters and search based on price range and ratings. Overall wine ratings over the world were visualised on map were color coordination was given for different rating categories and hovering will give us the average ratings with variety. These visualizations will help the user to buy the best wines in a reasonable price.