

Uncorking Insights: Exploring the World of Wine through Data Visualizations

Matt Moersen mpmoers@clemson.edu C54761488
Akhila Nangineedi anangin@clemson.edu C32054414
Mohammad Alshurbaji malshur@clemson.edu

Repository: https://github.com/sumanthnangineedi97/CPSC6030_Project.git

Basic Info

This dataset encompasses nearly 130,000 wine reviews from Wine Enthusiast, serving as a valuable resource for training an AI model. Much like a master sommelier, a well-trained AI model can leverage descriptions of unfamiliar wines provided by tasters to make educated predictions about their production year, likely region, and winery. This dataset is likely employed for the analysis of the wine industry, as well as for the examination of reviews and ratings. It offers a wealth of information about diverse wines, their characteristics, and the environments in which they are produced and reviewed.

Background and Motivation

The motivation for selecting this dataset stems from a curiosity about wine industry analysis, given its status as a popular and dynamic market. Additionally, the dataset offers insights into the various wine varieties and their prices from around the world. The dataset presents several intriguing problem statements that can be addressed through a variety of data visualizations, including maps, bar charts, and treemaps. These factors collectively drove the decision to choose this dataset.

Project Objectives

Question-1: How many varieties and wines are there in each province of a chosen country?

Learning-1: To determine the quantity of varieties and wines in each province and visually highlight the provinces with a higher number of varieties.

Question-2: Is there a correlation between the price and ratings of wines (based on variety) in a chosen country?

Learning-2: To determine whether the price of wine is influenced by its rating.

Question-3: Which wineries are the industry leaders in a specific country?

Learning-3: To identify the top wineries in a chosen country that produce various types of wines and create a visual comparison based on the count of wines they produce.

Data

The Wine Reviews dataset was downloaded from Kaggle.com (<https://www.kaggle.com/datasets/zynicide/wine-reviews>). The data in the dataset (winemag-data-130k-v2.csv) was collected from wineenthusiast.com in November 2017. It contains 129,971 reviews on the website

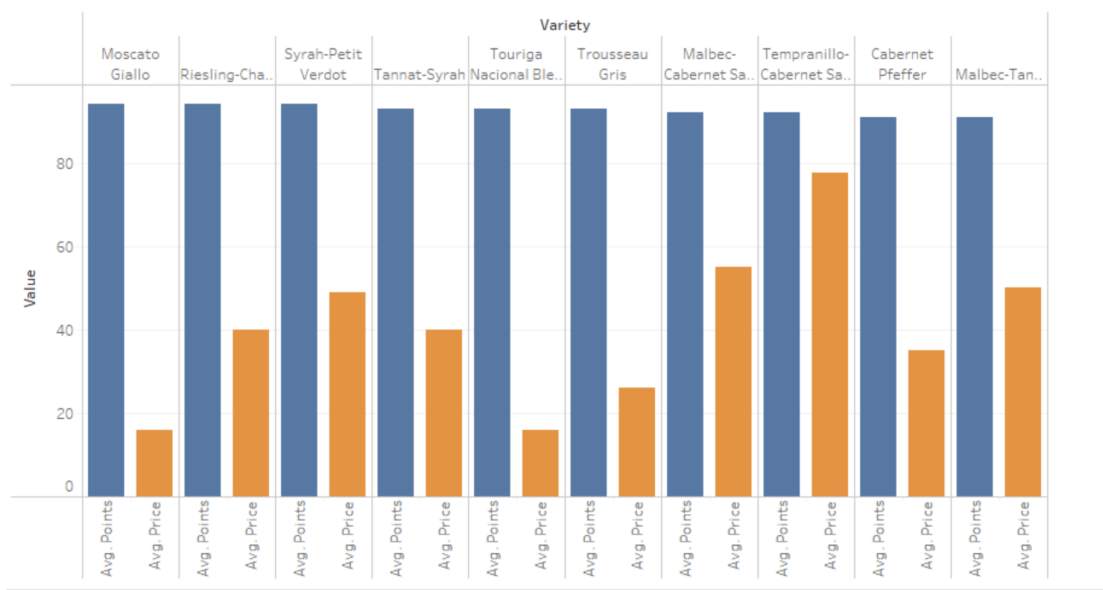
Data Processing

- Selected only the following columns in the dataset: country, points, price, province, taster_name, title, variety, and winery.
- Eliminated records with NULL values in the country, price, and taster_name columns.
- Removed duplicate records.
- Will likely ignore the taster_name column in future.

Visualization Design

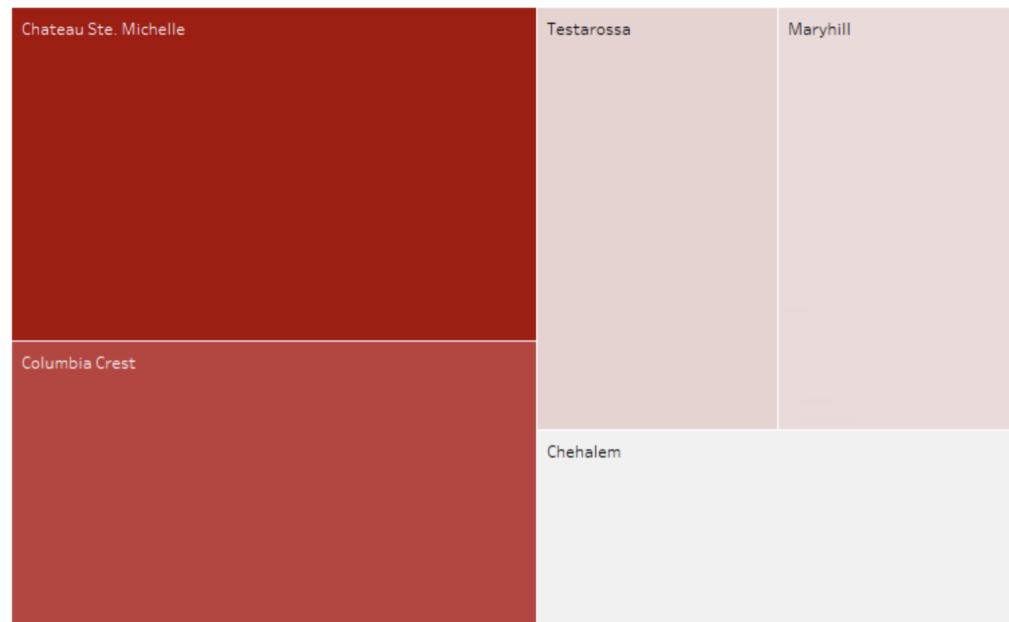
- DV-1: Heat Map displaying Count of wines and varieties of all provinces in a selected country.

- Correlation b/w Top 10 varieties with high avg point



- DV-3: Treemap displaying top 5 wineries in a selected country.

Top 5 leading wineries



Must-Have Features

Title: Provide a title for the chart that summarizes the purpose of the visualization and the context of the data being presented.

Legends: Use a legend to explain the color-coding scheme and any symbols or patterns used within the charts. This helps users understand the meaning of colors and shapes.

Labeling: Label elements with the name or label of the category or item it represents. These labels should be clear, concise, and positioned for easy readability.

Axis Labels: Label both the X-axis and Y-axis with clear and descriptive titles. The X-axis should describe the categories or groups, and the Y-axis should indicate the unit of measurement.

Interactivity: Depending on the platform or context, consider adding interactive features such as tooltips that provide additional information when users hover over or click on the bars.

Filter: To make all three visualizations interactive, a filter will be added to allow the selection of a country, which will modify the data in all three visualizations.

Hover Effects: Adding interactive features like tooltips that appear when you hover over cells can make the charts more user-friendly. These tooltips can display additional information about the data in each cell.

Optional Features

Legend Customization: Allow users to customize the legend to show relevant information, such as category names, values, or color scales.

Filtering and Sorting: Add interactive filters to allow users to focus on specific elements or criteria within the charts. Implement sorting options for data reordering.

Search: Allowing the users to search for countries in map visualization.

Project Schedule

Process	Period	Members
Data Processing, Basic Website Design(HTML + CSS), Data Visualization (Tableau	Week 1 (Oct. 7 - Oct. 13)	Mohammad, Matt, Akhila
Implement the Data Vis in the website using D3JS (JS, CSS, HTML)	Week 2(Oct. 14 - Oct. 20)	Mohammad, Matt, Akhila
Enhance the Website using the D3JS, and make the user interface more easier for the user	Week 3(Oct. 21 - Oct. 27)	Mohammad, Matt, Akhila
Final Touches	Week 4(Oct. 28 - Nov. 5)	Mohammad, Matt, Akhila