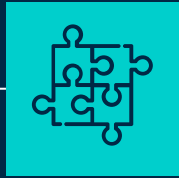


FCA

WINE ENTHUSIAST ANALYSIS



01

INTRODUCTION

- Introducing questions
- Importance of Data

TABLE OF CONTENTS



02

QUESTIONS & ANALYSIS

- Addressing tasks
- Visualizing results



03

CONCLUSIONS

- Improvement of results
- Summary

INTRODUCTION

1. Which vineyard produces the best wine?
2. If you had to pick three wines to recommend using only this data, which would you pick?
3. Which factors (price, country, variety, taster) are most important in determining the overall score of a wine?
4. If I typically enjoy a wine that is “dry” and has “citrus” flavours, which variety of wine should I try?

Analysis to be conducted using python to work on my skillset outside of R & SQL



WHICH VINEYARD PRODUCES THE BEST WINE?

WHAT ARE WE LOOKING FOR?

- Vineyards that, when grouped, return the highest amount of mean points and a low standard deviation

Vineyard name	Times appeared in dataset	Mean points
Cerretalto	1	100
Cristal Vintage Brut	1	100
Barca-Velha	2	99.5
Clos du Mesnil Brut Blanc de Blancs	1	99
Cuvée Constance	1	99
CÃ d'Morissio Riserva	1	99
Elevation 1147 Estate	1	99
Nacional Vintage	2	99
Rare Vineyards	2	99
Sorì Tildin	1	99



KEY POINTS:

- Very low number of observations
 - Questions consistency
 - No standard deviation for single observations

Table 1.1: Output of the top 10 vineyards with the highest mean points and their reciprocal number of times they appear in the database.

REFINING THESE RESULTS FURTHER

HOW?

- Looking for vineyards with observations > 5
- Observing standard deviations to account for the measure of spread
- Understanding the variety of wines the vineyard produces to understand their flexibility in the industry

Vineyard name	Times appeared in dataset	Mean points (1 d.p)	Standard deviation (1 d.p)	Number of varieties
Sassicaia	7	96.4	1.6	1
Kiedricher Grafenberg Riesling Trockenbeerenauslese	6	96.3	1.2	1
Bionic Frog	6	95.5	2.4	1
Clos Saint Urbain Rangen de Thann Grand Cru	11	95.4	0.9	3
En Cerise Vineyard	6	95.2	2.2	1

Table 1.2: Output of the top 5 vineyards with the highest mean points that appear more than five times in the database, their reciprocal mean points, standard deviation, and number of varieties of wines they produce

Best for points considering consistency: Sassicaia

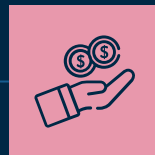
Best vineyard for producing a variety of wines: Clos Saint Urbain Rangen de Thann Grand Cru

PICKING 3 WINES TO RECOMMEND

WHAT FACTORS AM I LOOKING FOR?

COUNTRY

A top-tier country of origin, to experience great tastes from all over

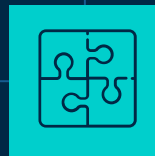


PRICE

A wine that isn't too expensive

POINTS

Wines that score highly, to exceed expectations



VARIETY

3 wines of different varieties to give a range of insight

PICKING 3 WINES TO RECOMMEND

COUNTRIES

- Splitting countries into five tiers by their mean points
- Allows me to refine by only top-tier countries

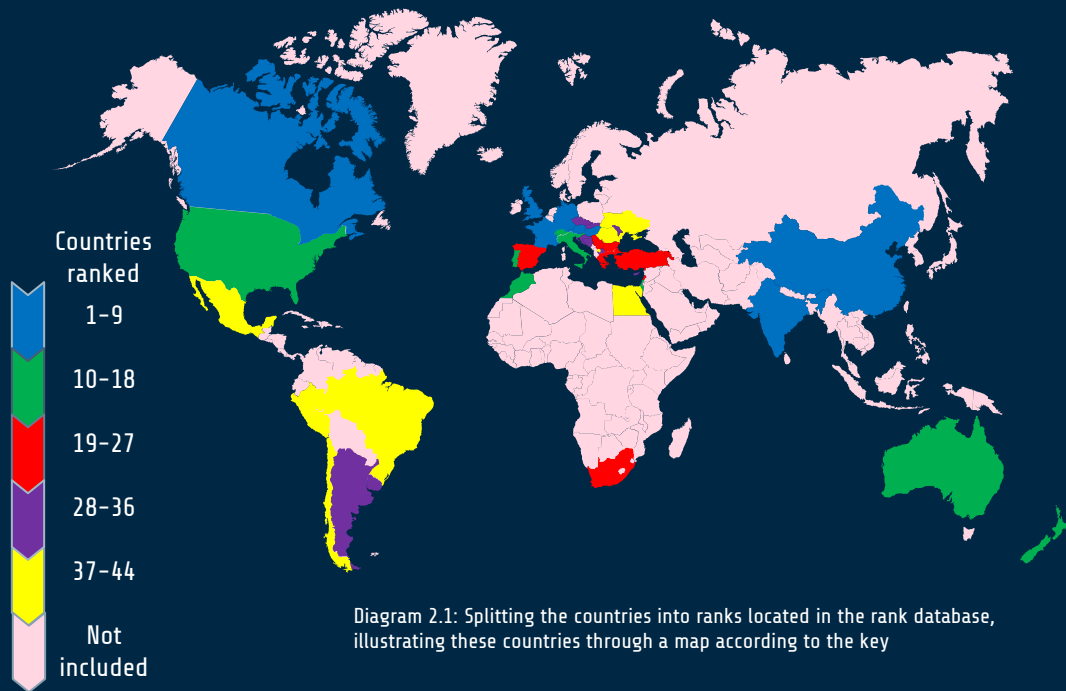


Diagram 2.1: Splitting the countries into ranks located in the rank database, illustrating these countries through a map according to the key

- The rank database in CSV form:
bestcountryq2results.csv

PICKING 3 WINES TO RECOMMEND

ACHIEVING THESE RESULTS

- Looking for wines from a top-tier country
- Looking for wines with a price <£50
- Looking for wines from different varieties
- Looking for wines with points >95

From this result, I could choose three wines which were all of different varieties and different countries of origin, that were top performers in the dataset while remaining true to the budget requirements I set.

PLOT OF PRICE & POINTS BASED ON GIVEN REQUIREMENTS

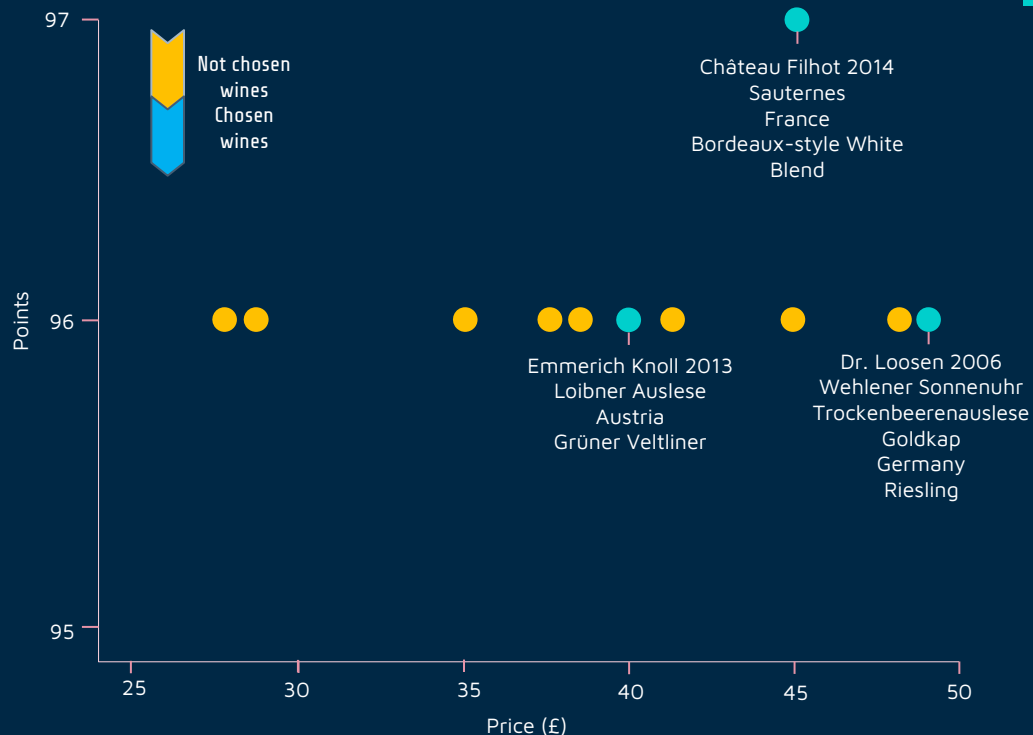


Figure 2.2: Plotting the price on the x-axis and points on the y-axis for the wines that satisfy the conditions: being from a top-tier country, price less than £50, and points greater than 95. The wines labelled fit the conditions that they are of different varieties and from different countries

WHICH FACTORS ARE MOST IMPORTANT IN DETERMINING THE SCORE FOR A WINE?

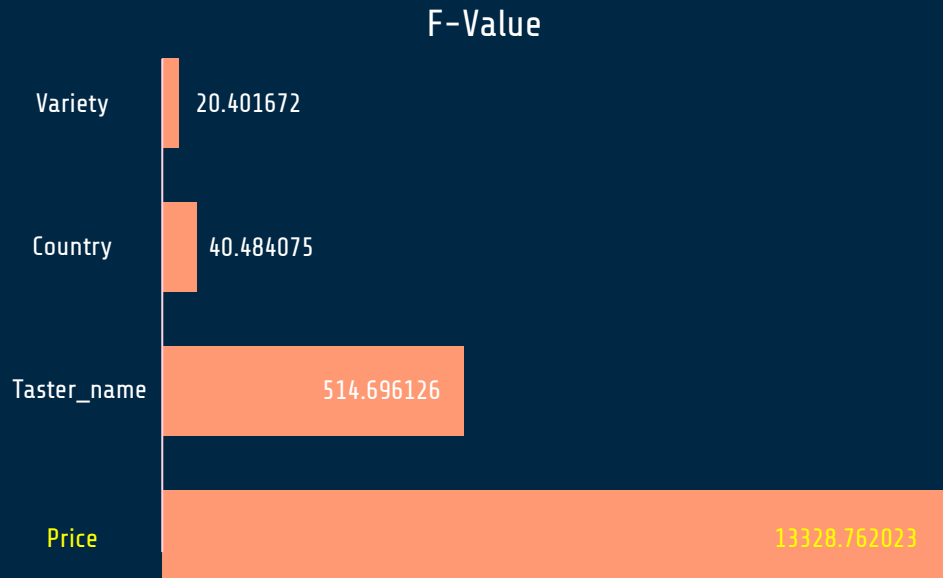


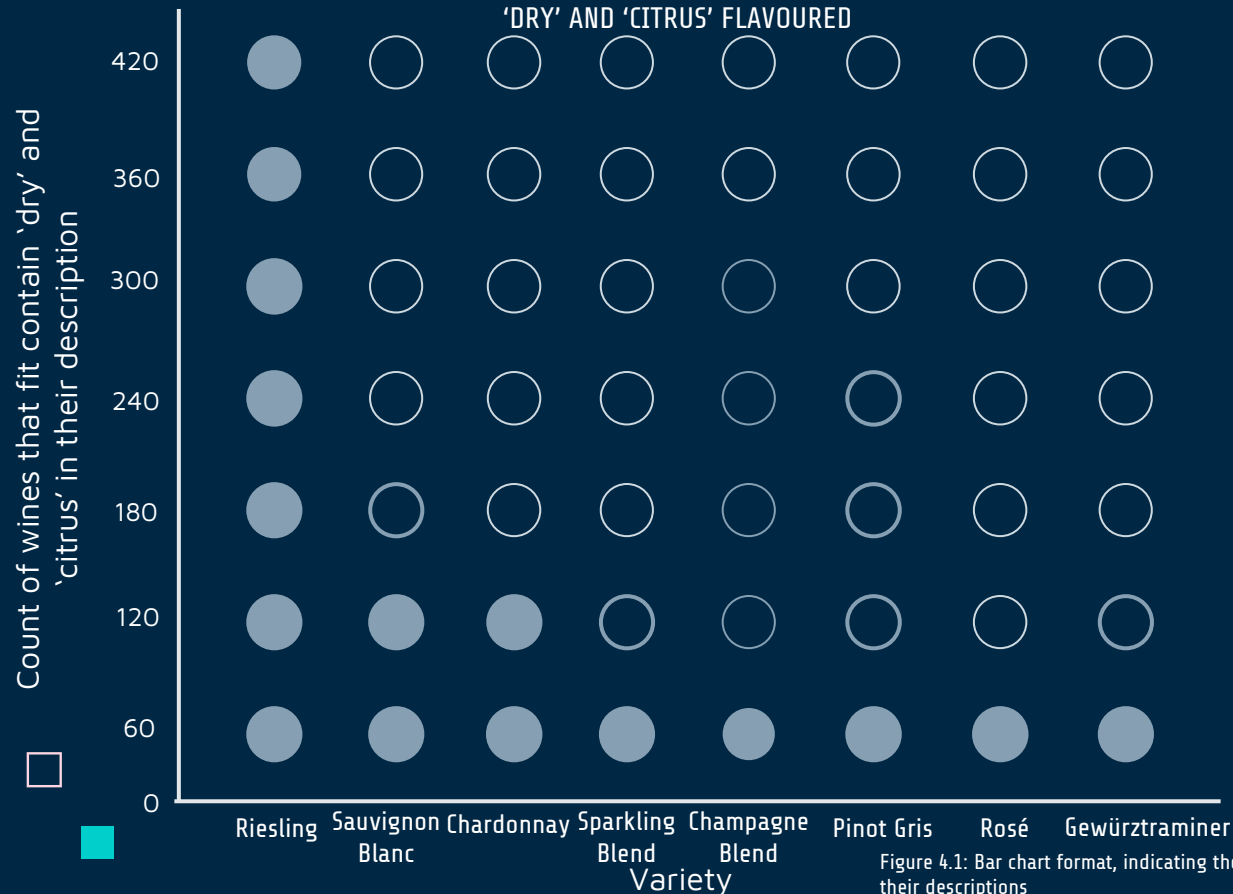
Figure 3.1: Shows the F values for variety, country, taster_name, and price. These values indicate their significance in determining the points for a wine

UTILIZING AN ANOVA TEST

- ANOVA model is used to determine which factors (price, variety, taster_name, country) are significant in determining the number of points a wine has.
- The ANOVA model compares the means of the response variable (points) for different levels of each factor to determine if there are significant differences between those means.
- The ANOVA model produces an F-value and p-value for each factor, which allows you to determine which factor is most significant in determining the points for a wine.

DRY & CITRUS VARIETY OF WINE

BAR CHART SHOWING THE NUMBER OF WINES OF EACH VARIETY THAT ARE
'DRY' AND 'CITRUS' FLAVOURED



APPROACH

- Filter wines by descriptions including 'citrus' and 'dry'
- Group by variety & count the number in each variety

Clear majority in favour of the Riesling variety

Figure 4.1: Bar chart format, indicating the varieties of wine that yield the highest count of 'dry' and 'citrus' strings in their descriptions

CONCLUSIONS

- Utilising machine learning to improve results
 - Supervised & unsupervised learning techniques
- Although the dataset is extensive, data could be collected in a more consistent manner resulting in less cleaning needing to be done
 - Also, enhancing the dataset by vineyards collecting data on assets like the wine's pH level, alcohol content, soil composition, and more could lead to more specific insights



TOM DAVIES