



Best Value Wine

Steve Williamson

IBM Capstone Applied Data Science Final Project



Business Problem

- The wine trade is a very competitive and complex because:
 - Consumers have many choices for purchasing wine:
 - Supermarket
 - Specialist wine store
 - Subscription wine clubs
 - The price variance for a bottle of wine is large (£4 to over £1,000 a bottle)
 - The Wine Producer network is extensive:
 - Approximately 7 million wine producers exist worldwide
- New wine retailers need to find a competitive edge
- Through Data Science, identify wineries, which produce **high quality** wine, and sell it at **low cost** (relative to other, better known wine brands)

Data Sources

Data Source	Description and Purpose
Wine Ratings	<p>A data set of 129,000 wines, which have been rated by expert wine tasters. Also includes price, country and region of production.</p> <p>This will be used for statistical analysis.</p>
Google Geo-location data	<p>Google API which provides Geo-coding data (i.e. latitude and longitude) based on an address or region.</p> <p>This will be used to get the geo-coordinates for each high quality wine region.</p>
Four Square Venues	<p>Contains details of a range of venues. A Winery is a type of venue.</p> <p>This will be used to discover new wineries and extract relevant details (i.e. rating and url)</p>

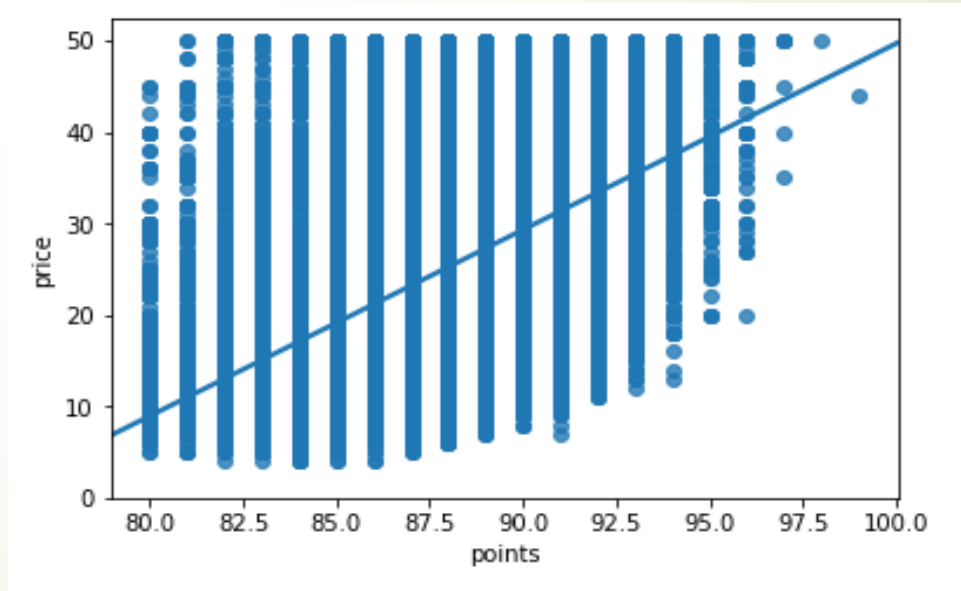
Summary of Statistical Analysis

- Weak positive correlation between Price and Quality:
 - Quality is measured by Points (as awarded by wine taster)
- Outliers (very expensive wines) and any wine over £50 was removed from the data set

```
#Look again at correlation  
df_wine.corr()
```

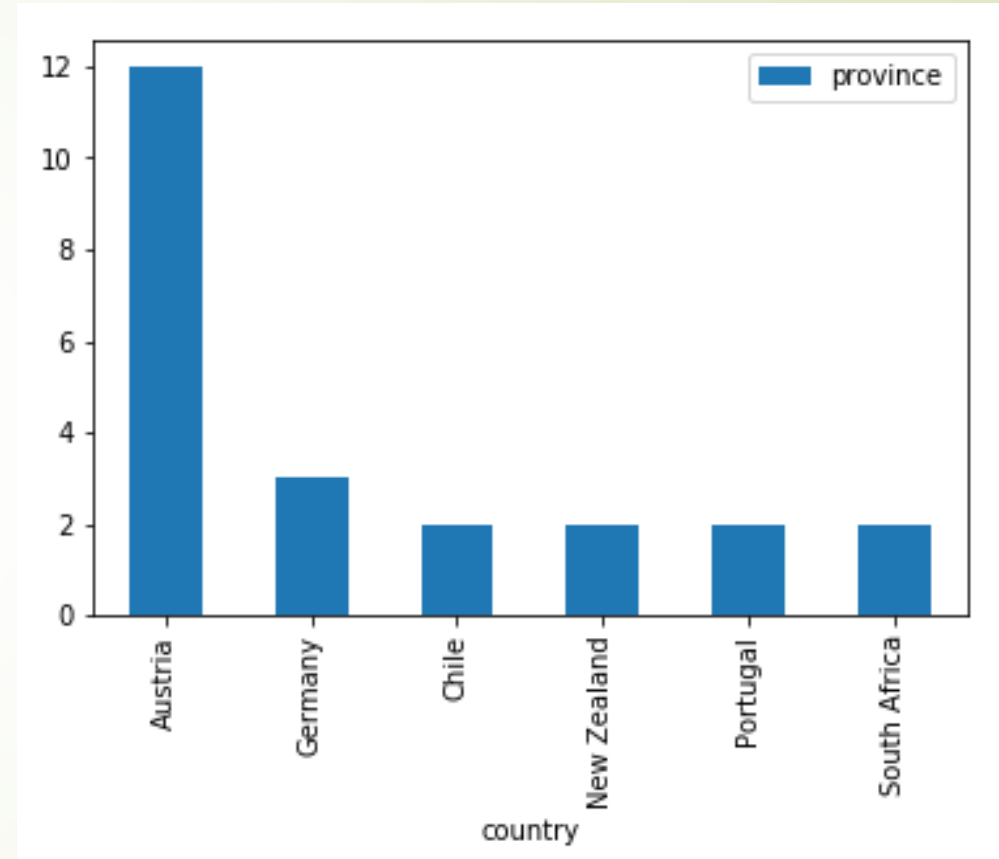
```
]:
```

	points	price
points	1.000000	0.497471
price	0.497471	1.000000



Summary of Statistical Analysis - 2

- Initial analysis of Top wine producing regions, revealed that country alone is not a good predictor (e.g. Portugal appears in Top 5 and Bottom 5)
- Quality rating of each region was derived by using the **Mean** value or Points for all wines in the region
- Countries with few wineries or regions are not a reliable indicator. These were removed.
- Chart indicates top countries and number of Provinces that have a high quality rating (as determined by mean points score)

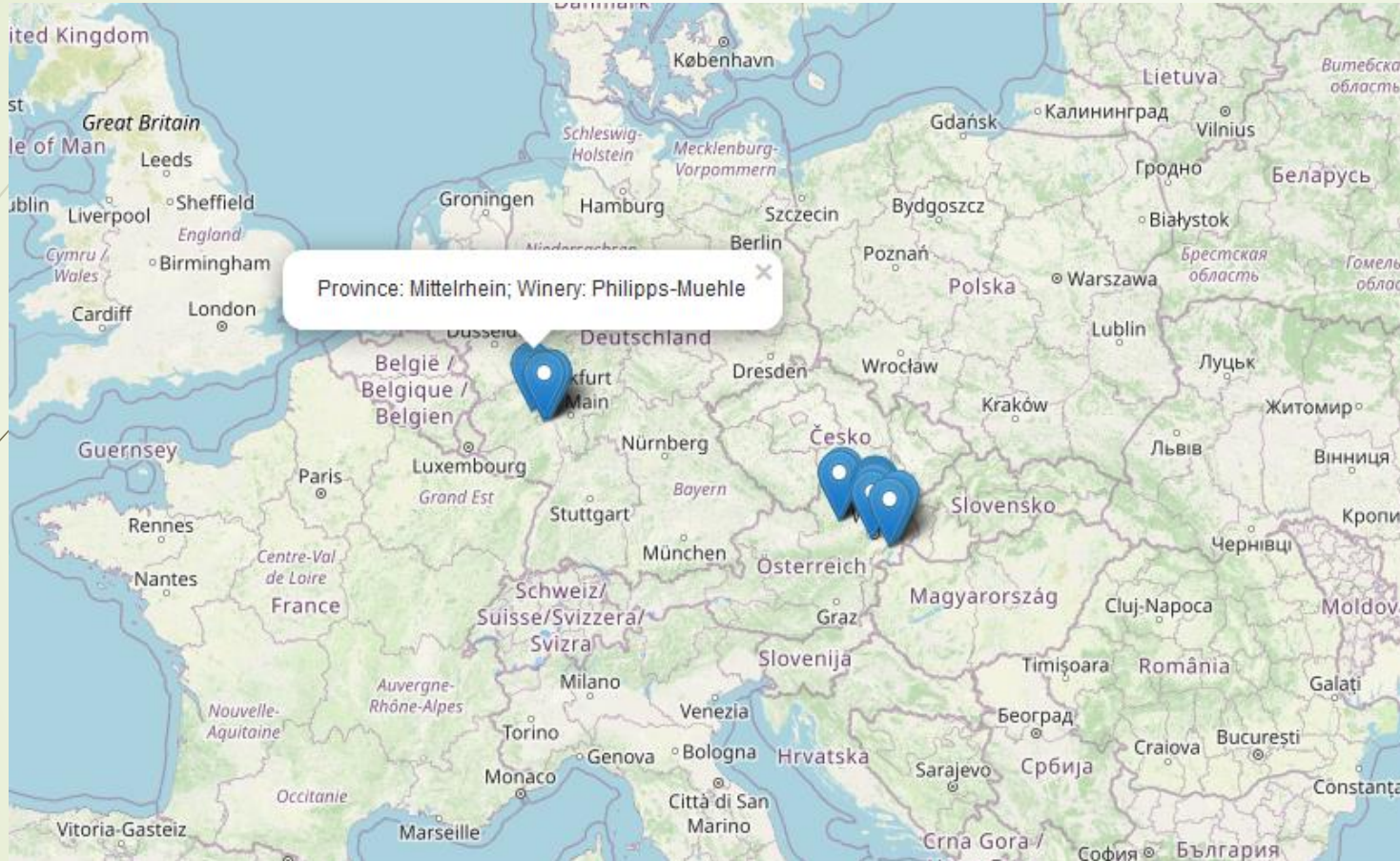


Top Wine Producing countries

Where are the Top wine Regions?



Where are the Top Wineries



Interactive world map, plots the Top 50 wineries. This includes:
Name of winery and Web site url



Next Steps

- Continue with New Business Feasibility Study:
 - Research the Top 50 wine producers(i.e. wine varieties, distribution channels, any distributors in the UK, price range)
 - Develop a cash flow model to assess the profitability of importing and selling these wines
- Continue to evolve the Data Science model to further identify high quality wine producers:
 - Analyse elevation of vineyards as a factor in determining quality