Submitted By: Divyang Deep Tiwari

Summary of Wine Dataset Analysis:

The journey of this analysis was quite fun and with some ups and downs. It mainly happened in two approaches.

Firstly, just for experimental purpose, I tried training model on features other than review_title and review_description which resulted in a huge failure. Yes, it gave maximum possible model accuracy of about 30% (using Decision Tree Classifier). Other Classifier like Gradient Boosting and Random Forest were also implemented on it but they gave even less accuracy than Decision Tree.

So, it was clear that removing review attributes is not gonna help. Then, Natural Language Processing was introduced in the picture. Review_description column was cleaned by removing stopwords, punctuation marks and retaining useful words. After that, top 10000 occuring words were assigned as the features to be extracted. Logistic Regression model were then used on the cleaned dataset. The validation of the model on test set resulted in a accuracy of around 70%. That is a quite significant rise in the accuracy as compared to previously obtained 30%.

Actionable Insights:

- California is the most repeated province in the dataset. It means it is definitely a good market as compared to others. Similarly, Pinot Noir is the most frequent variety of wine. It could be possible that people prefer this variety over others.

- 'Nebbiolo' and 'Grüner Veltliner' are the varieties that scored highest mean points (around 90). Whereas, 'Pinot Grigio' and 'Rosé' scored the least (~ 86).

- Average price values for particular variety were found to be very luctuating with minimum at 15.34 (Portuguese White) and maximum at 71.6 (Champagne Blend)

- Average price of wines in Italy (46.57) is high as compared to other frequently occuring countries. Considering the profitability, one could make a decision to increase the selling of wines with high points in Italy.

- Another interesting finding is that the winery 'Louis Latour' is the second most repeated winery in the dataset that belongs to France. While the average price of wines in France is around 43.37, the average price for this particular winery is 131.50. That's a really high surge.

A businessman can combine the above informations and decide to sell wines made by 'Louis Latour' at Italy or California. Moreover, if possible he can order varieties like 'Nebbiolo' or 'Champagne Blend' which either is highly rated or highly priced.

This is the analysis that has been carried out on the dataset in the time provided.