Summary

1.Model Used: I have used the Random Logistic Regression model.

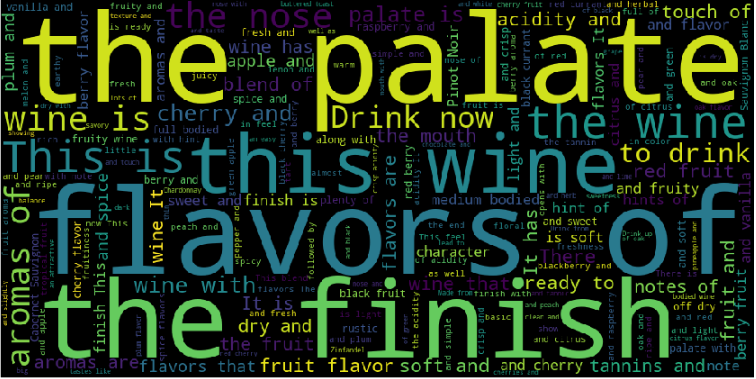
2.Features Extracted: I have extracted the Review Description attribute for the prediction as it contains the reviews on whose basis, I have predicted the variety of the wine.

3.Model accuracy in train: The model is giving an accuracy of 70.77% in the train dataset.

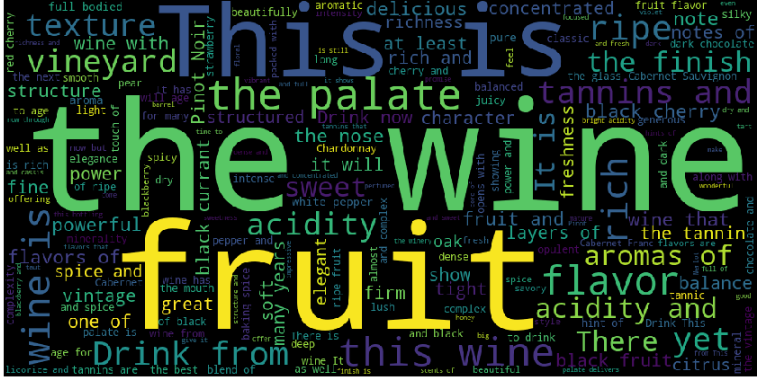
4.Data visualisations with inference’s:

a) Analysing the reviews:

Top wines:



Bottom wines:



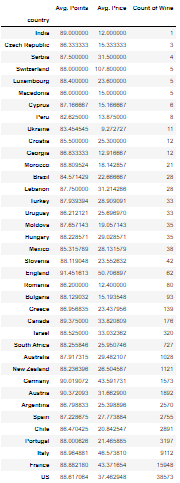
Looking at the words that are mostly used in the reviews of the highest rated wines (>=95 points)

and the lowest rated wines (<=85 points) the results are not very informative. If we exclude the basic common words

(e.g. wine) or the neutral words (e.g. note), the 3 words that mostly appear in the

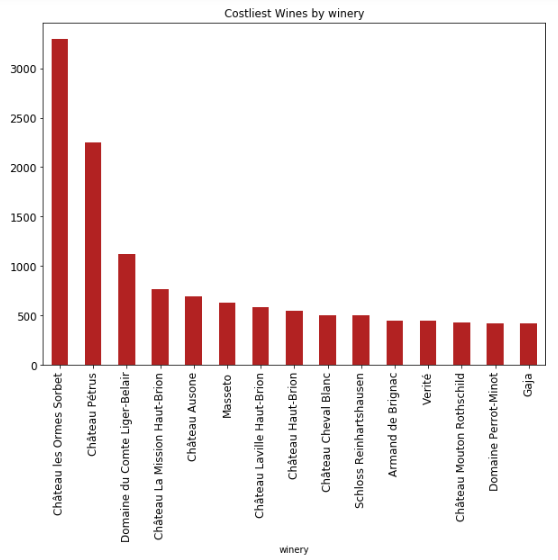
higher rated wines are: fruit, tannin, rich, aroma, power. For the lower rated wines the words are: dry, sweet, light/simple, soft.

b) Analyze wine from countries:



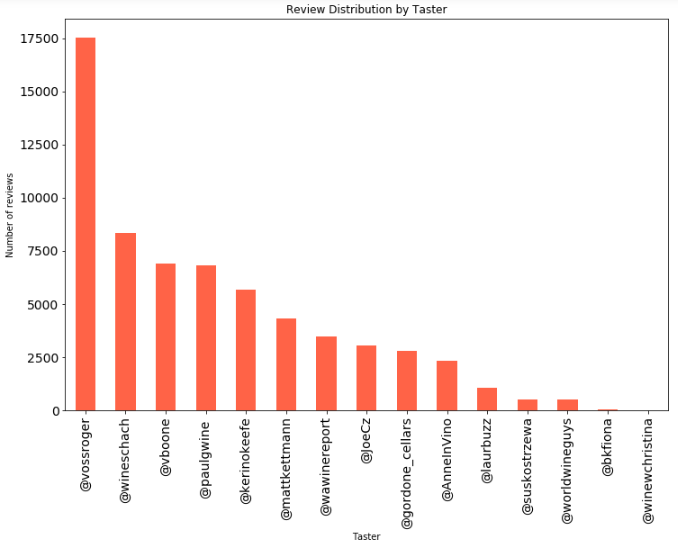
We can infer that US provide us with the maximum number of wines.

c) Analyze costly wines by Winery:



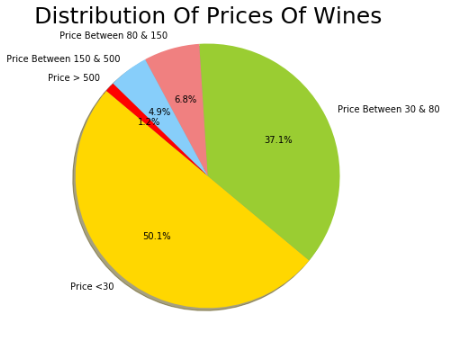
The Chateau Ies Ormes Sorbet winery has the costliest wine.

d) Analyze review distribution by taster user name:



We can see from the above graph that taster with user name @vossroger has shared the greatest number of reviews.

e) Visualising Distribution Of Wines In Various Price Groups:



We see that most of the wines produced are in the price range of 0 to 30.