**Summary**

* Initially, I started with the EDA process. In this, I planned to take out the important insights and trends that the data was showing. I saw that the ‘points’ were lying between 80 to 100 and this shows that the data is right skewed i.e. shifted to the right side of the distribution.
* Also, ‘region\_1’ and ‘region\_2’ columns were having a major number of null values which was not good for the model and replacing them was not going make any sense out of it.
* Further, I checked that our response variable (dependant variable) i.e. ‘variety’ was having 28 unique classes and we have to predict that. So, to do that I converted the categorical data into numerical form from ‘variety’ and ‘winery’ which were important for the model.
* I also plotted some graphs to know more about the features present in the dataset, and it turns out that ‘ratings’ and ‘price’ were showing some kind of a relationship. Also, I plotted a heatmap to further know about the correlation of the data.
* Then, I used the Stop-words and PorterStemmer from Natural Language Toolkit for the text analysis of the ‘review\_title’ column as to determine the sentiments of the users about the variety of wines. And, I used CountVectorizer (Bag of Words) as the model for the analysis.
* Then, the data was split into train and test and I also checked for the top feature names that were being given in the model.
* I used Multinomial Naïve Bayes Classifier algorithm for the dataset as it works well with text data, and it gave an accuracy of about 86.9 % approx. which is pretty good.
* To make the model more accurate I also checked it with another algorithm which was Passive Aggressive Classifier and I obtained and even better accuracy on this which was 92.2% approx.