➔ **The model used:** After taking care of the missing data, removing the stopwords, and vectorizing the string columns using TfidfVectorizer. LinearSVC performed the best on the validation dataset after modeling the dataset on various classification algorithms like Logistic Regression, Naive Bayes, Random Forest, Decision Tree, and RBF SVC.

➔ **Features extracted:** To deal with missing data in price, we've replaced the missing data with the mean of the entire 'price' column in the dataset. The missing elements in 'country' and 'province' are very small in number with respect to the size of the data. Therefore, here we have replaced the missing elements with the most frequent element in that particular feature. Further, we eliminated the features one by one by backward elimination to see if there's some irrelevant feature. **We found that the 'review_title' is sufficient enough to predict the 'variety' with an F1 score of 0.96**

➔ **Model accuracy:**

	F1 Score	Recall	Precision	Accuracy
Training (25%)	0.987	0.986	0.987	0.98732
Validation (25%)	0.964	0.965	0.965	0.96472

➔ **Top 5 actionable Insights from the Data:**

1. The ‘review_title’ feature is the most suited variable in the dataset to predict the variety. There were no instances of missing data in this column.

2. Online Wine Shop named “The Wine Land” should focus on ensuring the visitor or the customer reviews the products. They should make this as a mandatory field in their review form so that the variety is predicted successfully.

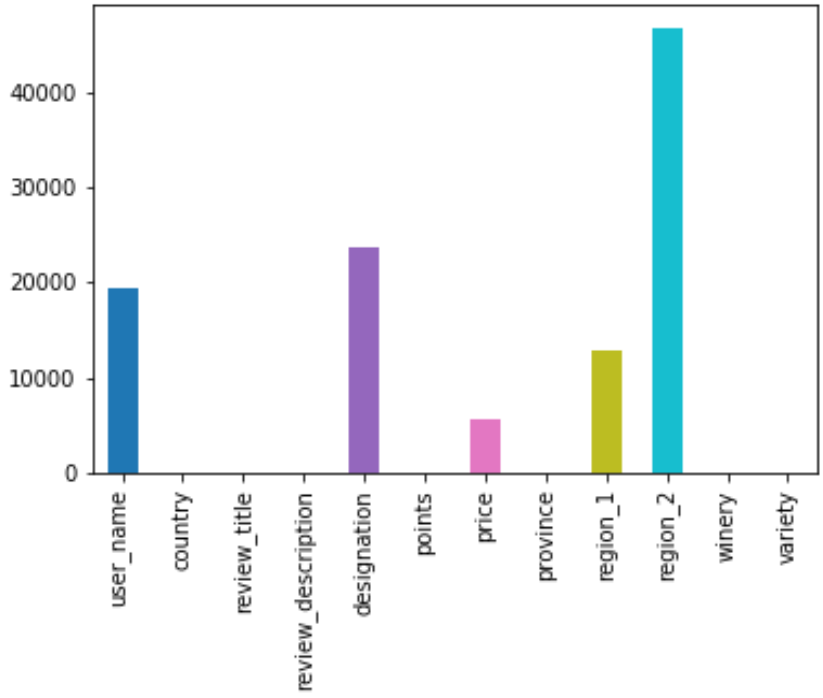
3. From feature elimination, it’s evident that fields like ‘province’, ‘region_1’, and ‘region_2’ are futile. Therefore, to get a detailed feedback form they can remove these irrelevant questions to push the customers to fully fill the review form with no missing columns.

4. Customers usually give a detailed review_description in the data which shows how engaging the products are.

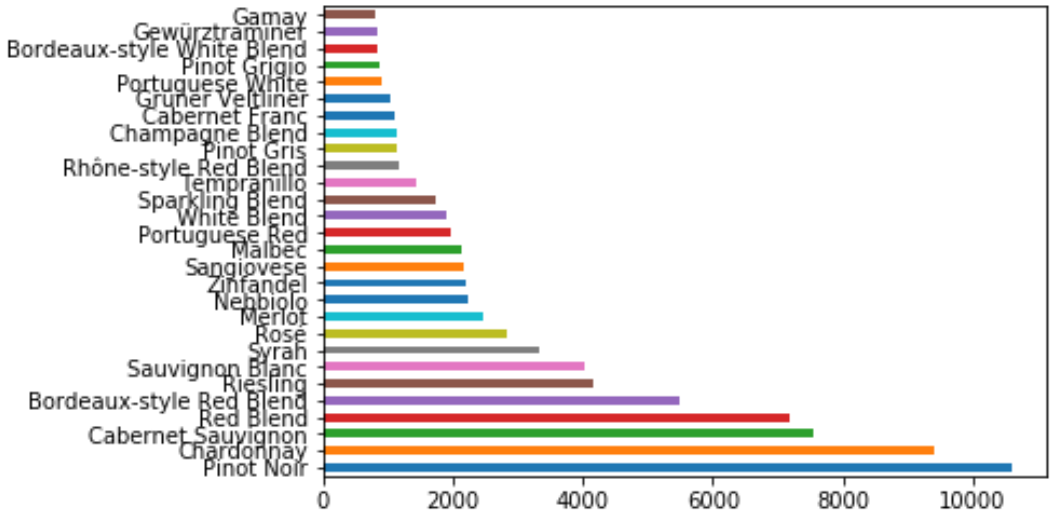
5. The minimum rating received is 88 (out of 100) which is very commendable as the dataset contains 80,000+ samples. Customers are loving the services offered by the company. (Visualization below)

➔ **Visualization of data:**

Missing Data in the features provided

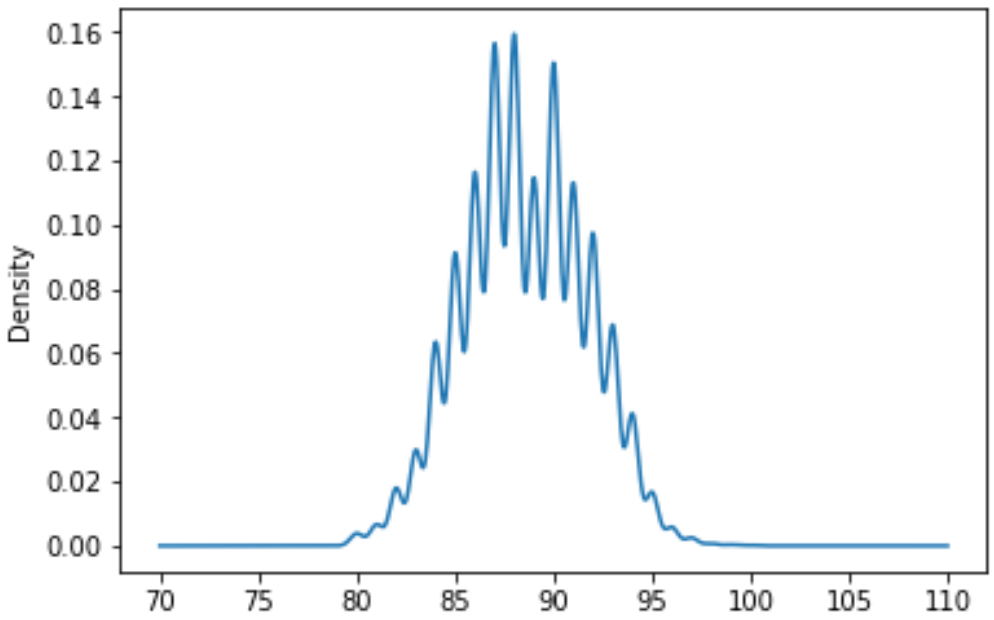


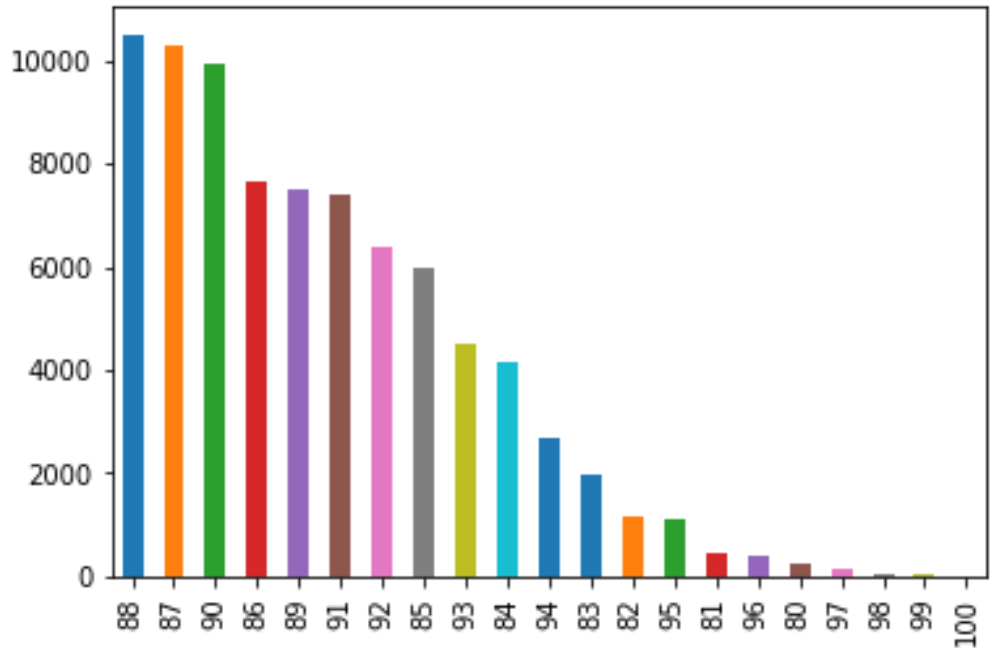
Output 'variety' classes bar plot



Average Rating: Are the users loving the service?

Ratings lie b/w 88 and 100.





Validation Results

	precision	recall	f1-score	support
Bordeaux-style Red Blend	0.884	0.895	0.889	1366
Bordeaux-style White Blend	0.906	0.736	0.812	197
Cabernet Franc	0.985	0.996	0.990	261
Cabernet Sauvignon	0.999	0.998	0.998	1933
Champagne Blend	1.000	0.983	0.991	287
Chardonnay	0.991	0.979	0.985	2399
Gamay	0.968	1.000	0.984	209
Gewürztraminer	1.000	1.000	1.000	218
Grüner Veltliner	1.000	1.000	1.000	252
Malbec	0.989	0.998	0.994	538
Merlot	0.995	0.985	0.990	608
Nebbiolo	1.000	0.998	0.999	548
Pinot Grigio	0.995	1.000	0.998	221
Pinot Gris	1.000	1.000	1.000	323
Pinot Noir	0.980	0.986	0.983	2630
Portuguese Red	0.994	0.988	0.991	483
Portuguese White	1.000	0.991	0.996	228
Red Blend	0.853	0.882	0.867	1826
Rhône-style Red Blend	0.889	0.681	0.771	282
Riesling	0.996	0.999	0.997	998
Rosé	0.985	0.978	0.981	723
Sangiovese	0.879	0.888	0.883	546
Sauvignon Blanc	0.981	0.988	0.985	954
Sparkling Blend	0.977	0.991	0.984	438
Syrah	0.989	0.998	0.993	842
Tempranillo	0.964	0.992	0.978	355
White Blend	0.964	0.974	0.969	494
Zinfandel	1.000	1.000	1.000	506
avg / total	0.965	0.965	0.964	20665