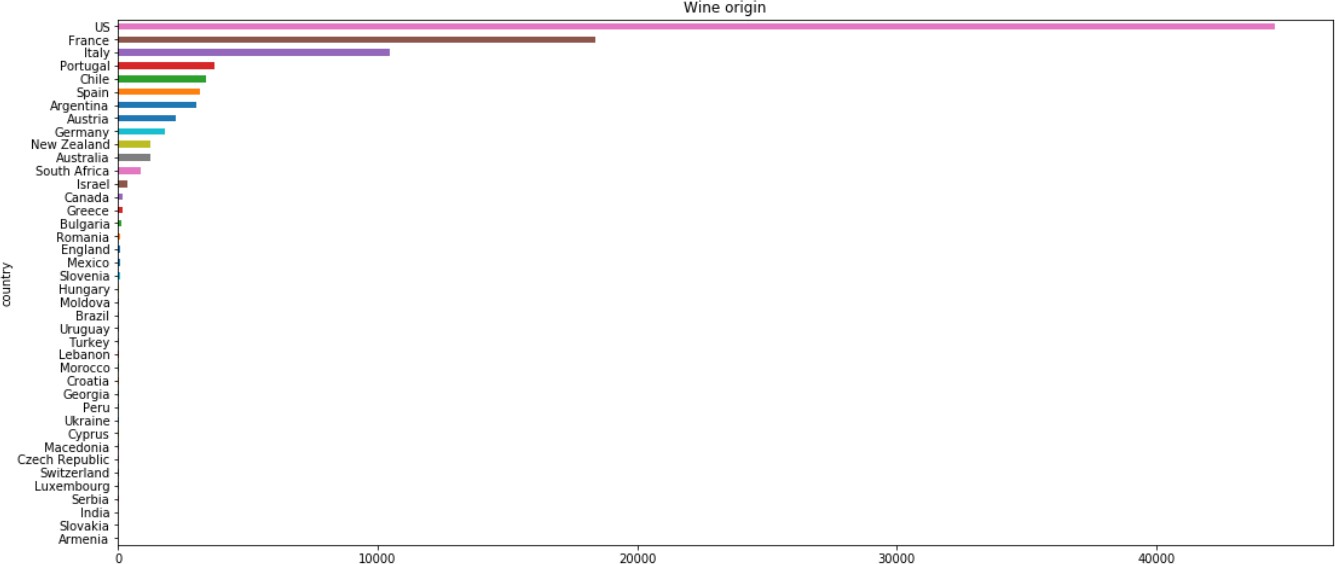
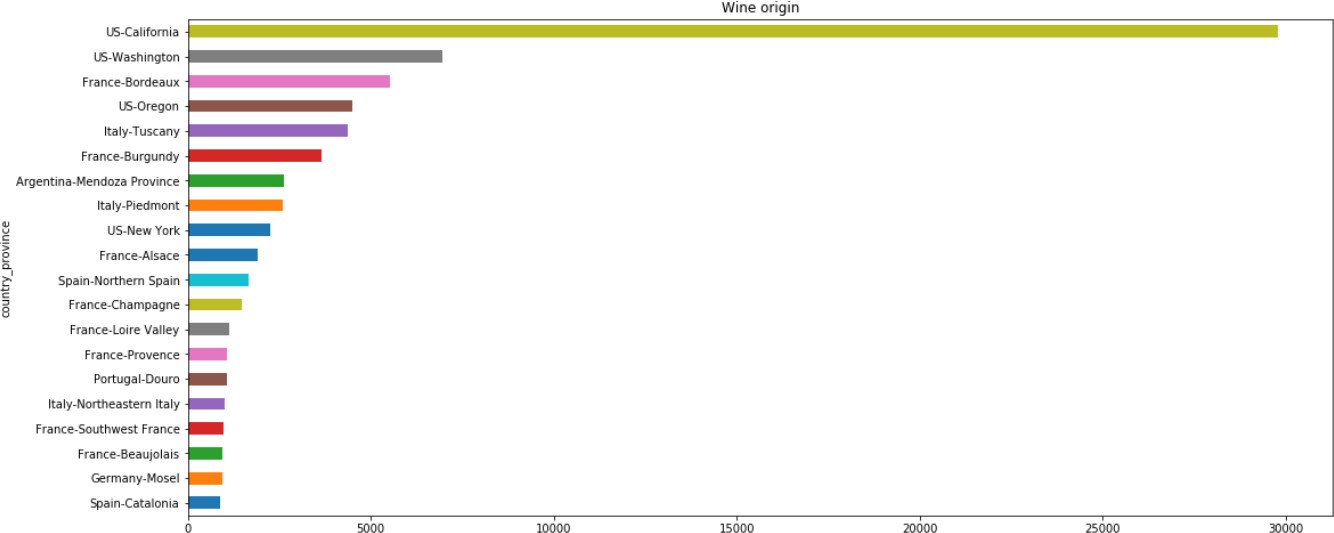
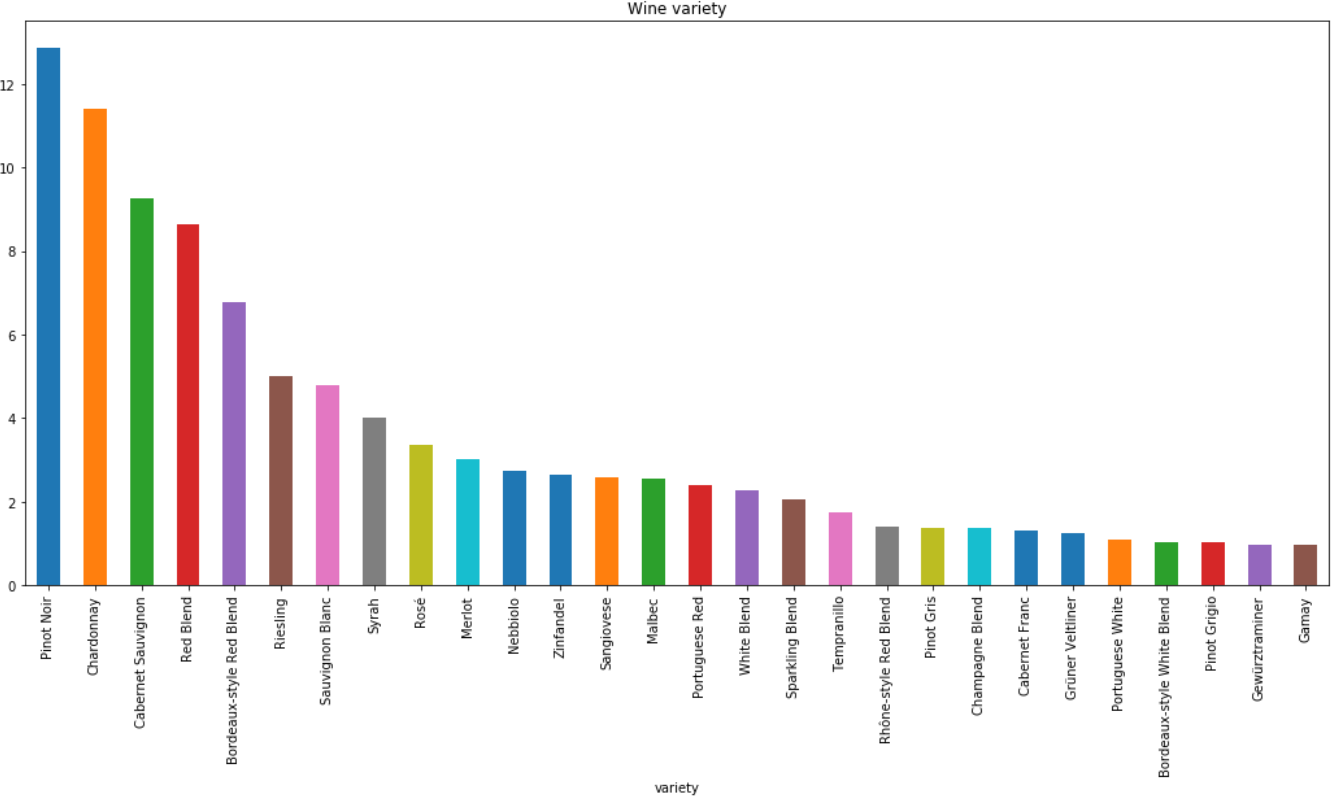
# Insights:

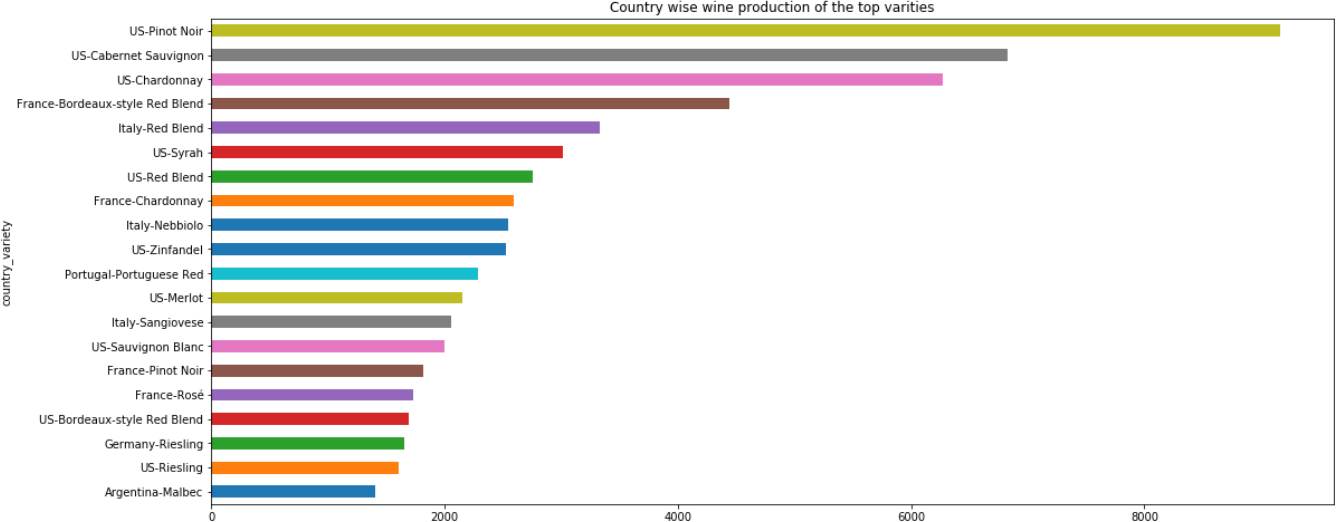
1. The Wine Land sales the highest number of wines produced in the US followed by France, Italy, Portugal, Chile and so on.



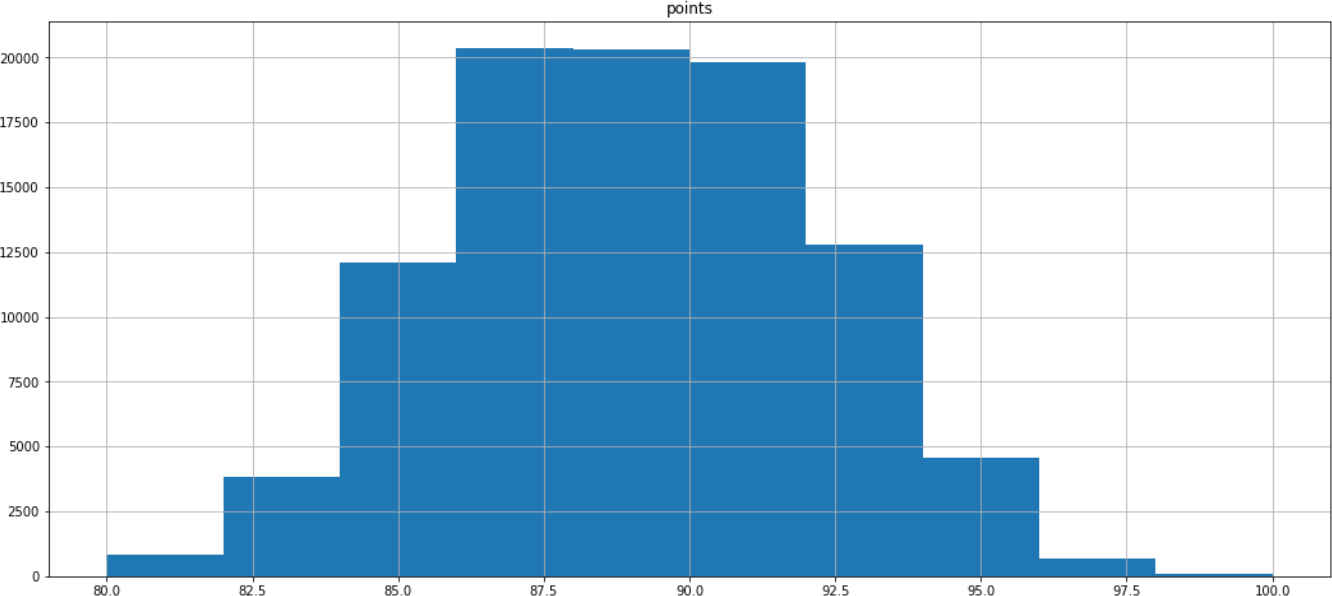
1. Province wise wine information for the top 20 countries.
2. The top 5 most popular wine varieties are as follows.
   * Pinot Noir
   * Chardonnay
   * Cabernet Sauvignon
   * Red Blend
   * Bordeaux-style Red Blend



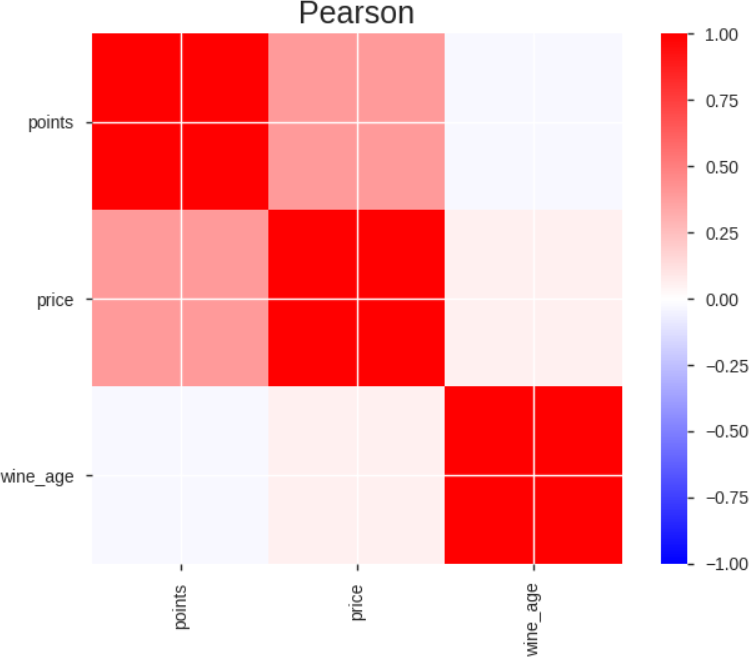
1. Country-wise information on the top variety of wines.



1. Ratings given by the users are normally distributed. On average a user gives a rating of 88.



1. Wine prices and points are not highly correlated. That means it is not likely that users are giving better ratings to costlier wines.



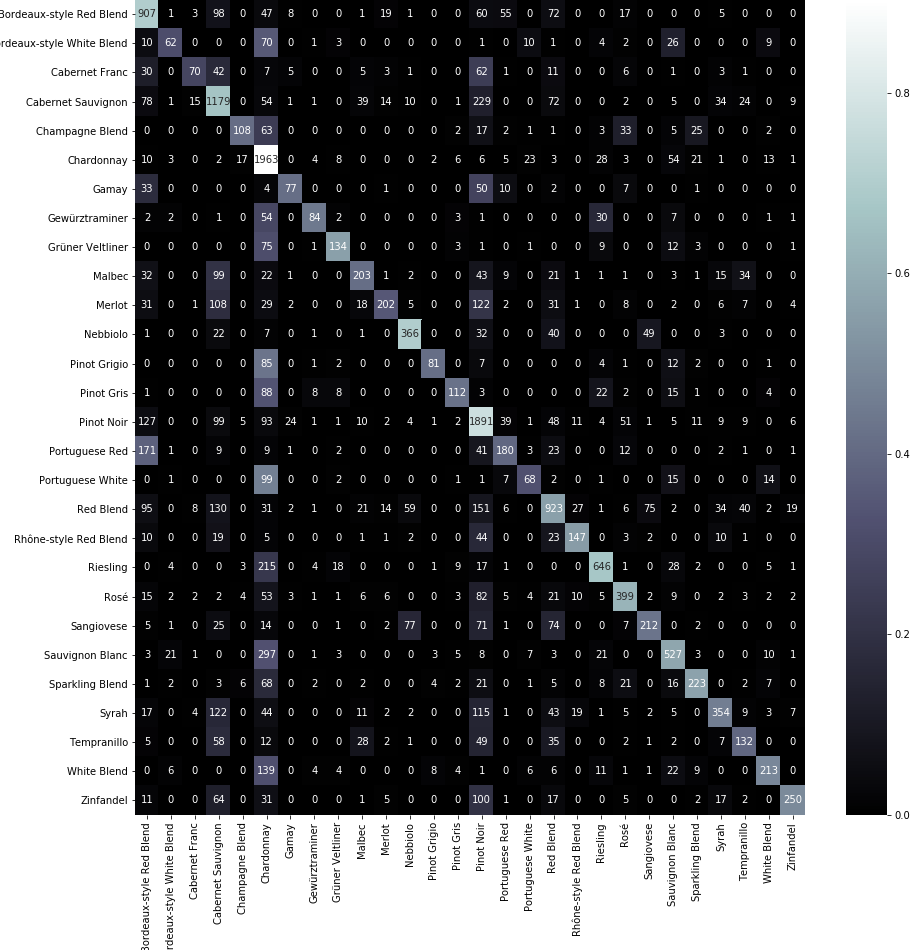
Created a new feature of wine\_age from review title column to check if the wine age has an effect on price and points. From the above plot, it is visible that it has no effects on points and very little effect on price.

# Performance metric of the predictive model:

1. Model Used: xgboost.
2. Feature: Created tf-idf feature from review description to build predictive model.
3. Classification report:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Precision | recall | F1-score | support |
| Bordeaux-style Red Blend | 0.57 | 0.70 | 0.63 | 1294 |
| Bordeaux-style White Blend | 0.58 | 0.31 | 0.41 | 199 |
| Cabernet Franc | 0.67 | 0.28 | 0.40 | 248 |
| Cabernet Sauvignon | 0.57 | 0.67 | 0.61 | 1768 |
| Champagne Blend | 0.76 | 0.41 | 0.53 | 262 |
| Chardonnay | 0.53 | 0.90 | 0.67 | 2173 |
| Gamay | 0.62 | 0.42 | 0.50 | 185 |
| Gewürztraminer | 0.73 | 0.45 | 0.55 | 188 |
| Grüner Veltliner | 0.71 | 0.56 | 0.62 | 240 |
| Malbec | 0.59 | 0.42 | 0.49 | 489 |
| Merlot | 0.74 | 0.35 | 0.47 | 579 |
| Nebbiolo | 0.69 | 0.70 | 0.70 | 522 |
| Pinot Grigio | 0.81 | 0.41 | 0.55 | 196 |
| Pinot Gris | 0.73 | 0.42 | 0.54 | 264 |
| Pinot Noir | 0.59 | 0.77 | 0.67 | 2455 |
| Portuguese Red | 0.55 | 0.39 | 0.46 | 456 |
| Portuguese White | 0.54 | 0.32 | 0.40 | 211 |
| Red Blend | 0.62 | 0.56 | 0.59 | 1647 |
| Rhône-style Red Blend | 0.68 | 0.55 | 0.61 | 268 |
| Riesling | 0.81 | 0.68 | 0.74 | 955 |
| Rosé | 0.67 | 0.62 | 0.64 | 644 |
| Sangiovese | 0.61 | 0.43 | 0.51 | 492 |
| Sauvignon Blanc | 0.68 | 0.58 | 0.62 | 914 |
| Sparkling Blend | 0.73 | 0.57 | 0.64 | 394 |
| Syrah | 0.71 | 0.46 | 0.56 | 766 |
| Tempranillo | 0.50 | 0.40 | 0.44 | 334 |
| White Blend | 0.74 | 0.49 | 0.59 | 435 |
| Zinfandel | 0.83 | 0.49 | 0.62 | 506 |
|  |  |  |  |  |
| micro avg | 0.61 | 0.61 | 0.61 | 19084 |
| macro avg | 0.66 | 0.51 | 0.56 | 19084 |
| weighted avg | 0.63 | 0.61 | 0.60 | 19084 |

1. Confusion matrix:



**Build API**: Below are the steps to build the API.

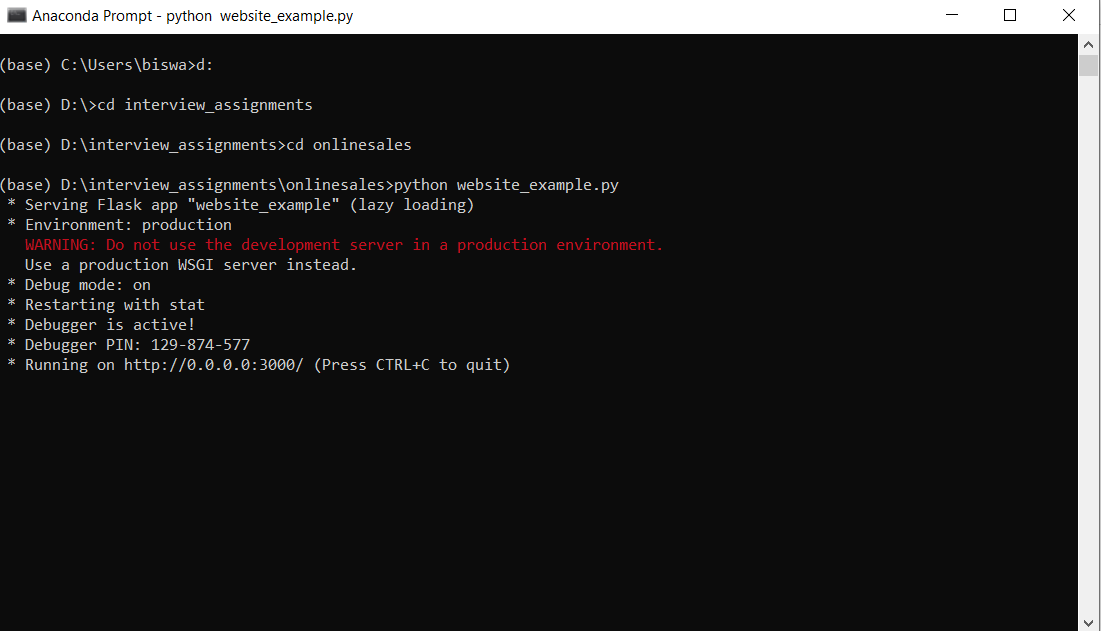
Step 1: Put the following files and folder along with the dataset in the present working directory.

Folder : templates

Python notebook : Build\_predictive\_model.ipynb Python script : website\_example.py

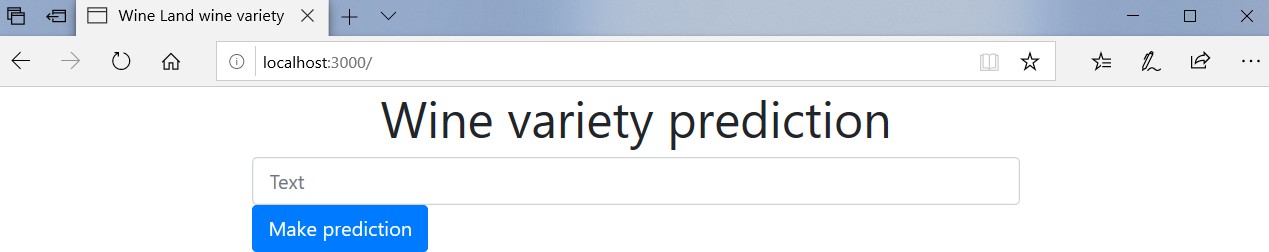
**Run API:** Below are the steps to run the API. Step 1:

Run the python script website\_example.py in the present working directory. This directory should contain saved models and other information that is required to make a prediction for new review descriptions.



Step 2:

Go to a web browser and type localhost:3000/ to open the following webpage.



Step 3:

Type in review and click the make prediction button to display the wine variety.

