Models Used:

- The main code is in Wine\_variety\_predictions.ipynb.

- CountVectorizer for converting words to vectors. LabelEncoder for Converting varieties column into integer categories.

- The predictions were made based on the ‘review\_descriptions’ column.

- Multi Layer perceptron for classification.Containing One Dense layer with 100 units and one output later with 26 units. Only 2 epochs needed because itwas overtraining after that. Got approximately 72 percent accuracy on the validation set and 91 percent accuracy on the test set.

- Initially tried to use RNN/LSTM using word2vec and Glove vectors but got only 25 percent validation accuracy.This may be due to the fact that the words used to describe wine are very specific and unique, may phrases are used specifically for wine. There is a very high vchance that these words are not present the the embeddings of word2vec and Glove vectors. One solution is to use the Robosomm Wine wheels to generate a custom wine embedding but this process would have taken much more time.

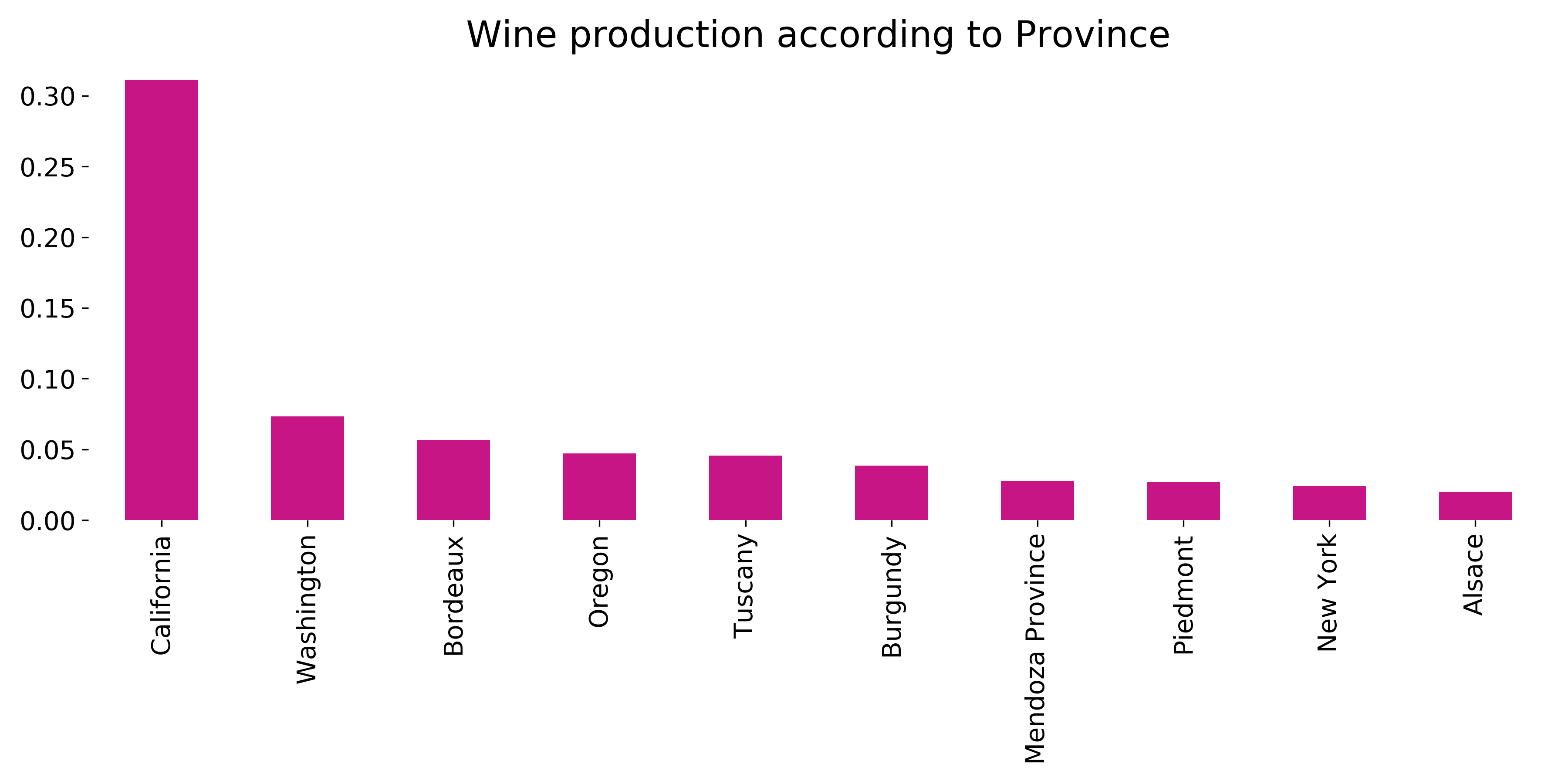
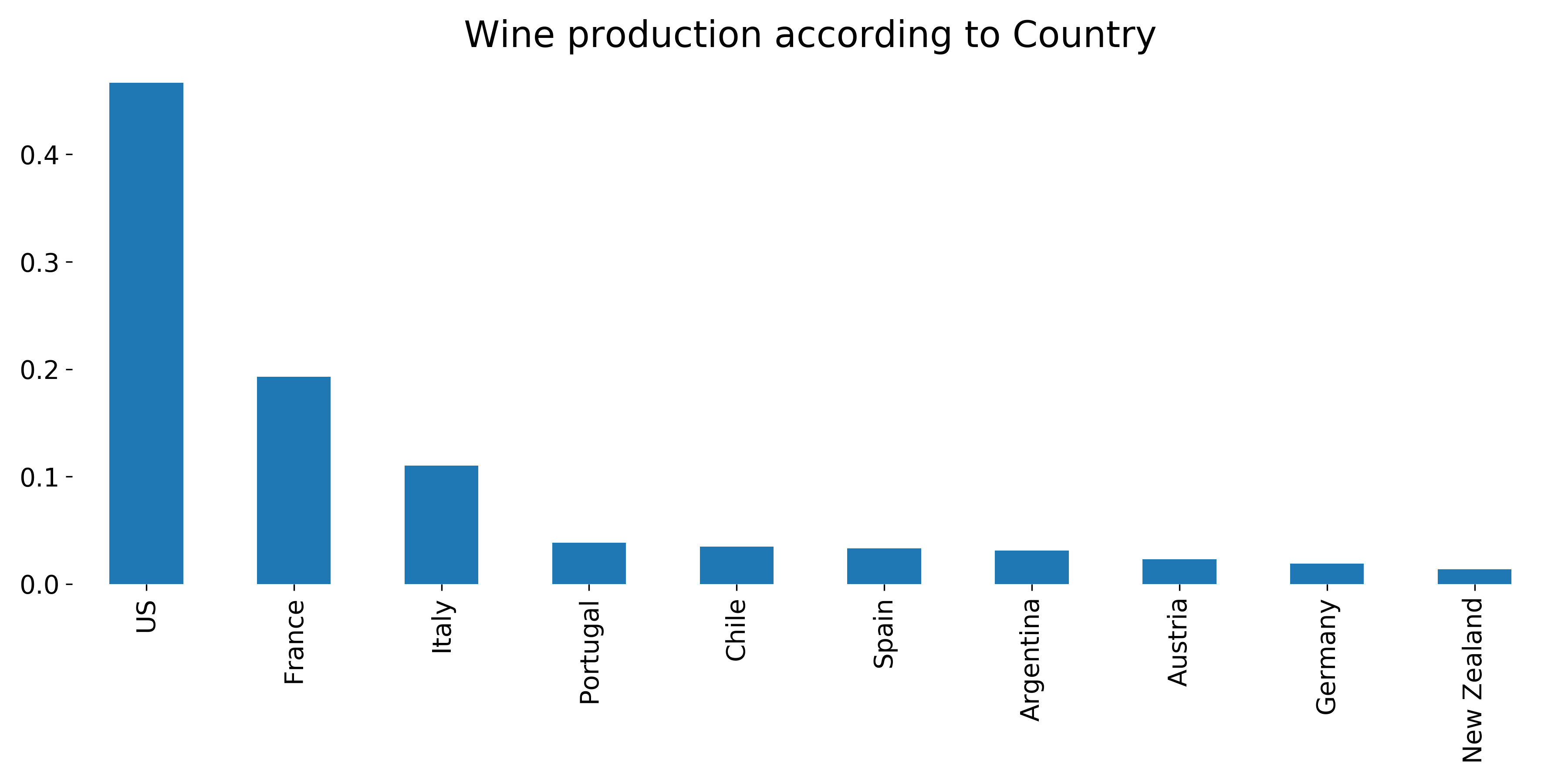
- Also tried another approach using Naive Byes for classification, but got 65 percent validation accuracy.

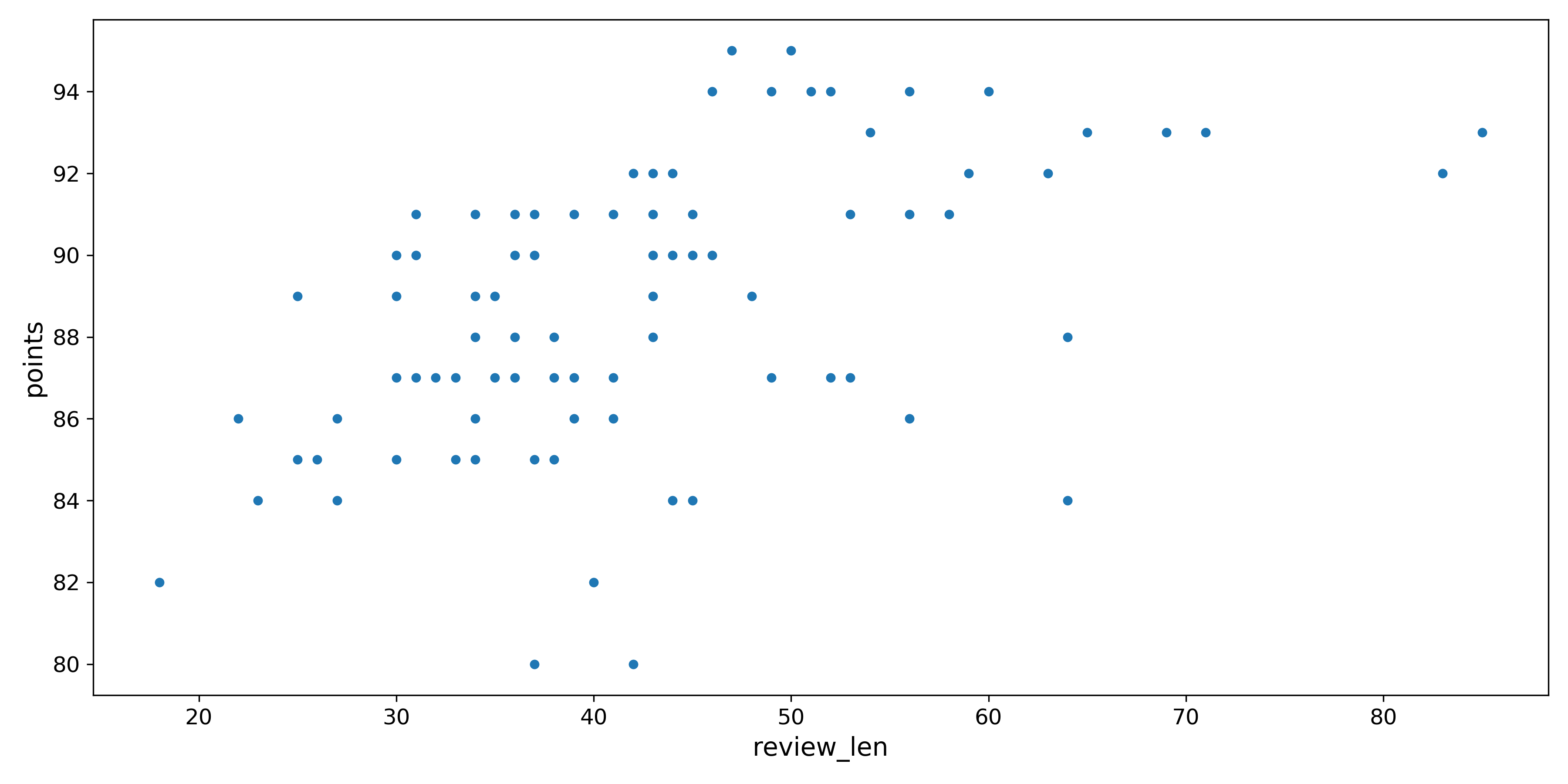
- According to the time limitation and performance, the best model among the three was the first one(MLP).

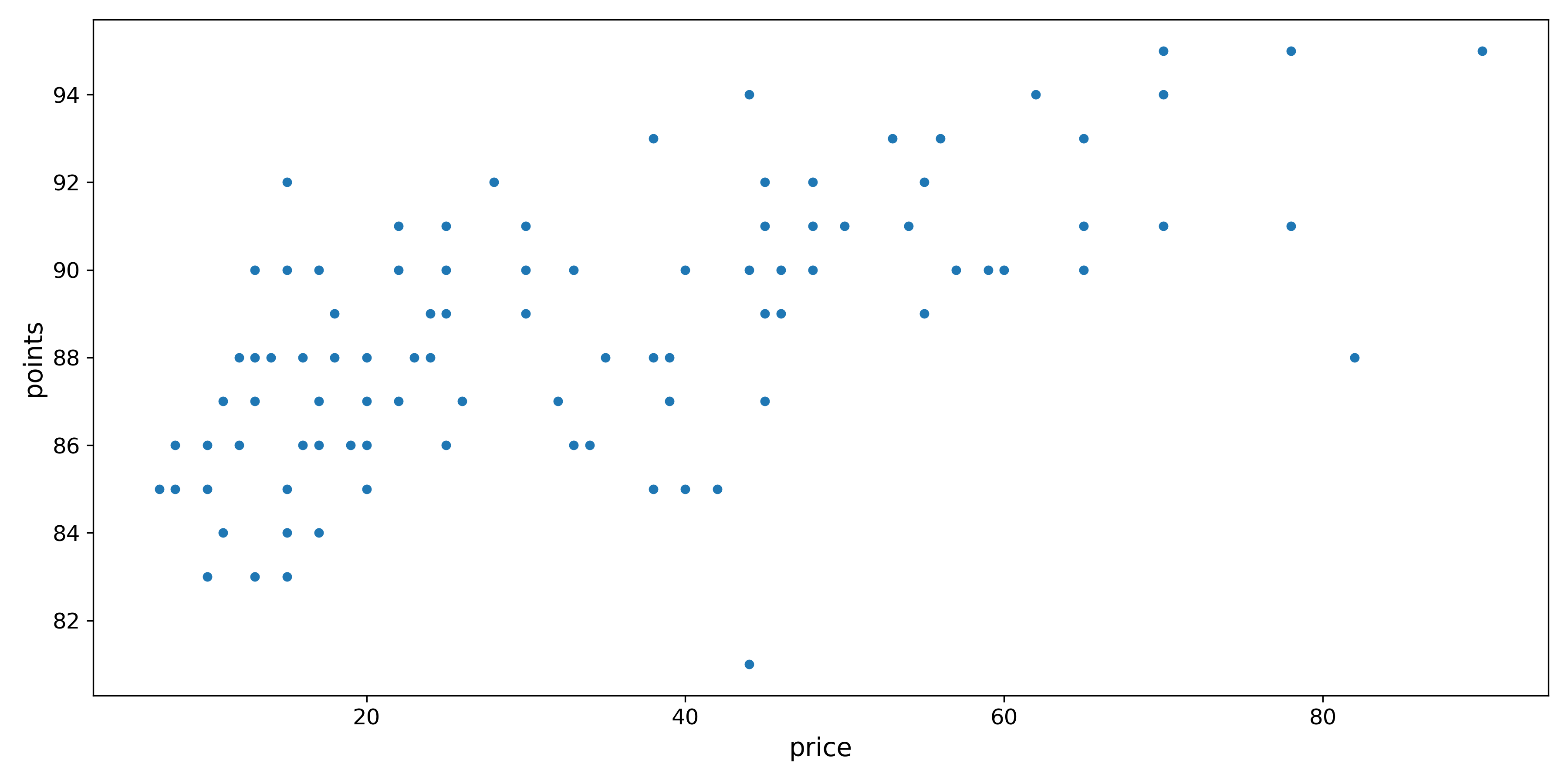
- Data visualization and data analysis notebooks are also submitted.

Top 5 insights:

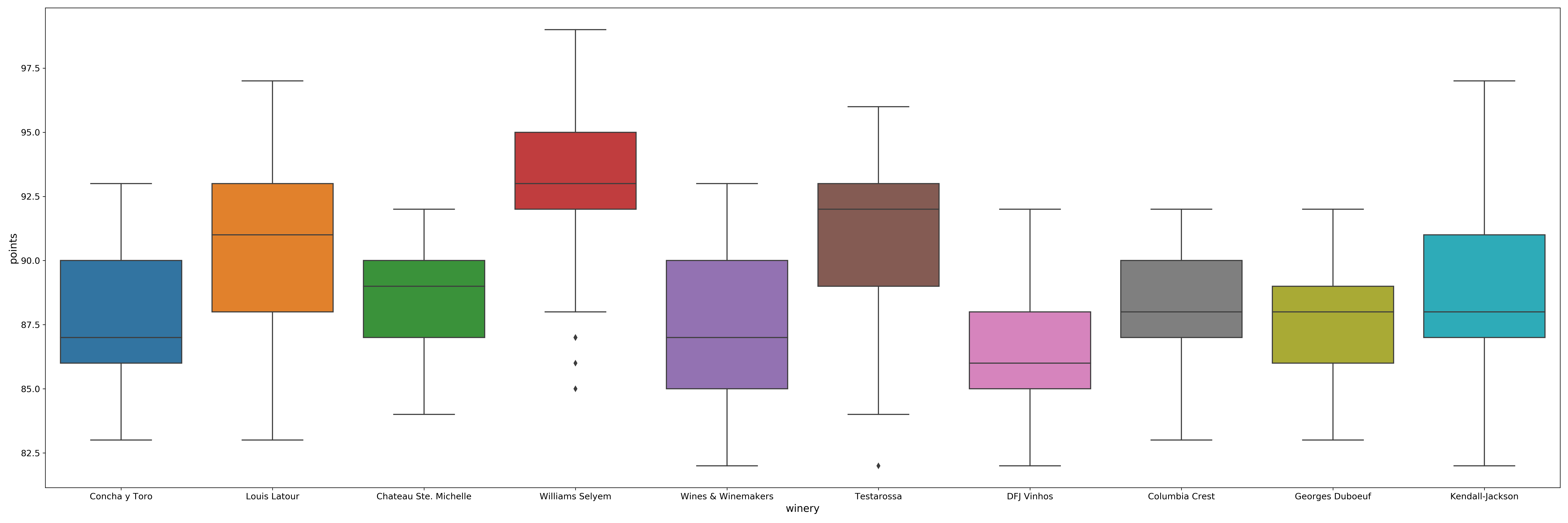
1) California produced most wine among provinces and US produced most wine among Countries.



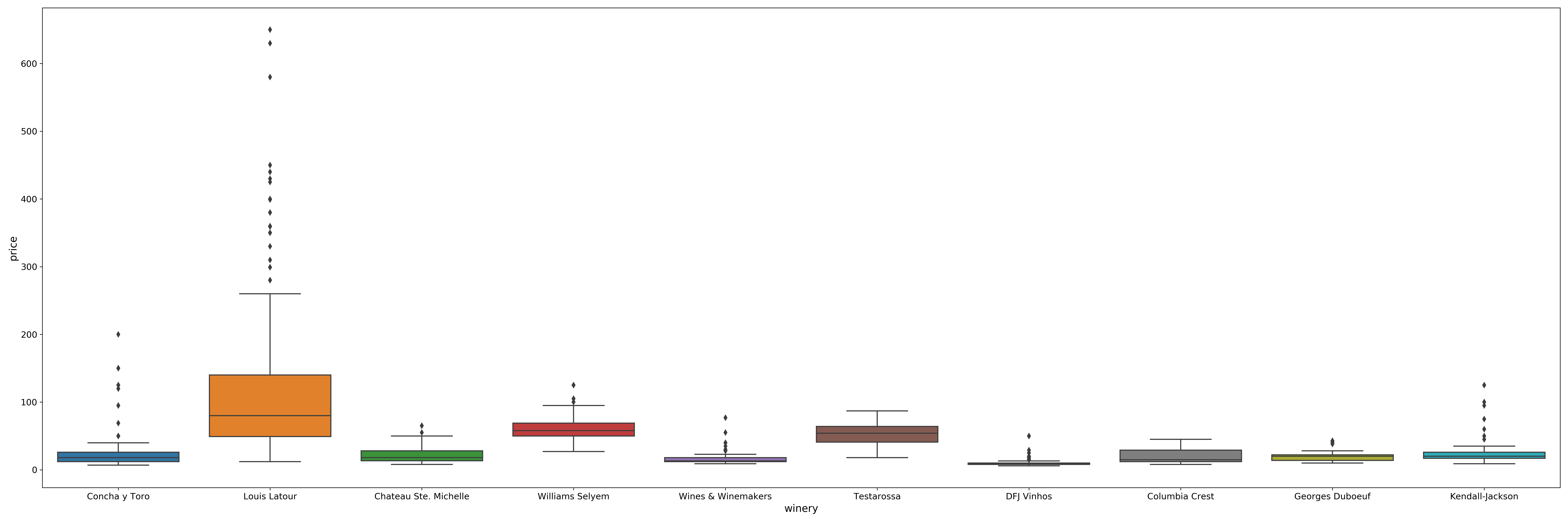
2) The points are proportional to the length of the review.

3) More expensive wines have more points. However, most people purchase wine between the price of 10 to 40.

4) William Selyem winery has the most number of points. So, according to customer reviews, it produces the best wine.



5) Louis Latour makes the most expensive wines.



ce price of 10 to 40.