



# Rating and Prediction Analysis

By: Shuyang Cui, Yulin Luo, Ziming Qi,  
Misha Reswick, Zhouyuan Yuan

ECE Winter 2023

# Outline



1. Motivation and Goals
2. Dataset
3. Methodology
4. Analysis
5. Prediction

# Motivation Goals: Questions We Have:

- In approaching a new industry, be one a producer or consumer, there are some questions that come to mind:
  - 1. How do we become informed about an industry?
  - 2. How do we then interpret different measuring tools?
  - 3. Who gives us industry feedback, and what characterizes that information? Is there bias?
  - 4. Are there any overall trends that further characterize what it means to be “successful” in our industry?
  - 5. How can we then make predictions about our field?

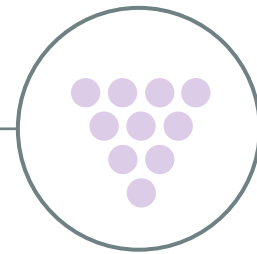
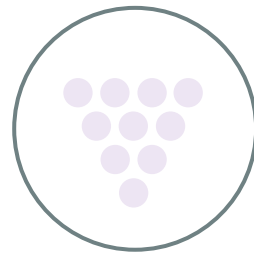
## 1. The Producer.



## 2. The Consumer.

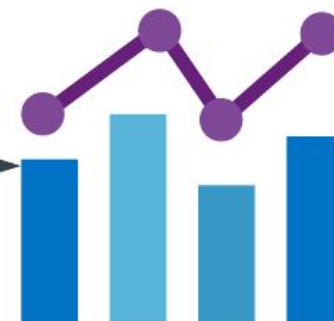
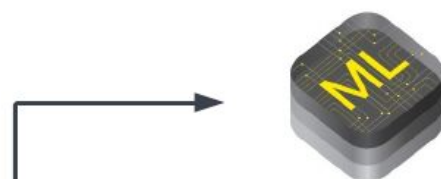
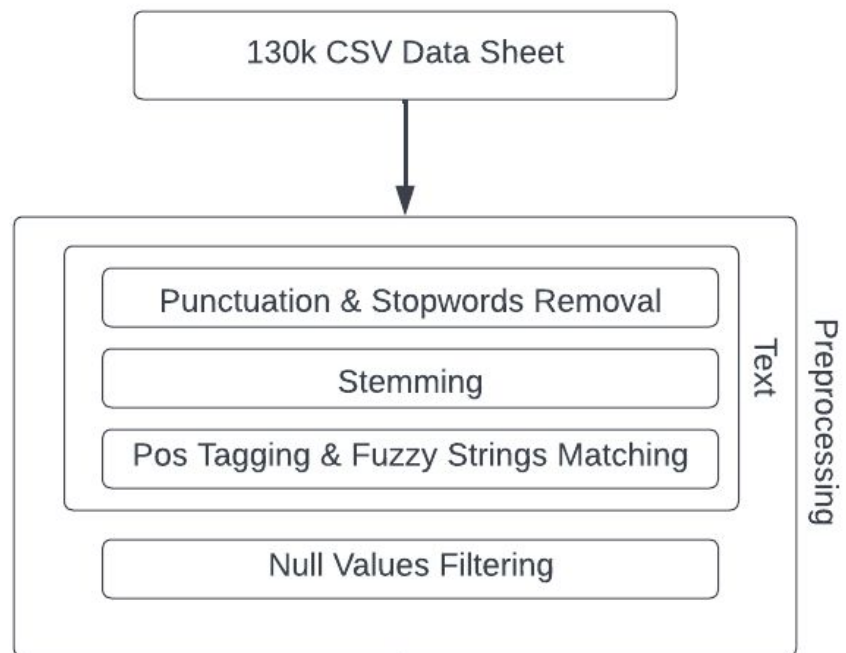


# **Dataset and Methodology**



# Dataset Overview:

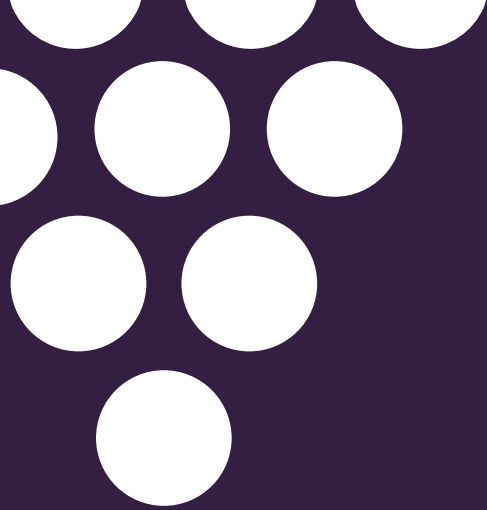
- ~130k rows
- 14 columns for reviews as follows:
  - country
  - description
  - designation (a location-based description of the wine)
  - points (in points 0-100)
  - price (in dollars)
  - province (overall state or region)
  - region\_1 (a more specific region within the province)
  - region\_2 (a further specific region within region\_1)
  - taster\_name
  - taster\_twitter\_handle
  - title (of the wine, such as with date)
  - variety (the type of grape)
  - winery



# Methodology







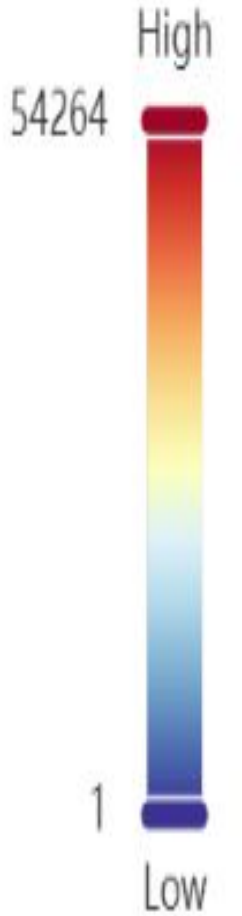
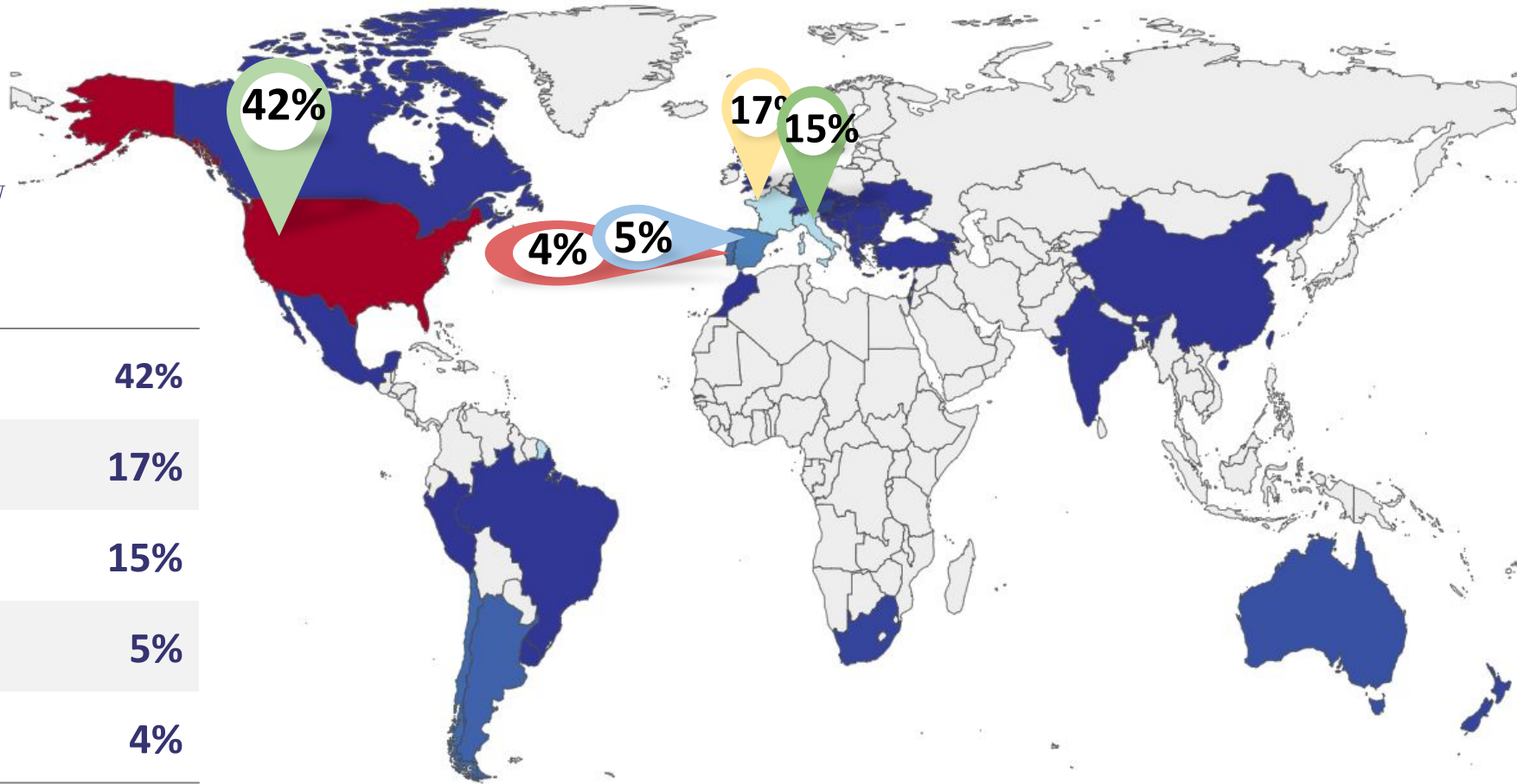
# Analysis

# Wine Counts Worldwide



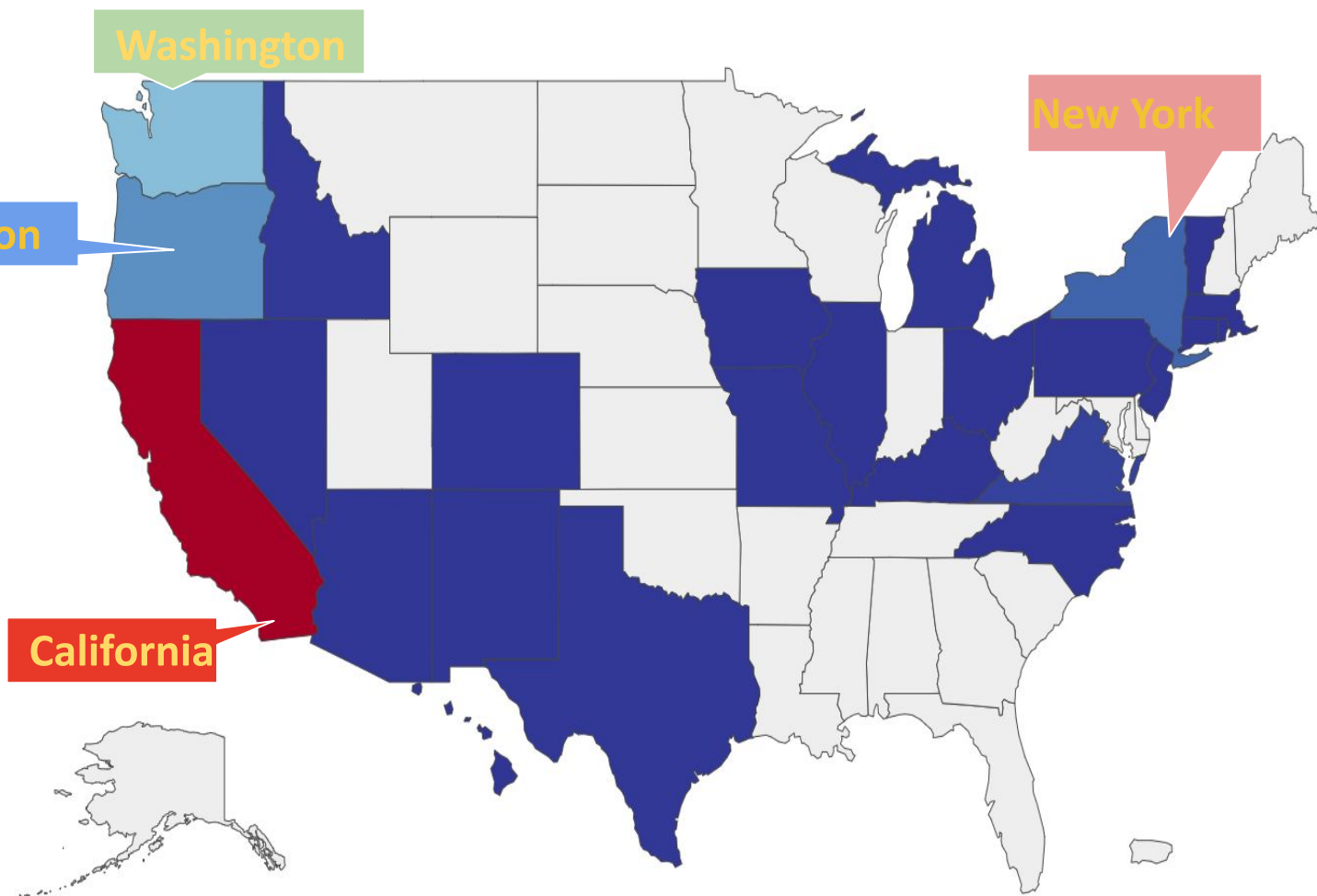
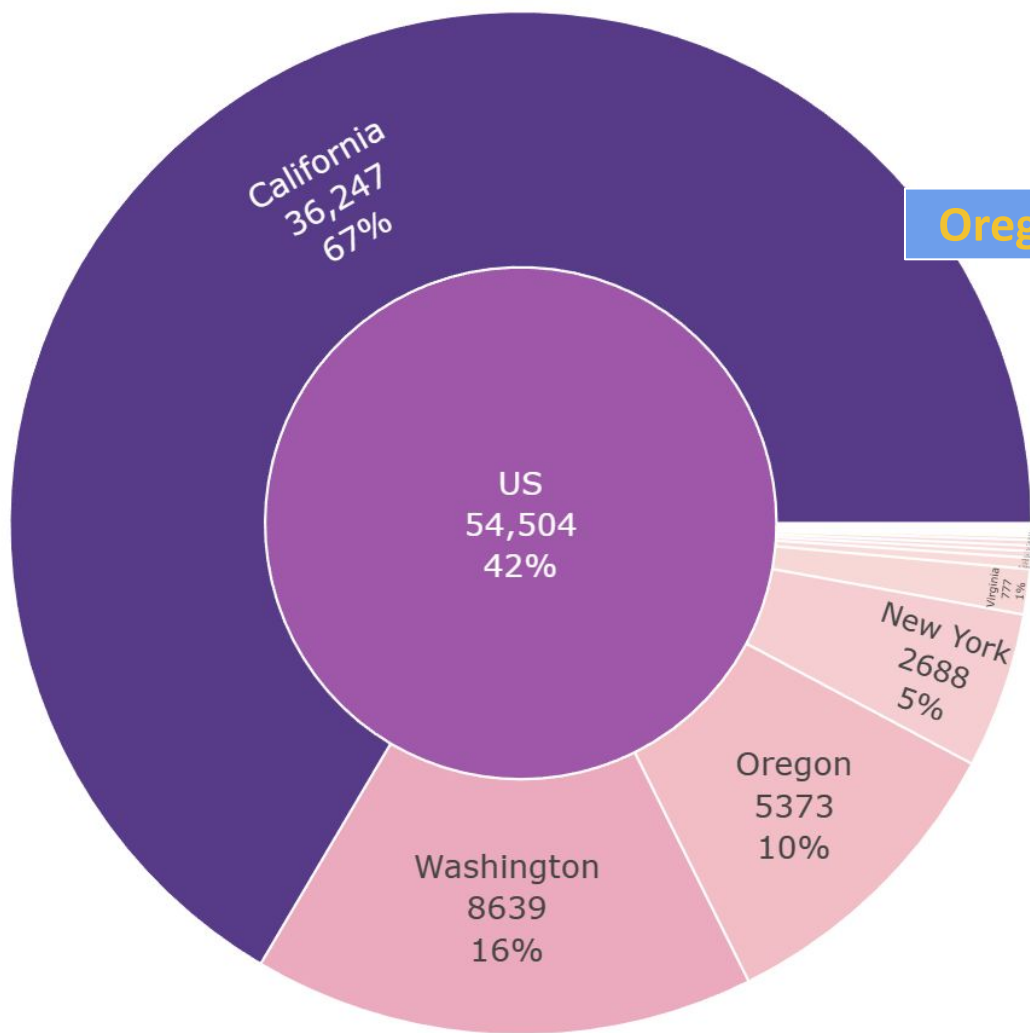
Review  
counts

United States	42%
France	17%
Italy	15%
Spain	5%
Portugal	4%

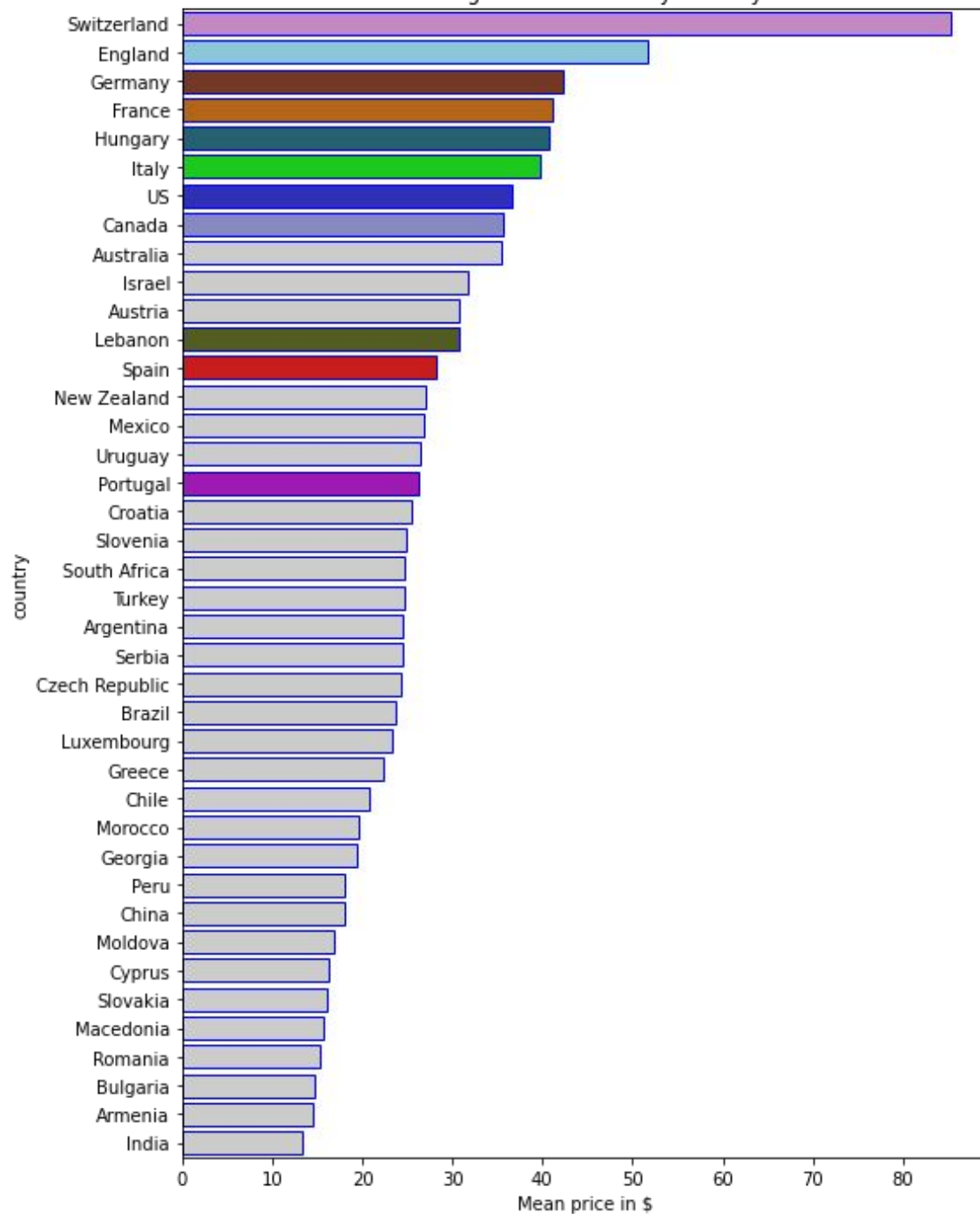


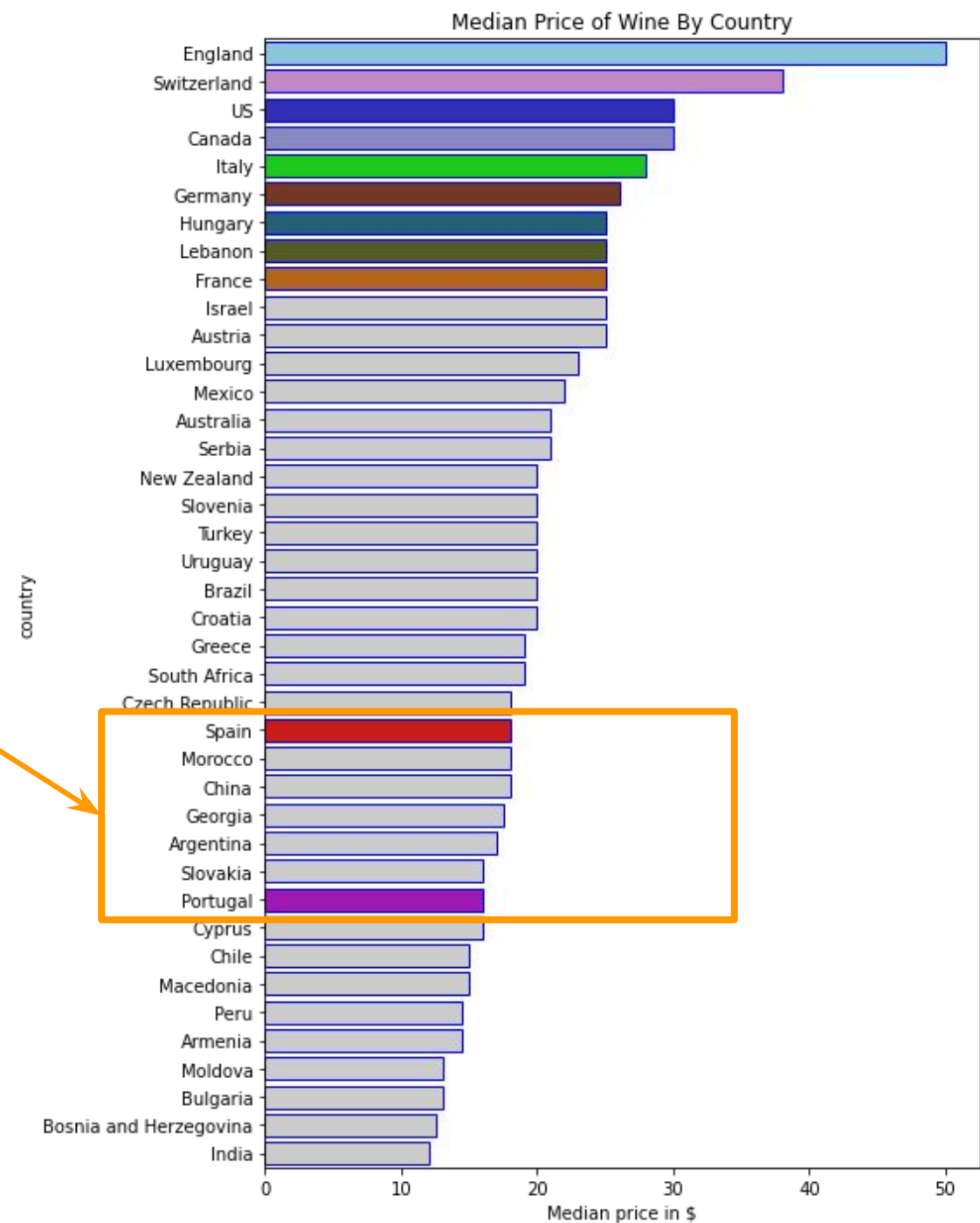
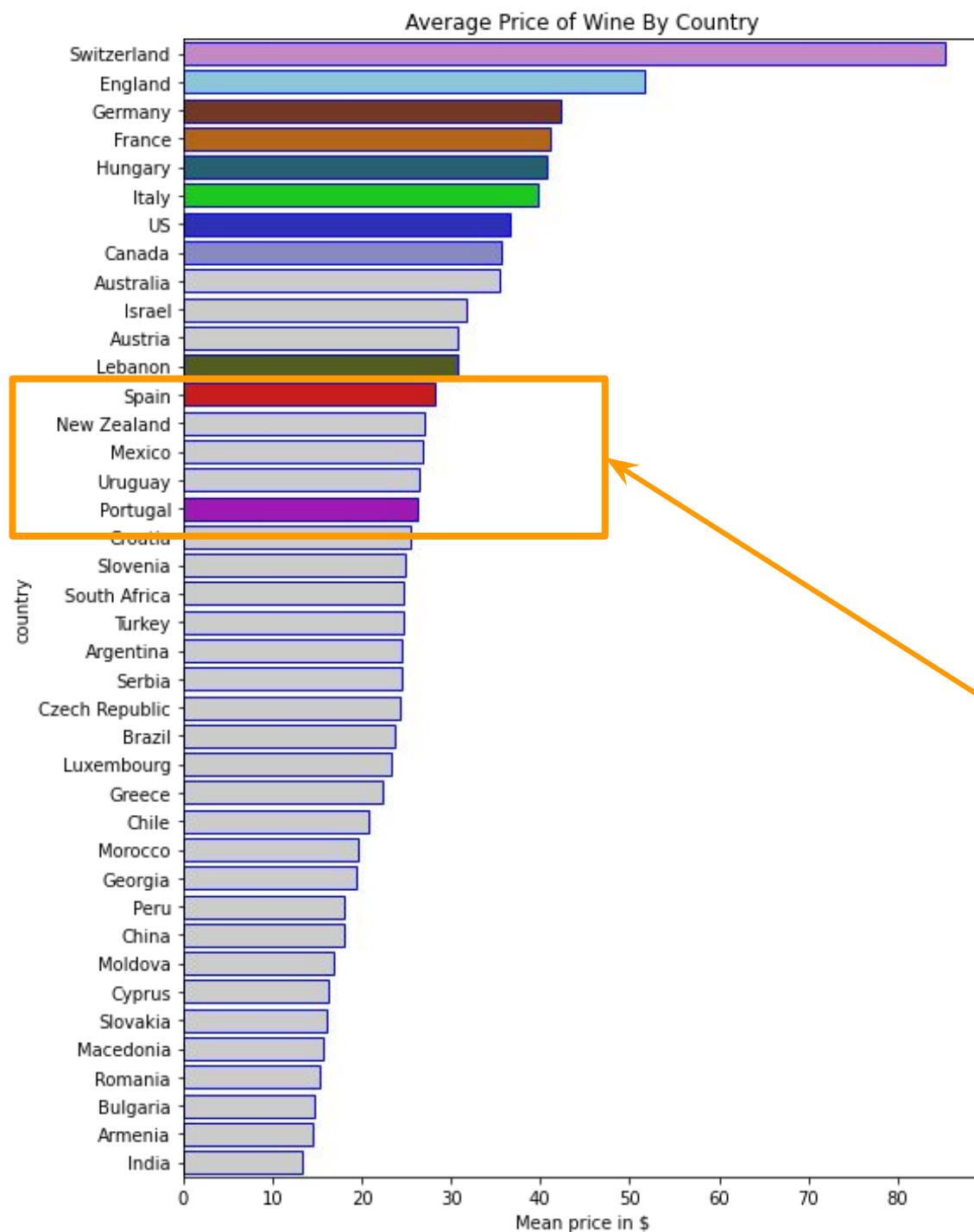


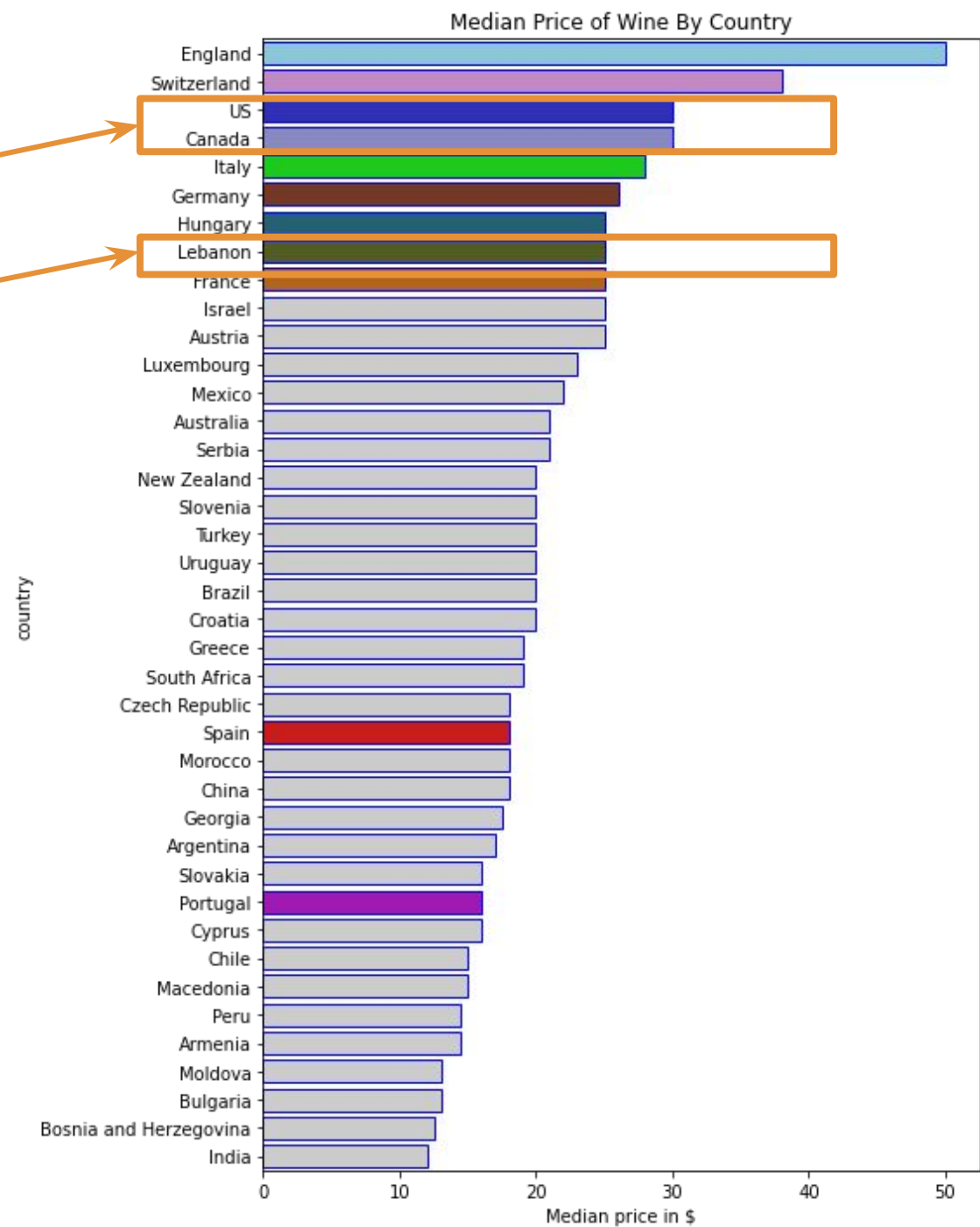
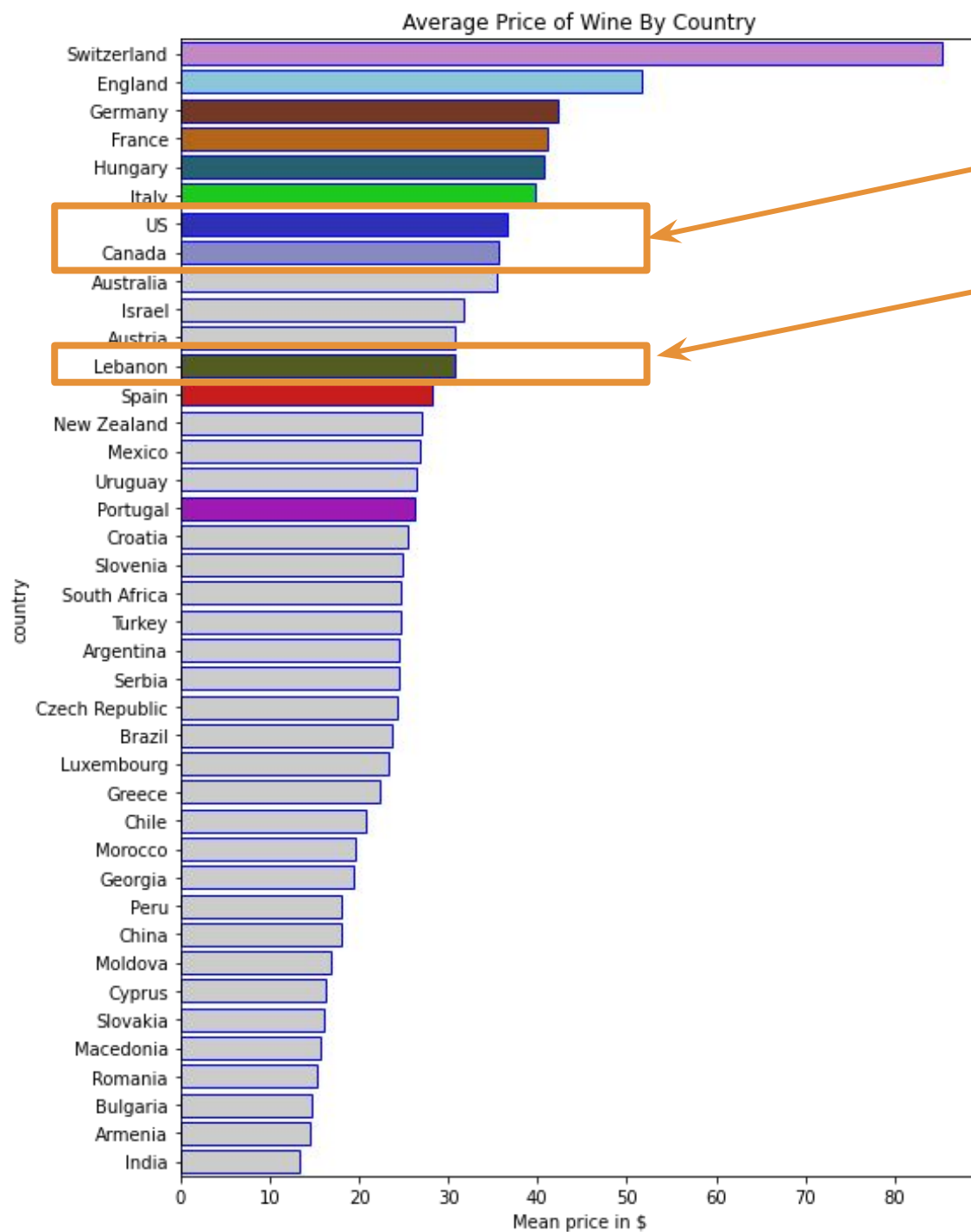
# Wine Counts in US

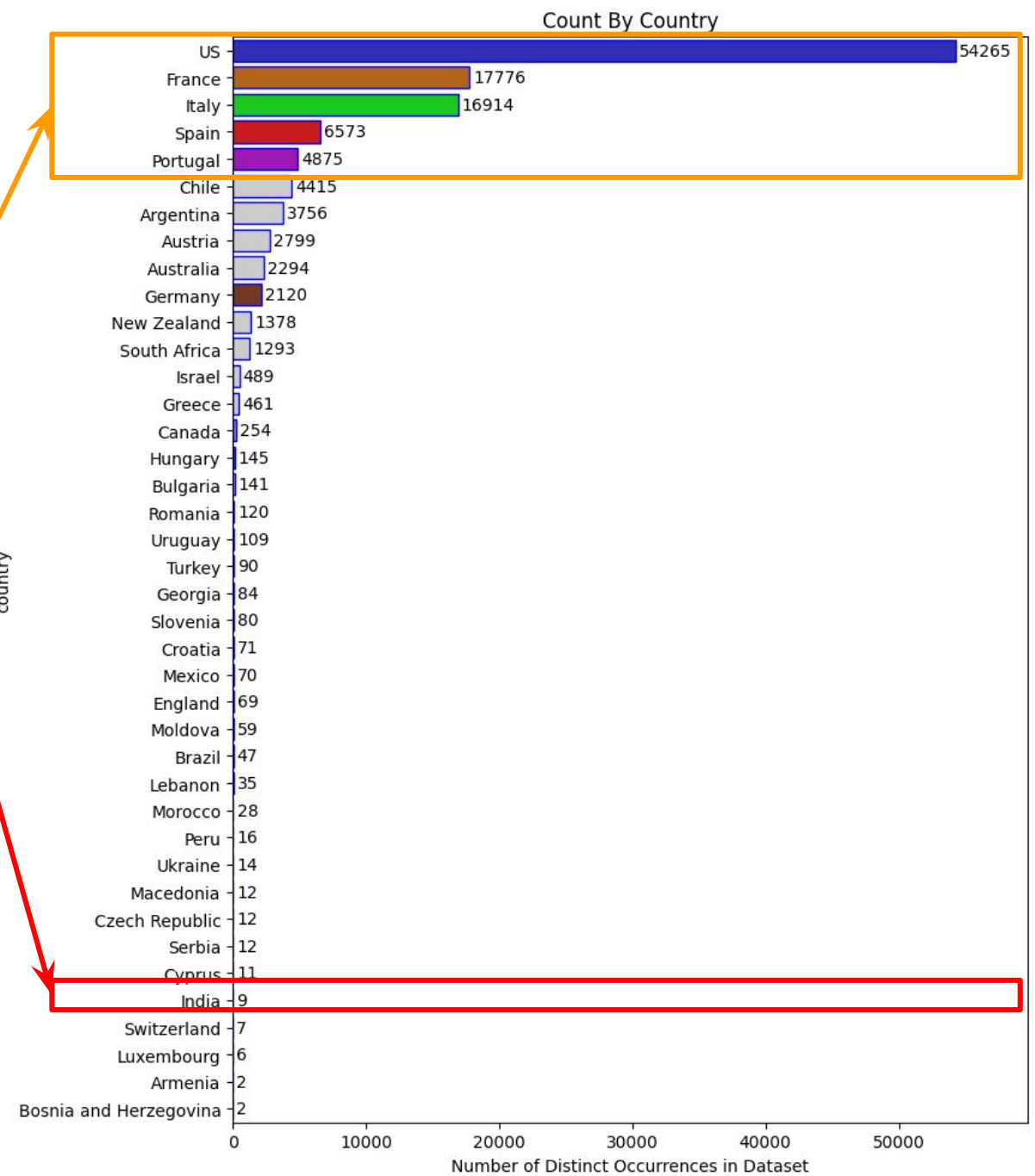
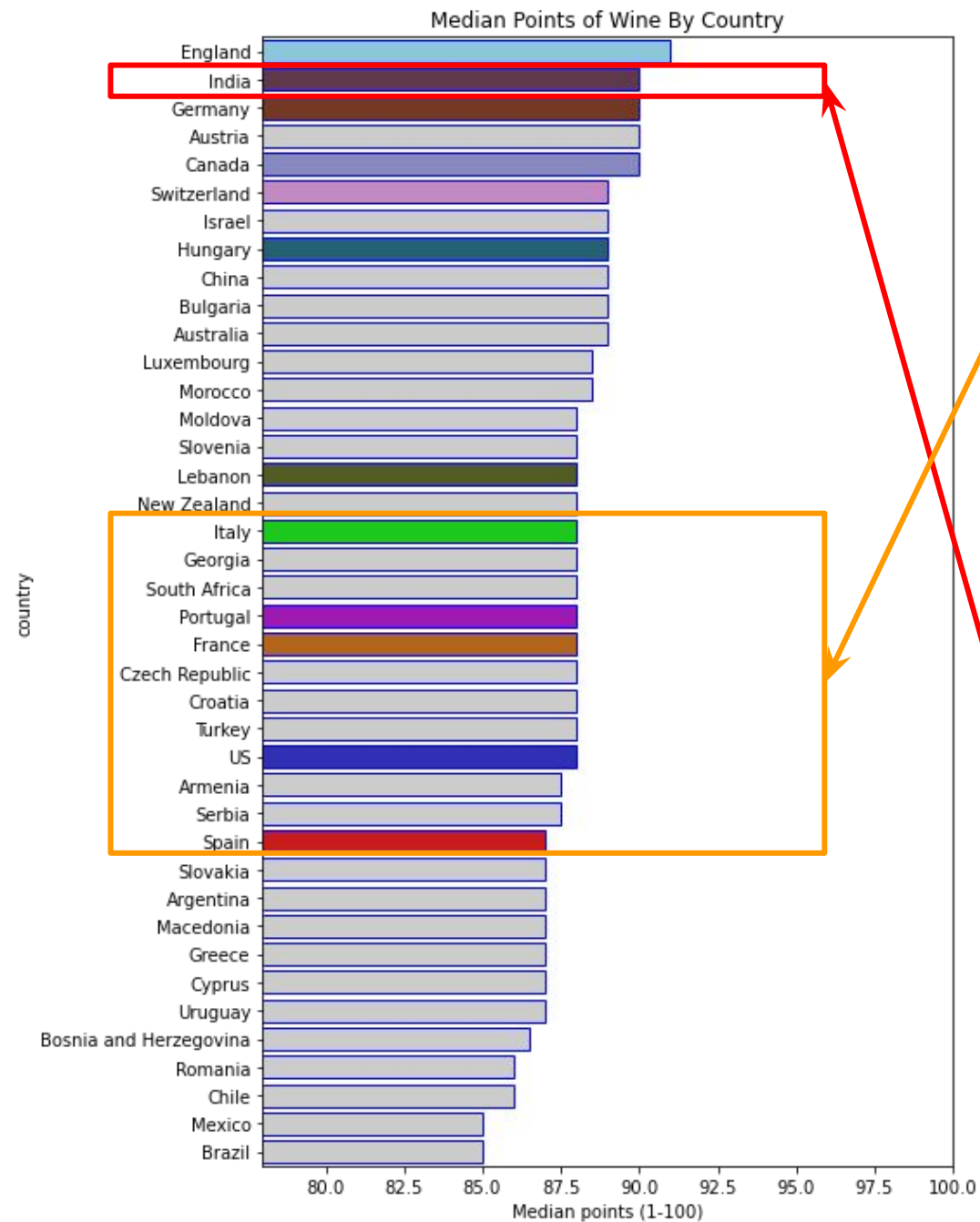


Average Price of Wine By Country





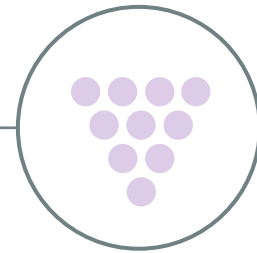
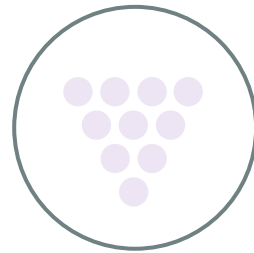


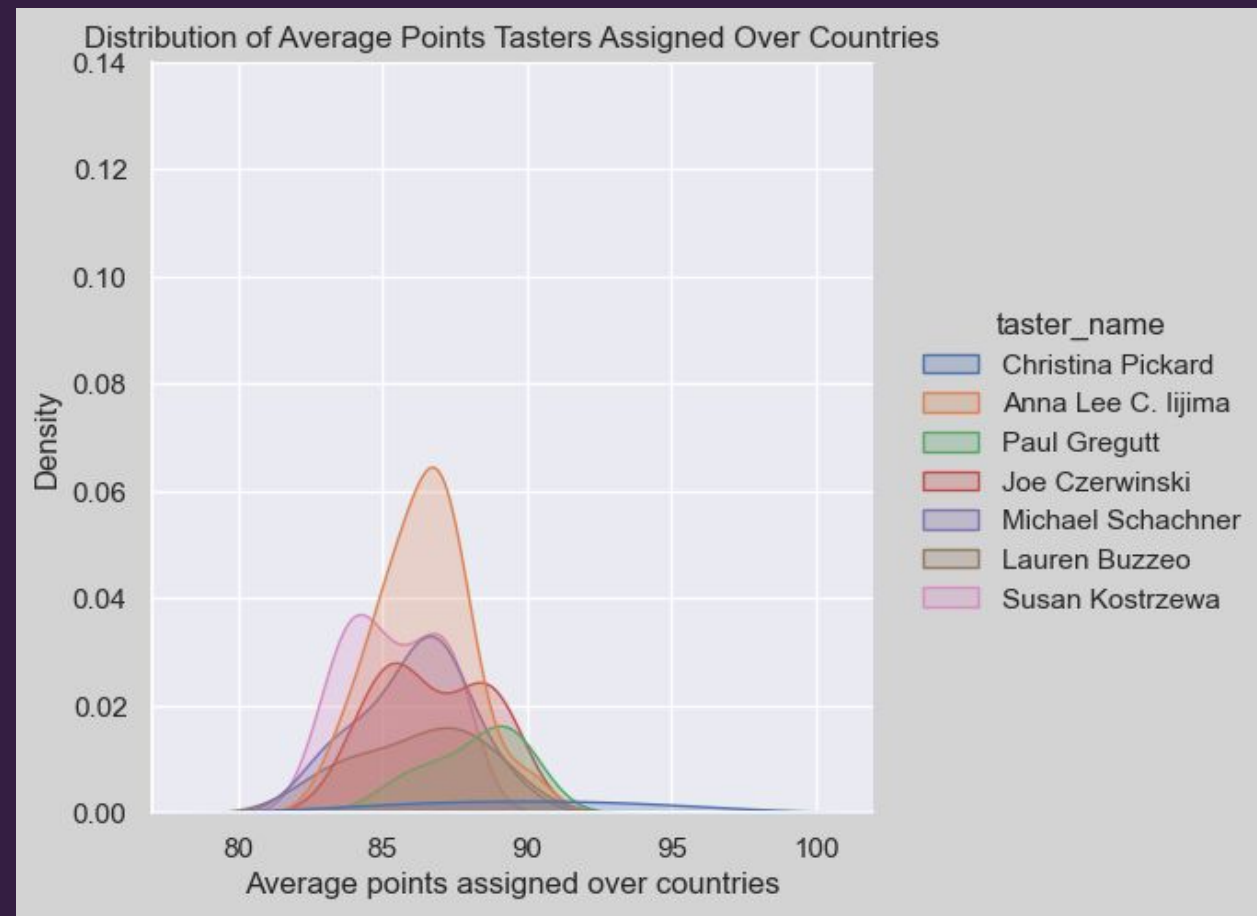
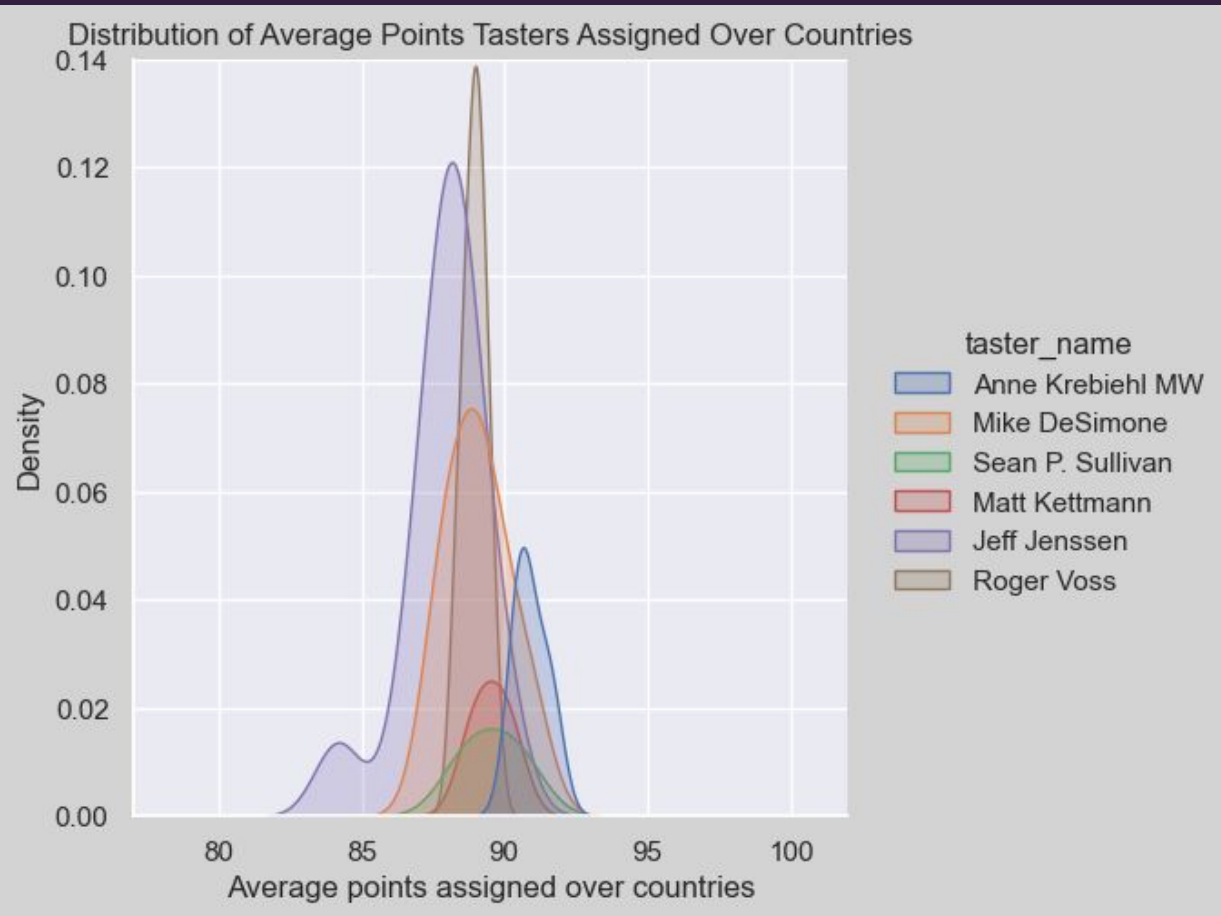




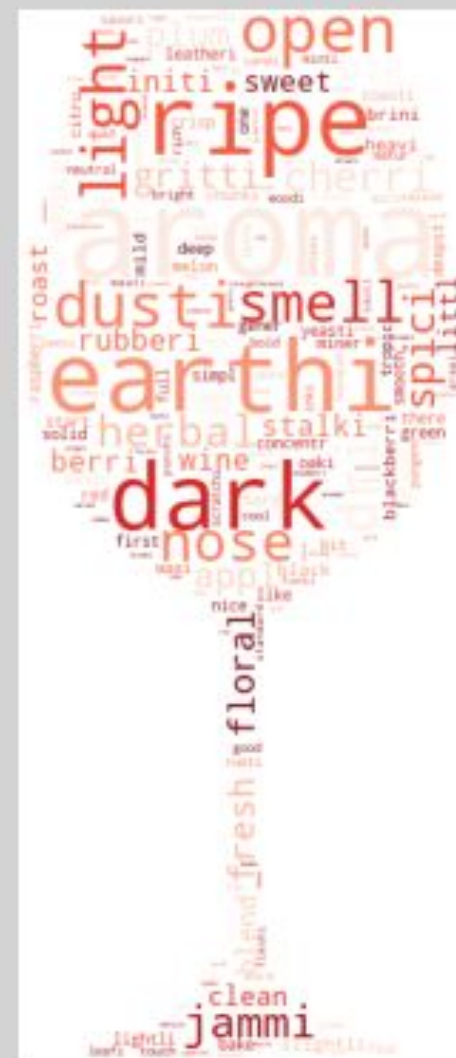
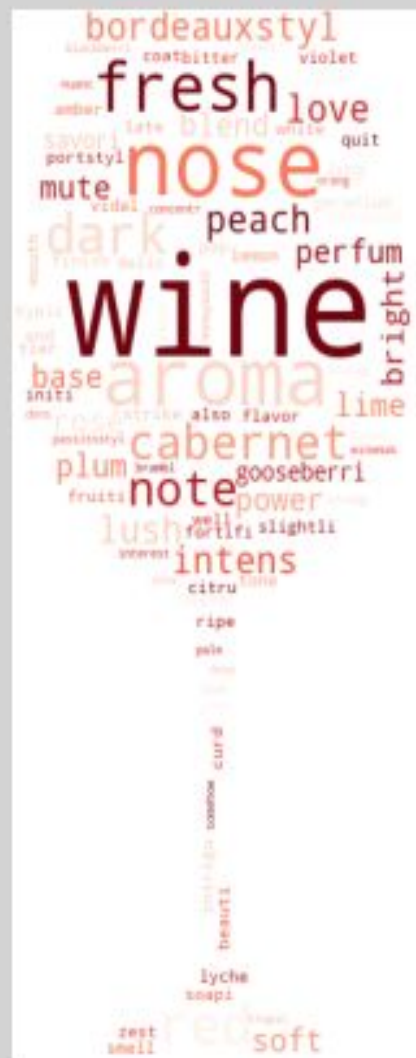
# Reviews' Analysis

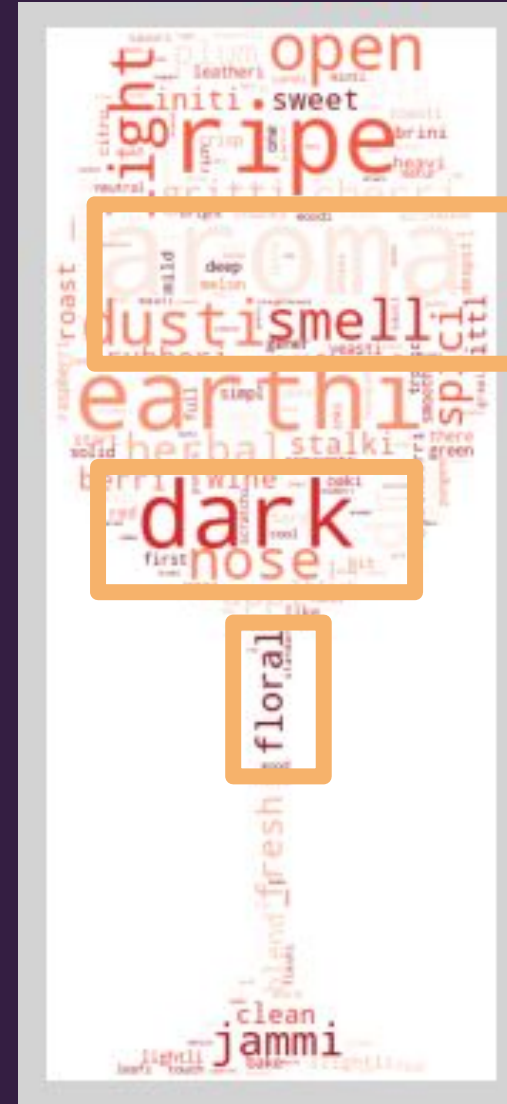
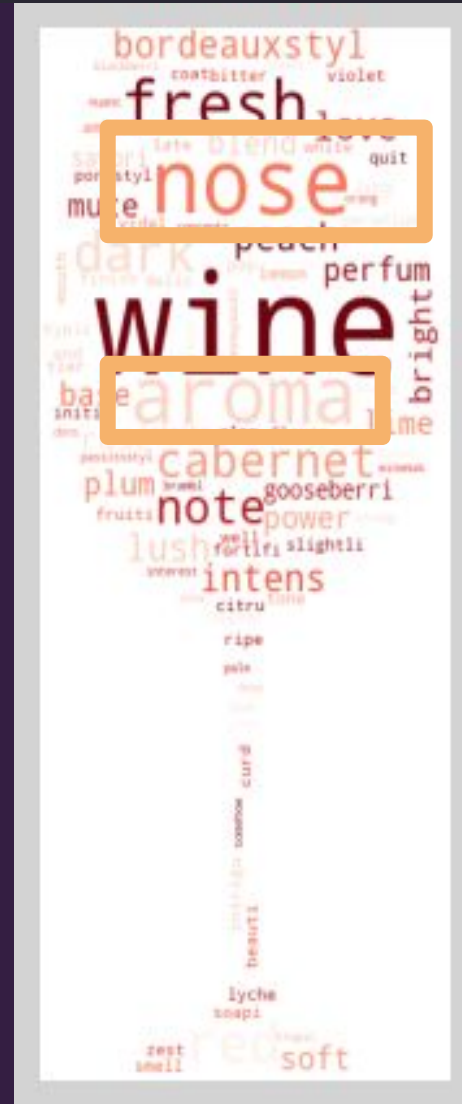
Word Frequency based on reviewers' descriptions  
& Rating Points





Notice the widths, even for tasters like Roger Voss with reviews in multiple countries!

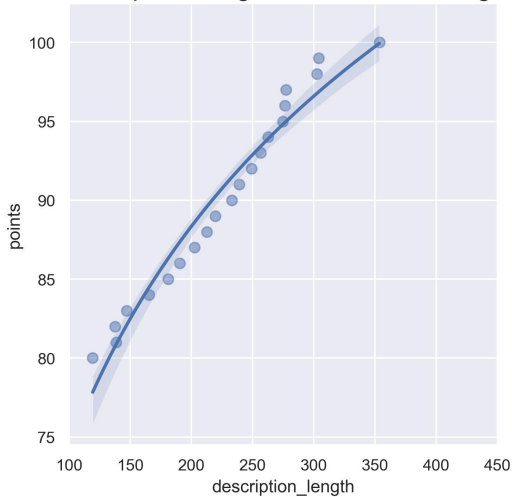




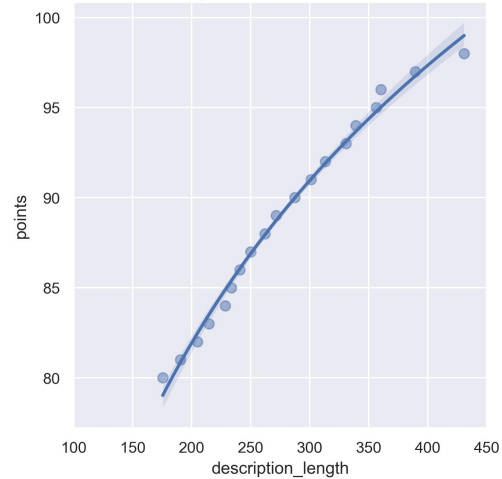


# Taster System Analysis

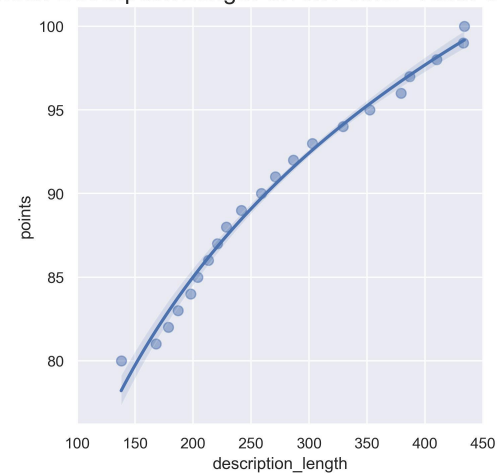
Mean Description Length Given Points - Roger Voss



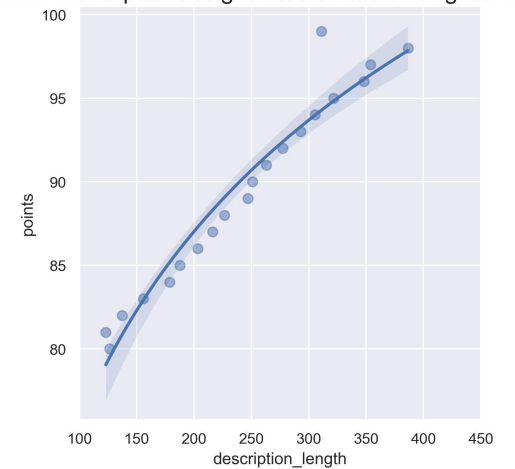
Mean Description Length Given Points - Michael Schachner



Mean Description Length Given Points - Kerin O'Keefe



Mean Description Length Given Points - Virginie Boone

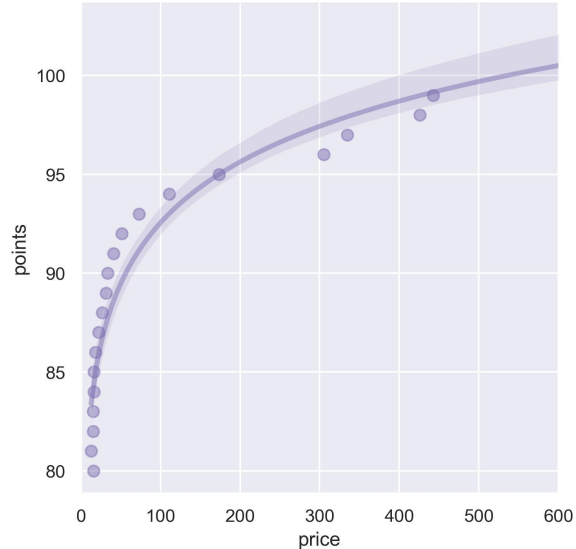


**Tasters tend to write longer descriptions when they give higher points!**

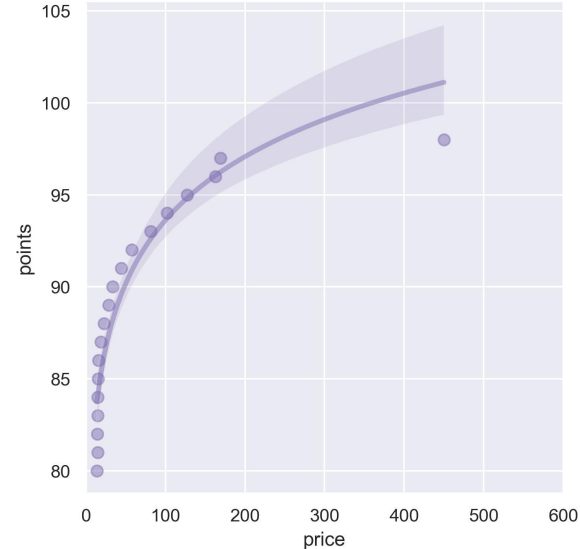


# Taster System Analysis

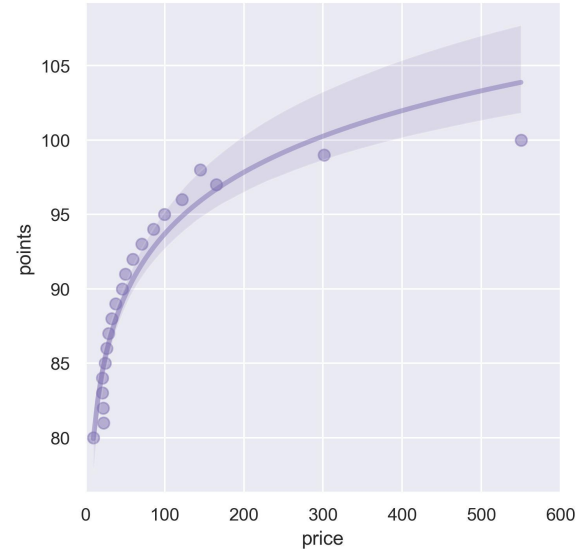
Mean Price Given Points - Roger Voss



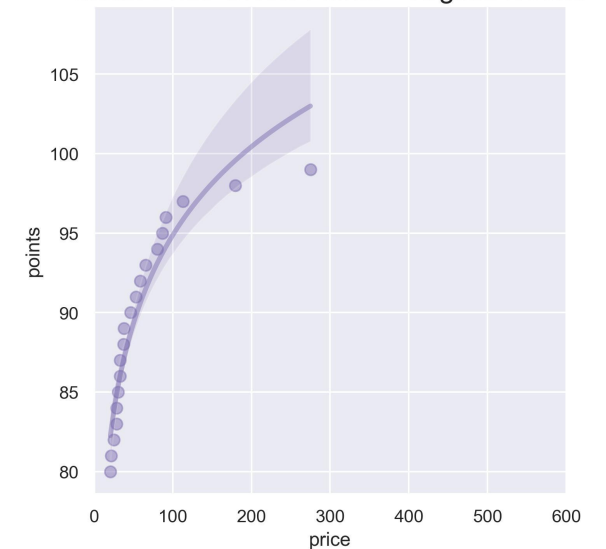
Mean Price Given Points - Michael Schachner



Mean Price Given Points - Kerin O'Keefe



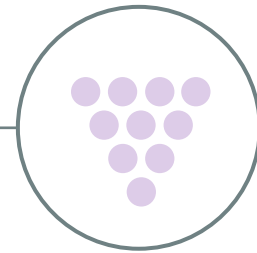
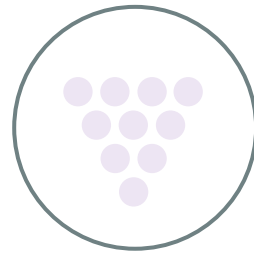
Mean Price Given Points - Virginie Boone



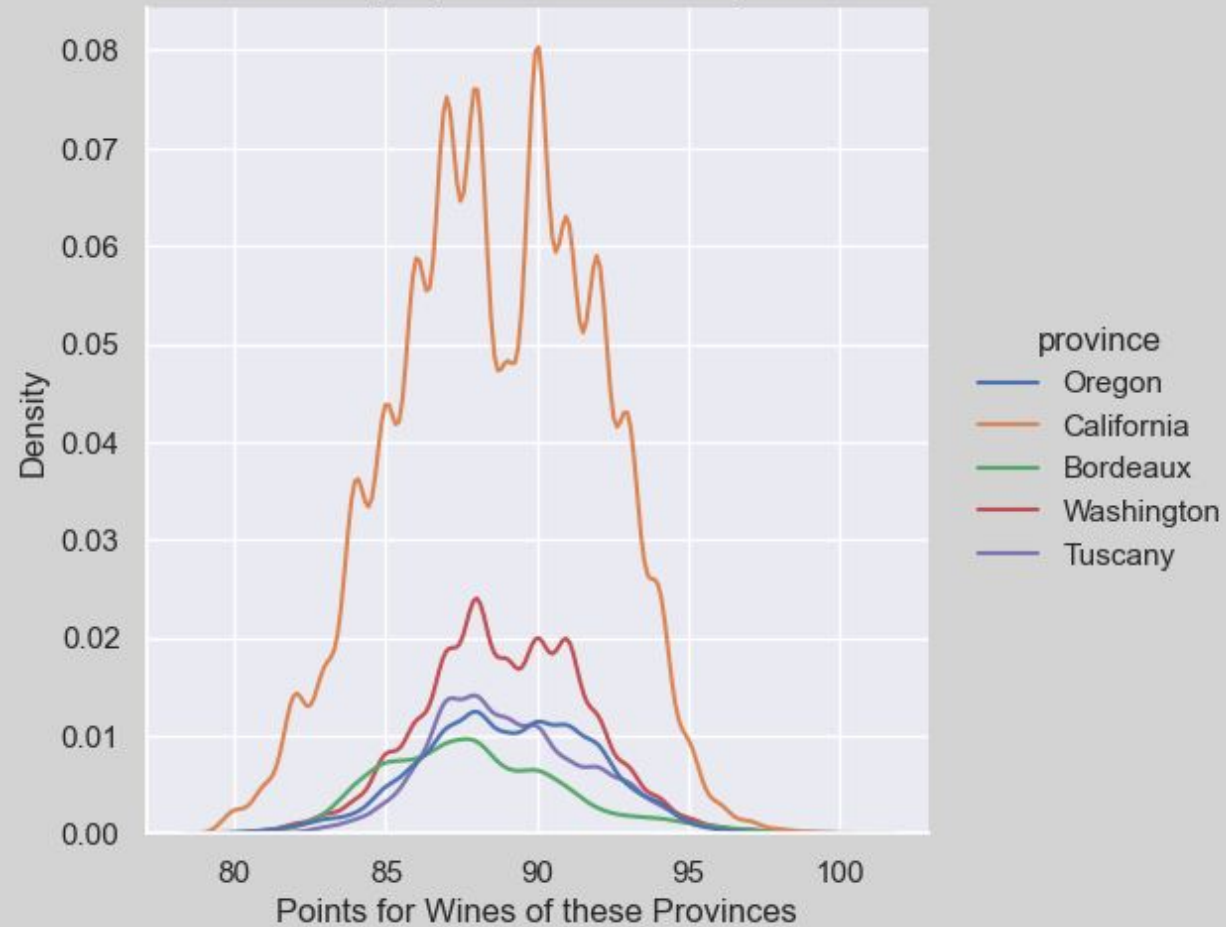
More expensive: more points, but diminishing returns!

# Overall Trends and Prediction

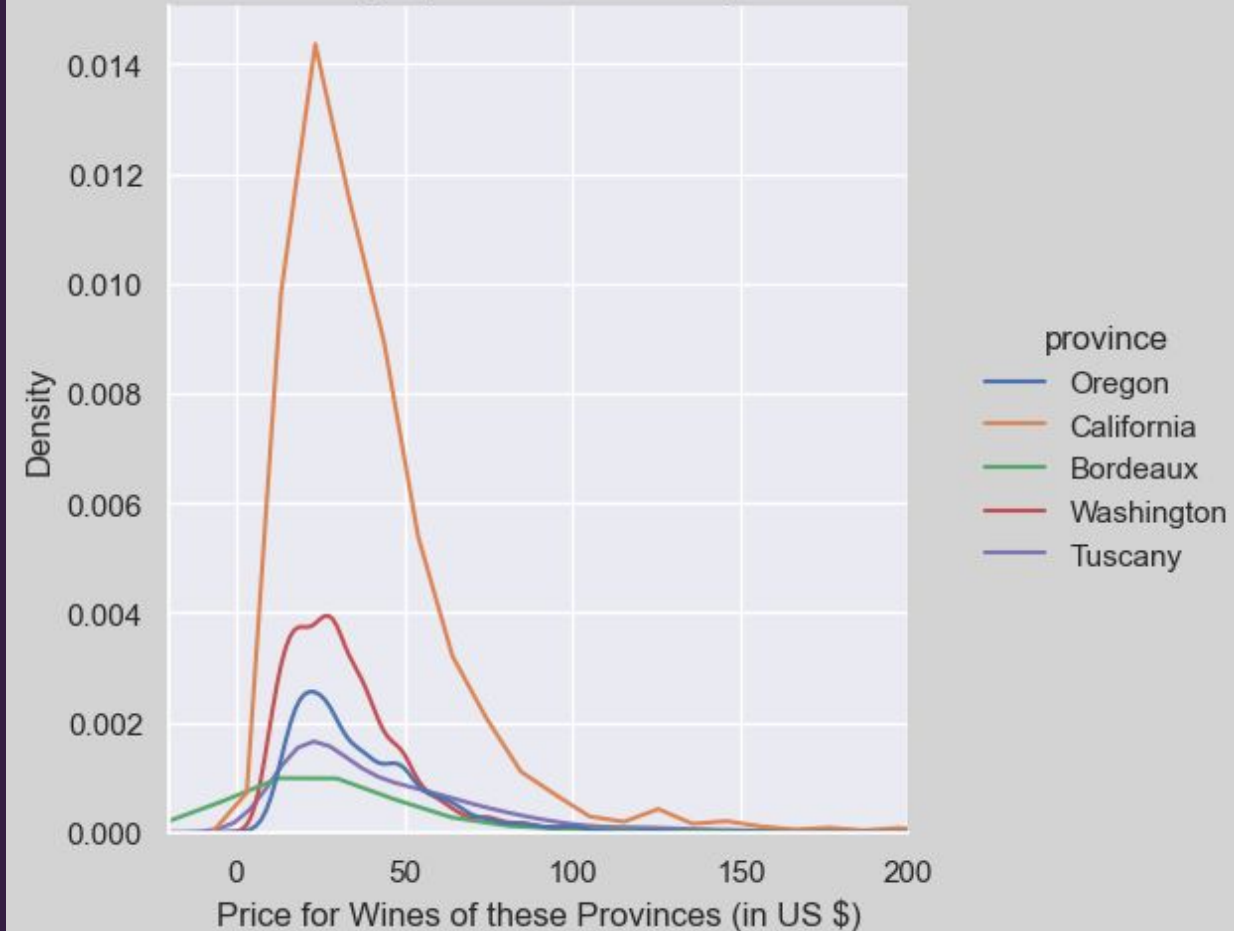
Points vs Price and their Prediction



Points Distribution By Top 5 Most Numerous Rated Provinces

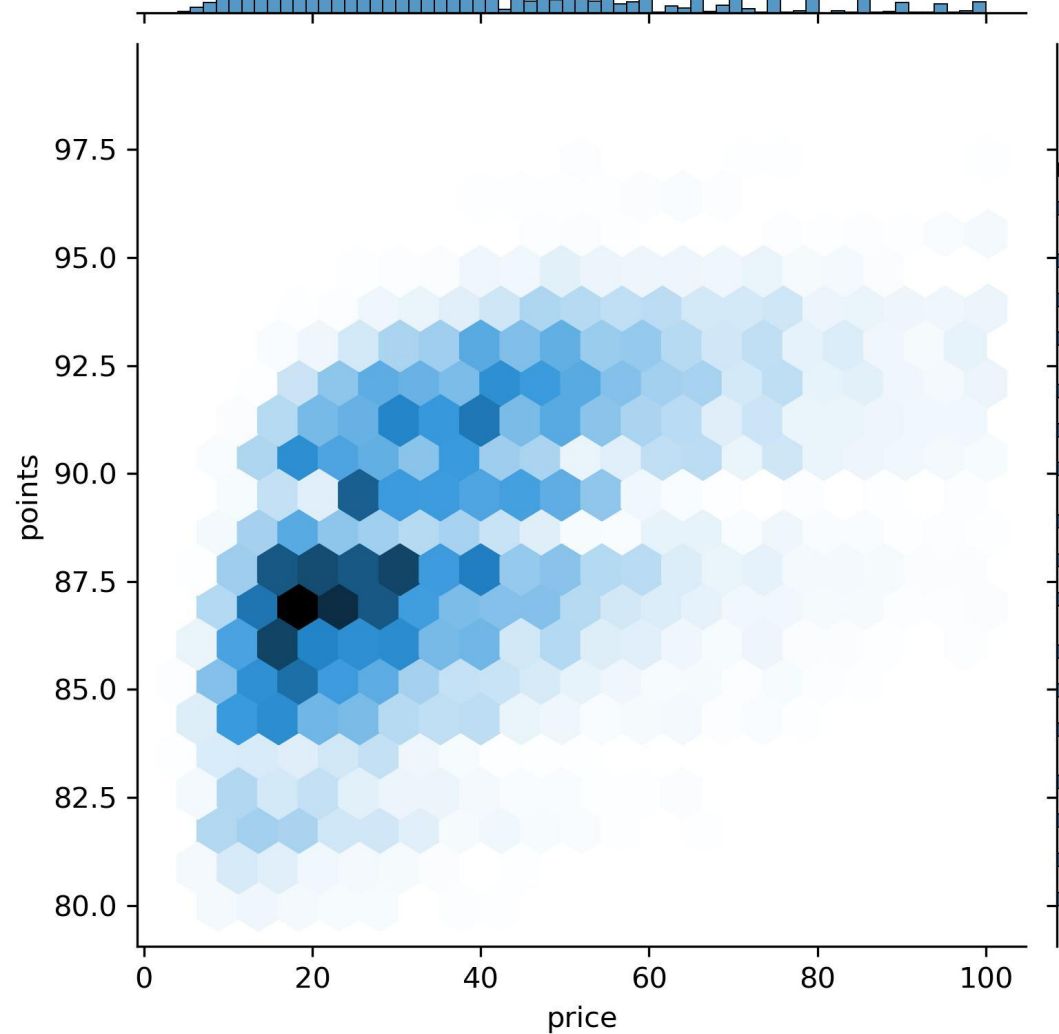


Price Distribution By Top 5 Most Numerous Rated Provinces

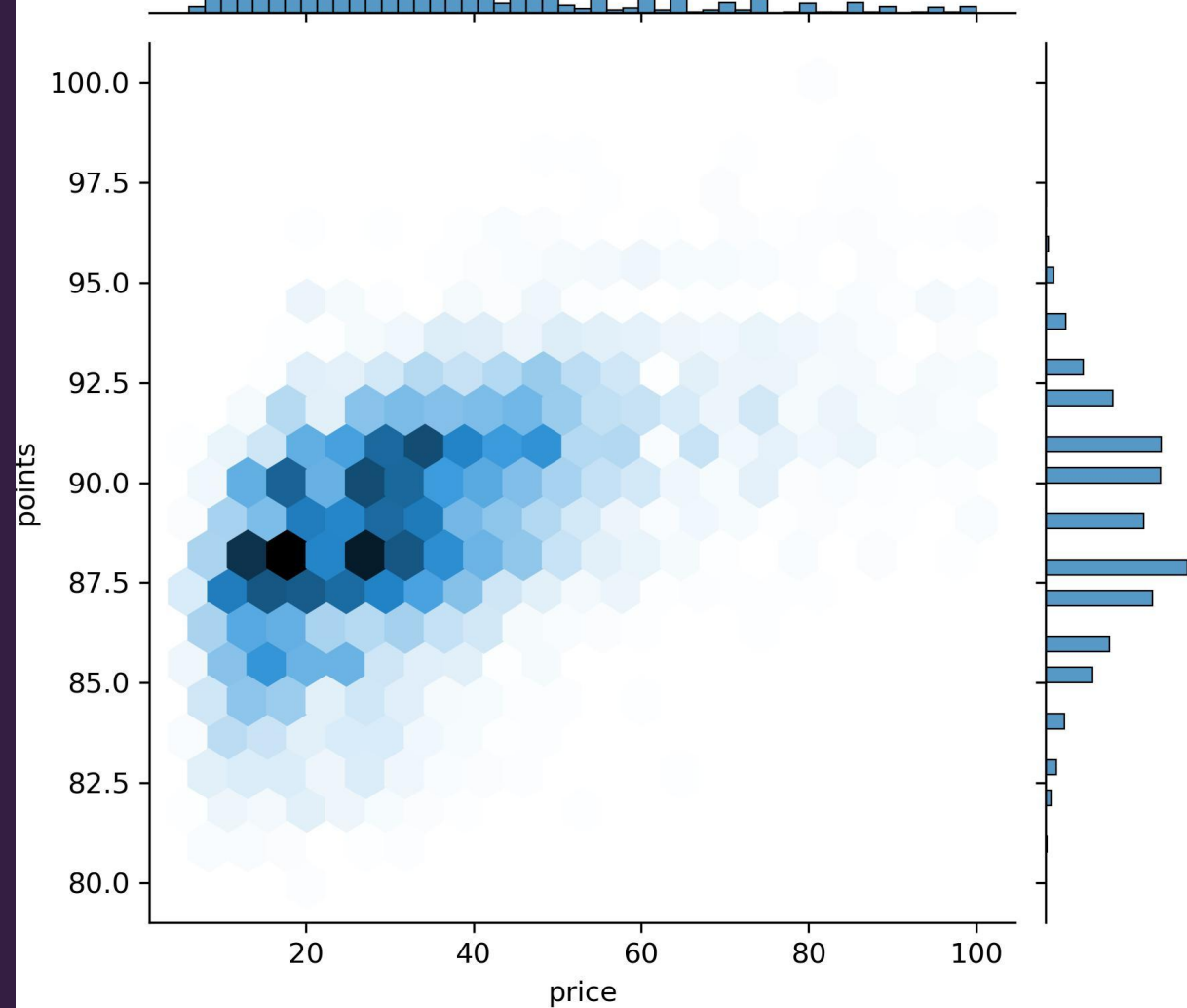


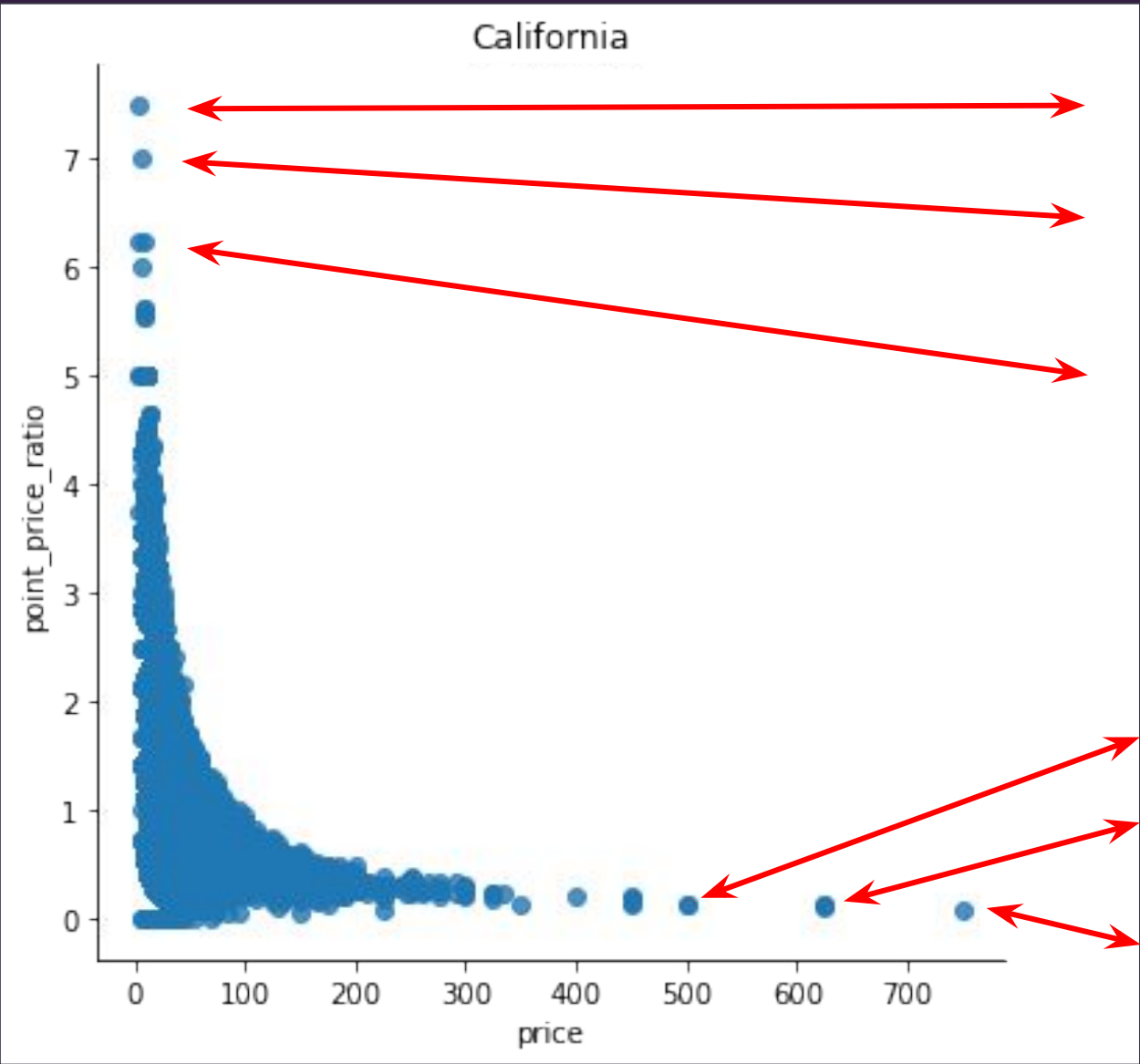
**Notice the general shapes: bell vs skewed right!**  
**This means more weighting of points toward lower price.**

Hex grid of price and point in California



Hex grid of price and point in Washington



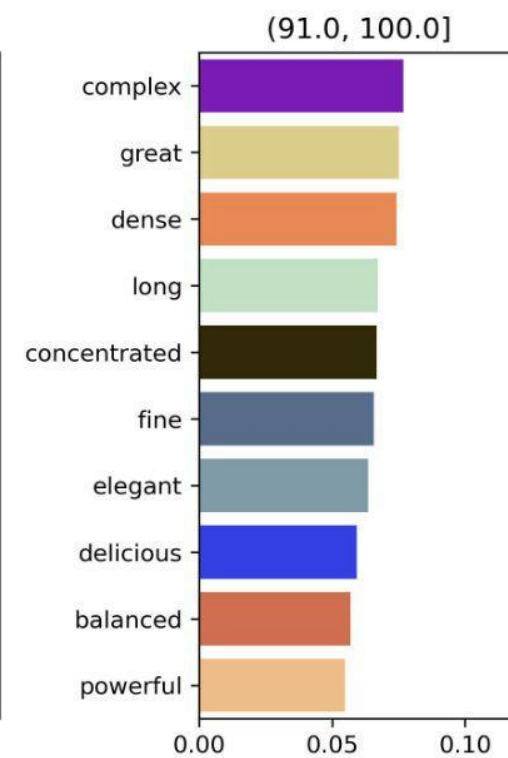
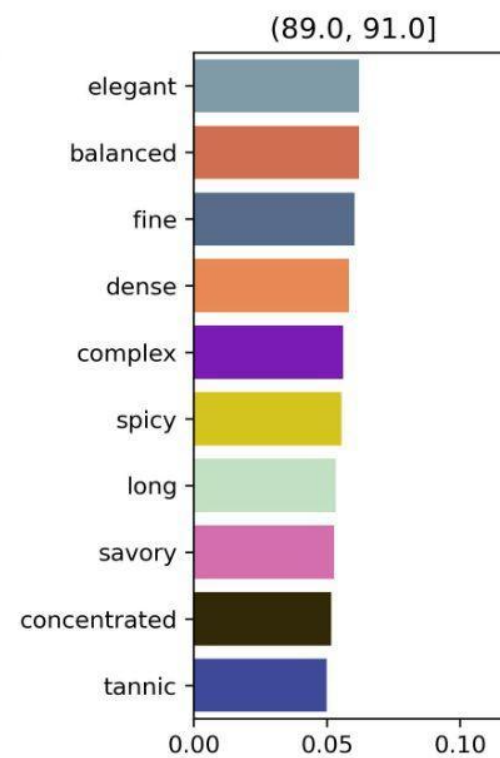


country	designation	province	variety	title	points	price	point_price_ratio
US	None	California	Merlot	Bandit NV Merlot (California)	86	4.0	7.50
US	Fizz	California	Sparkling Blend	Mancan NV Fizz Sparkling (California)	87	5.0	7.00
US	None	California	Chardonnay	Esser Cellars 2001 Chardonnay (California)	90	8.0	6.25
US	None	California	White Blend	Dancing Coyote 2015 White (Clarksburg)	85	4.0	6.25
US	Double Plus Estate Grown	California	Cabernet Sauvignon	Hewitt 2013 Double Plus Estate Grown Cabernet ...	93	500.0	0.130
US	Family Reserve	California	Cabernet Sauvignon	Yao Ming 2011 Family Reserve Cabernet Sauvigno...	94	625.0	0.112
US	Intrepid	California	Syrah	Law 2013 Intrepid Syrah (Paso Robles)	92	750.0	0.08



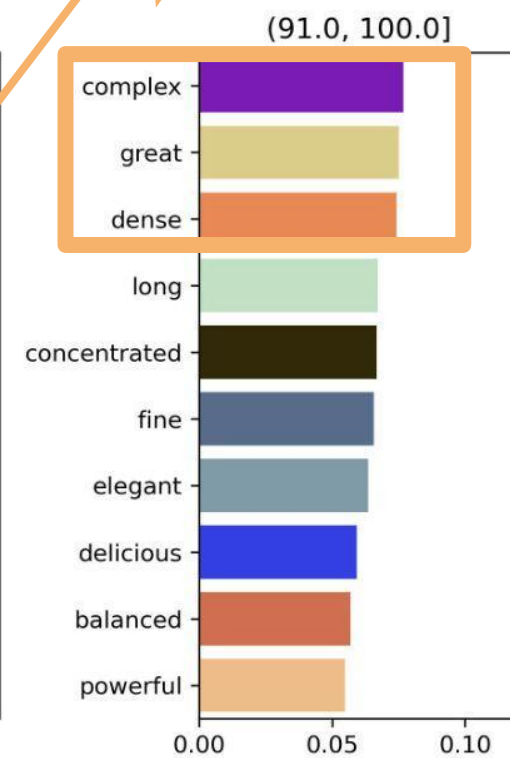
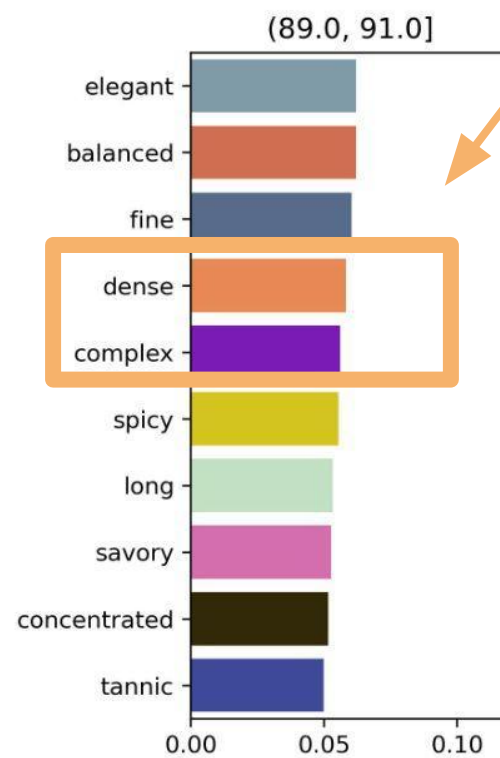


## Top Descriptive Words





Top Descriptive Words





Top Descriptive Words

(80.0, 86.0]



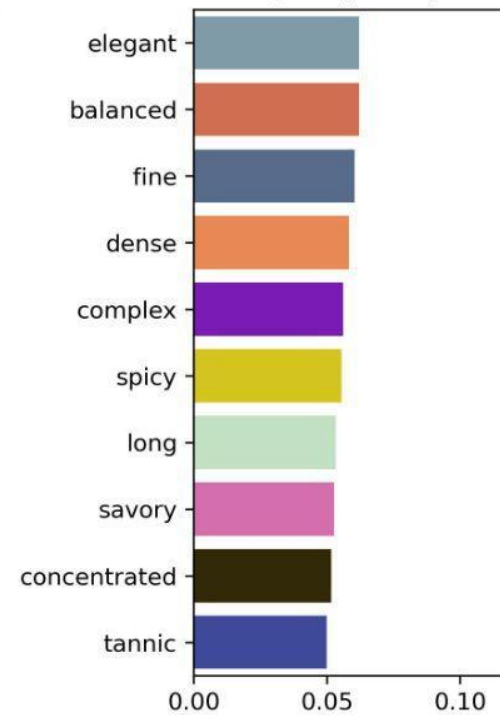
(86.0, 88.0]



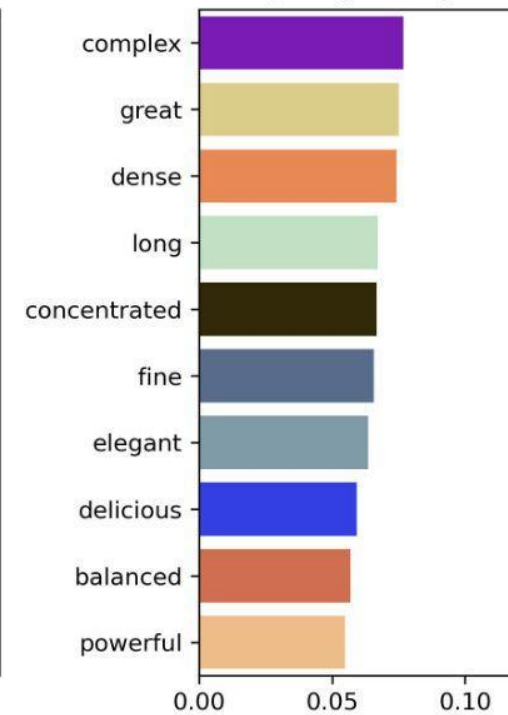
(88.0, 89.0]



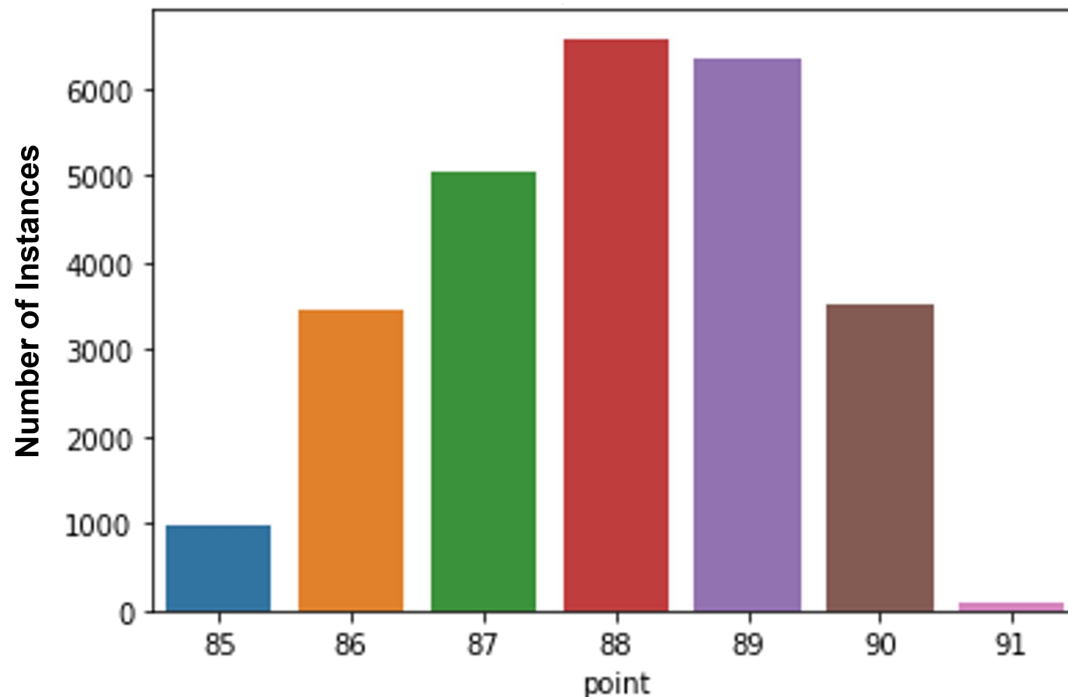
(89.0, 91.0]



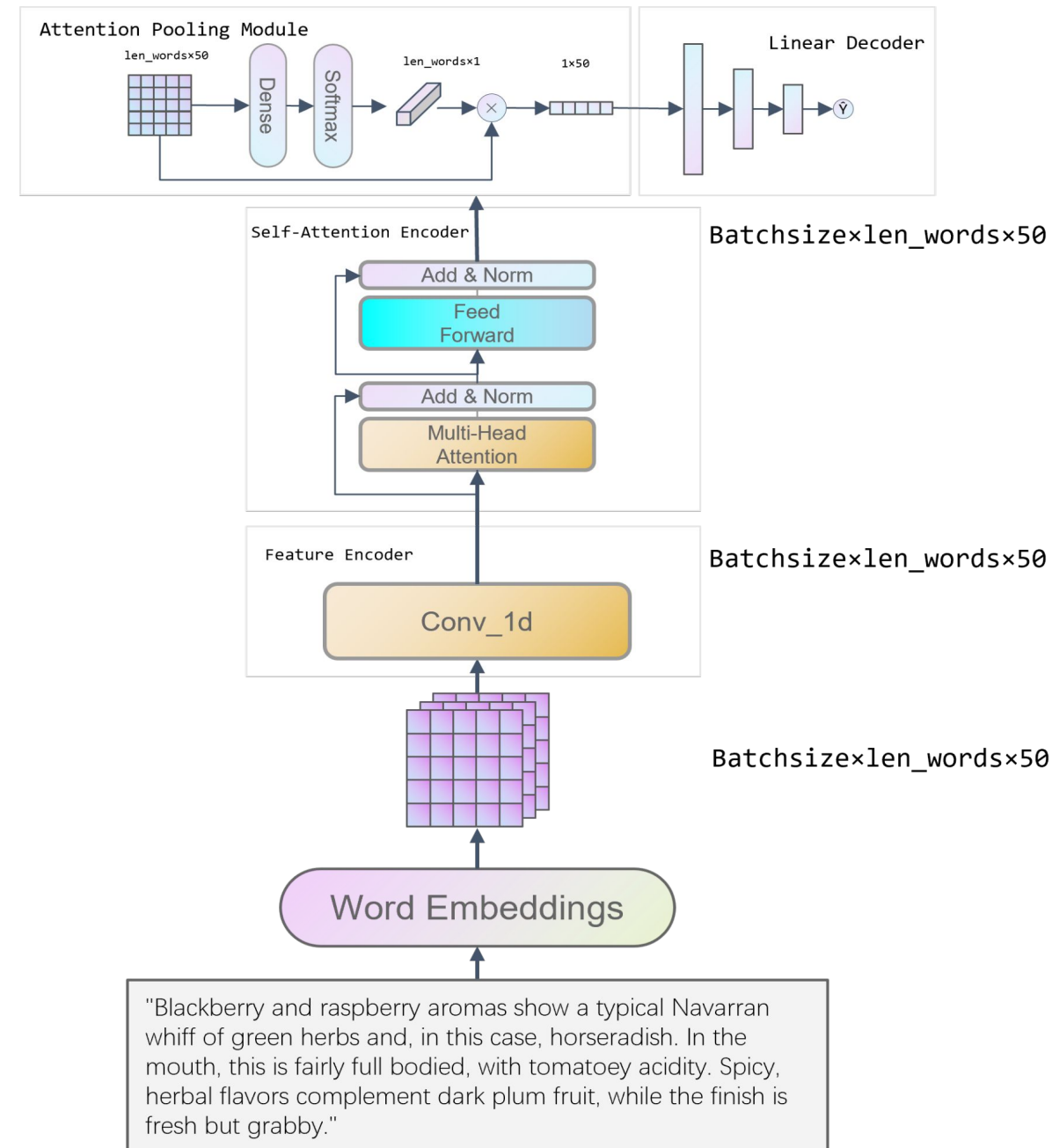
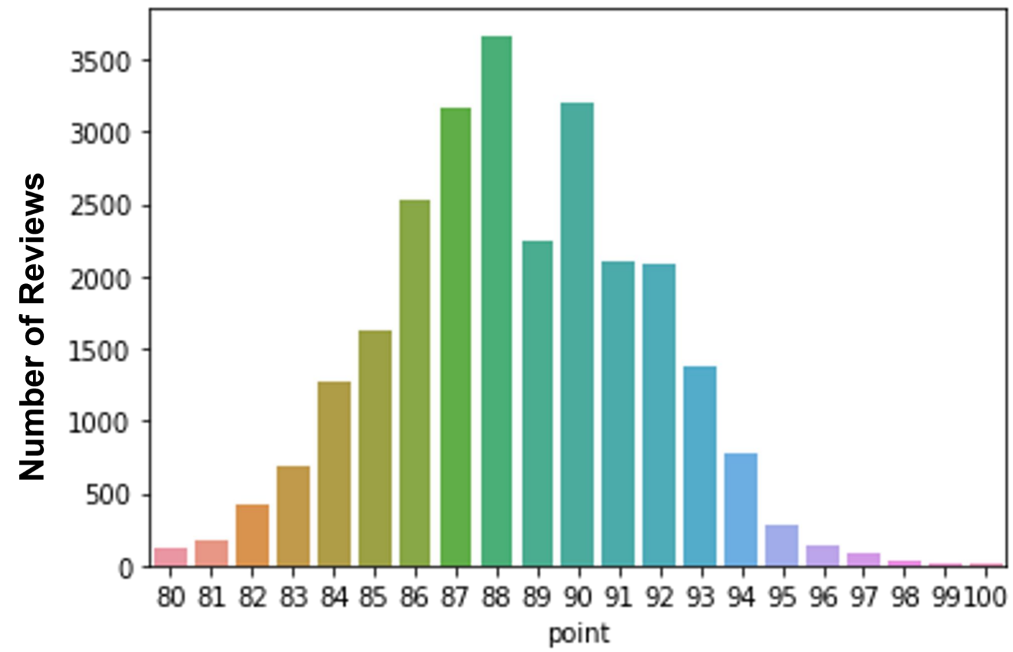
(91.0, 100.0]



Predicted Point Distribution



Actual Point Distribution





The background of the slide features a close-up, slightly blurred image of several wine corks. A semi-transparent purple rectangle is overlaid on the right side of the image. The text "Thank you" is written in a large, white, serif font, with "Thank" on the top line and "you" on the bottom line, positioned to the left of the purple rectangle.

**Thank  
you**

**Any questions?**