

# PRICING POWER IN WINE

---

BRADLEY LEVERGOOD

SEPTEMBER 17 2018

# MOTIVATION

---

## **For the large-scale wine purchaser:**

- ❖ For how much can you sell a given bottle of wine?
- ❖ What is a fair price to purchase wine shipments such that you maintain a good profit margin?

## **For Brad the average consumer:**

- ❖ How much should you pay for a given bottle?

# THE DATASET

---

A collection of wine reviews scraped from the **Wine Enthusiast Magazine's** website

❖ **14** Columns:

- **Country**
- Description
- **Designation**
- Points
- Price
- **Province**
- **Region I category**
- **Region 2 category**
- **Taster (reviewer)**
- **Taster twitter-handle**
- Title
- **Vintage**
- **Variety**
- **Winery**

❖ **129,971** rows

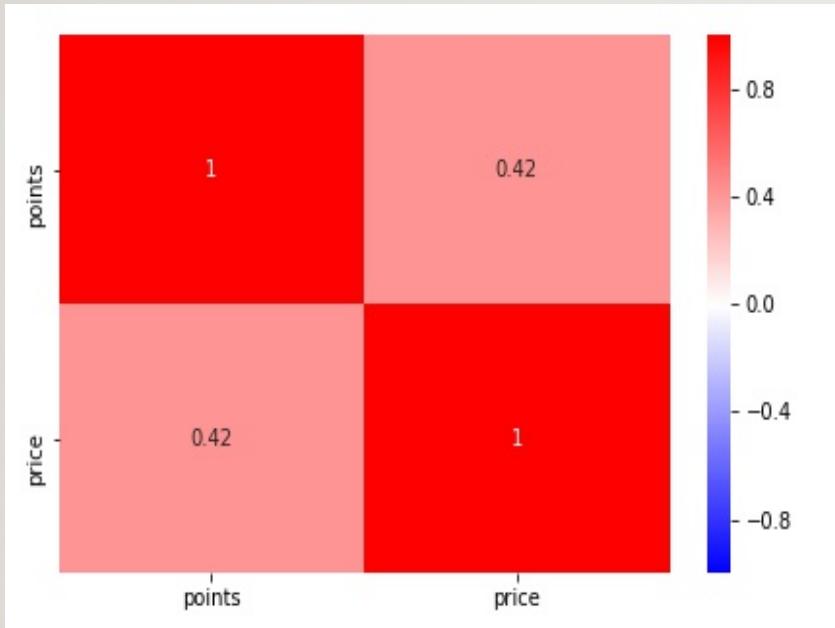
Assumptions:

- ❖ All null **country**, **variety**, and **price**-value entries are dropped
- ❖ All null **region**, **taster name**, **twitter handle**, and **vintage** values are replaced with place-holder values

Final Dataset size before model evaluation has **111,537** entries\*

# PRICE CORRELATIONS

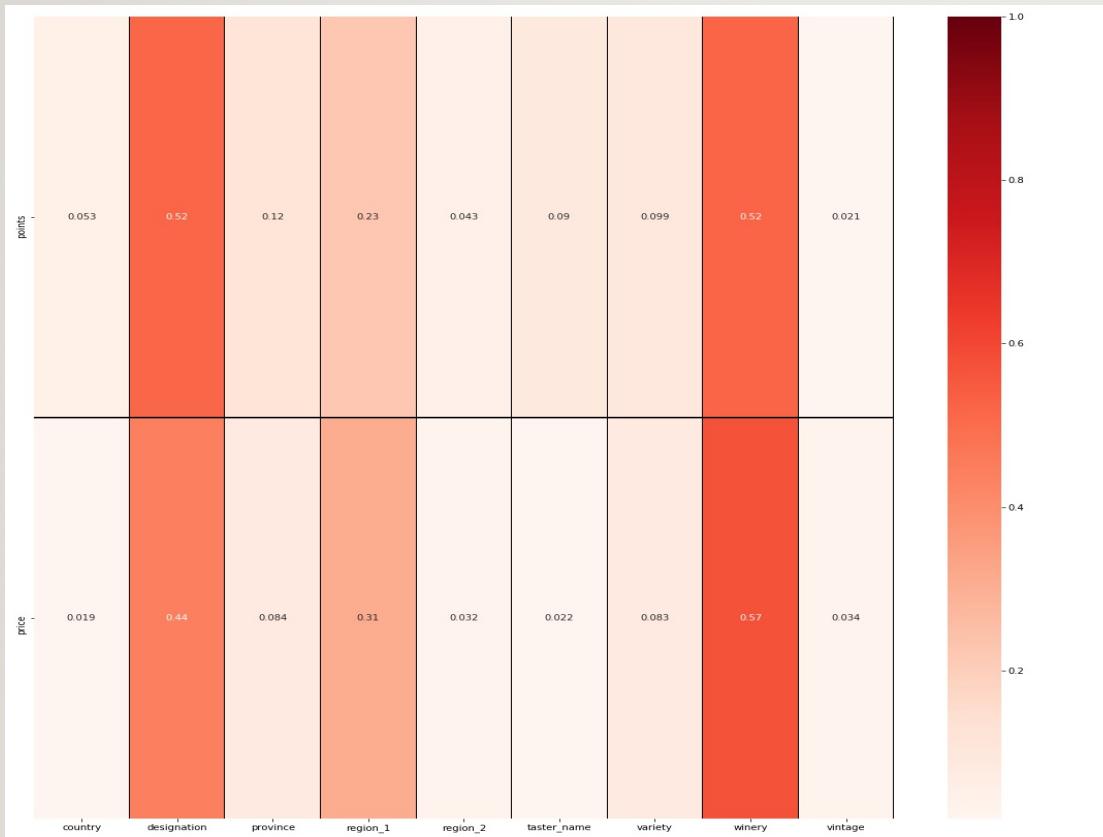
---



- ❖ Shows a moderate correlation between ratings and points
  
- ❖ Missing A LOT of information from categorical variables

# PRICE ASSOCIATIONS WITH CATEGORICAL INFORMATION

---

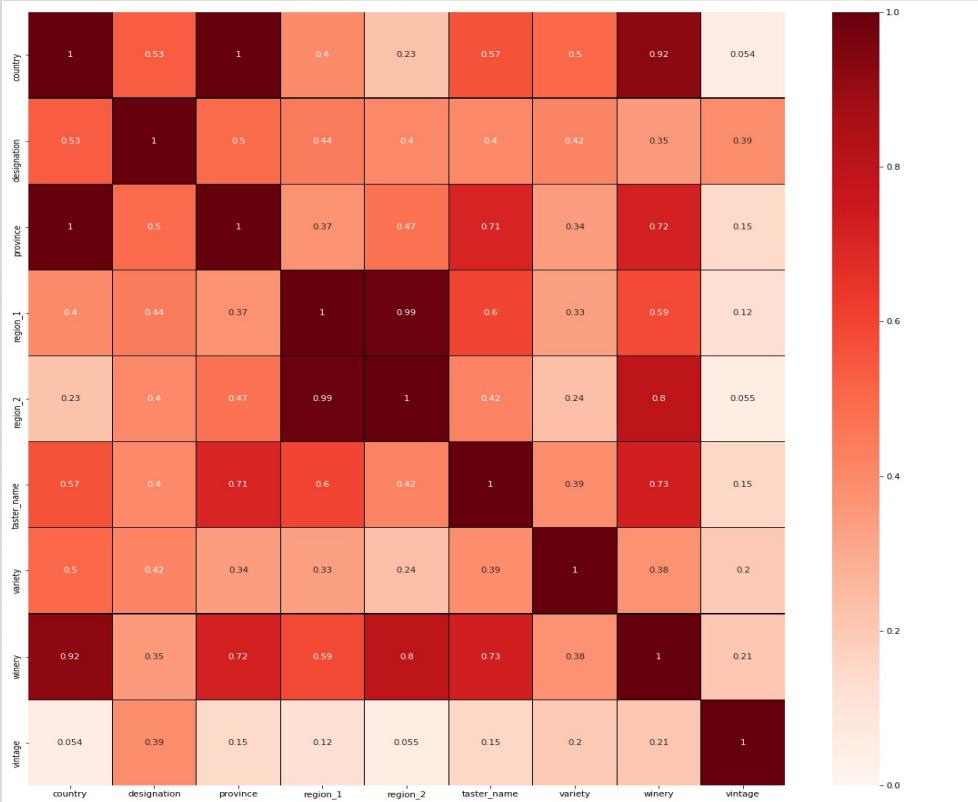


**Strongest Associations with price:**

- ❖ Winery
- ❖ Designation
- ❖ Region (larger)

# ASSOCIATIONS BETWEEN CATEGORICAL CHARACTERISTICS

---

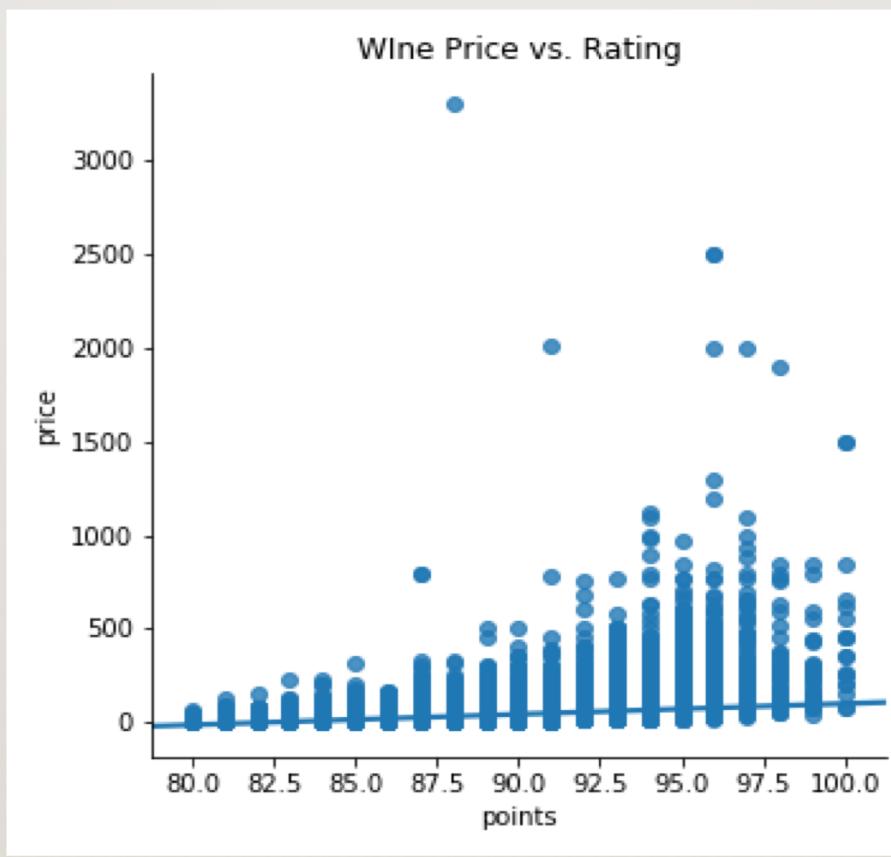


## Important Associations:

- ❖ Winery – Region: **0.59**
- ❖ Winery – Designation: **0.35**
- ❖ Region – Designation: **0.44**

# MODELING – LINEAR MODELS

---



# LINEAR MODEL SCORES

---

- **Linear Model Scores:**

- ❖ **Null RMSE: 40.23\***

- ❖ Designation

- RMSE: 40.98
    - R-squared: **0.04**

- ❖ Region I

- RMSE: 36.99
    - R-squared: **0.20**

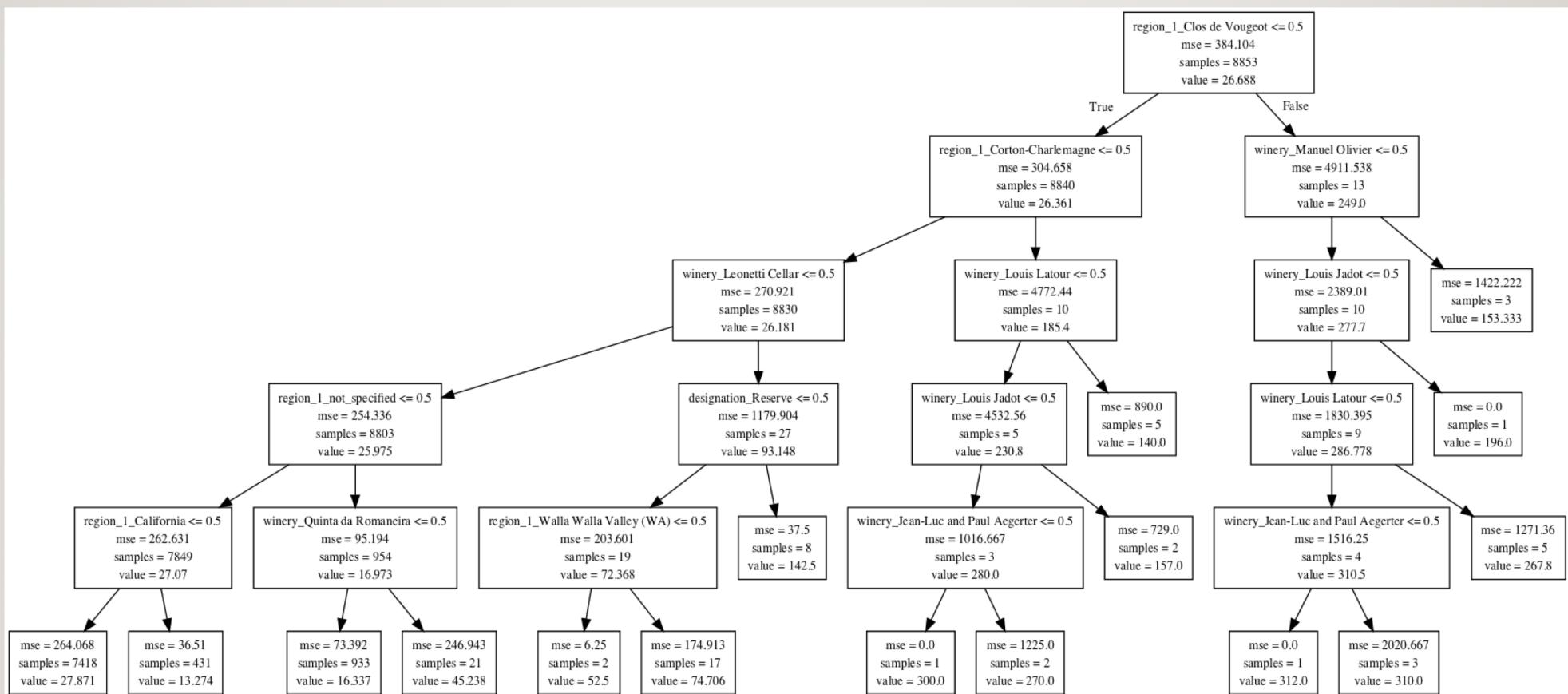
- ❖ Winery

- RMSE: 30.32
    - R-squared: **0.34**

- ❖ Combination

- RMSE: 14020216702308.39
    - R-squared:- **1.05e+24**

# MODELING – RANDOM FOREST MODEL



# RANDOM FORREST TRIALS

---

❖ Null RMSE: 17.66594385755445

❖ Designation, Region, Winery:

- RMSE: 8.08
- R-squared: **0.827**

❖ Designation, Region, Winery, Points:

- RMSE: 8.95
- R-squared: **0.74**

❖ All Available Features:

- RMSE: 8.07
- R-squared: **0.830**

# NEXT STEPS

---

- ❖ Run all forest models without data elimination
- ❖ Perform sentiment analysis to predict rating and price
- ❖ Identify descriptors associated with the highest rated wines for vendors

---

**Questions?**

# REFERENCES

---

- <https://www.kaggle.com/zynicide/wine-reviews>
- <https://www.winemag.com/>
- <https://towardsdatascience.com/the-search-for-categorical-correlation-a1cf7f1888c9>