

# WINE PRICING

Author: Steve Diamond



# SITUATION

## ➤ Background

- Data Science team at Princeton Wine Club
- Dedicated to drinking great red wine
- Funds are not endless (even in Princeton)

## ➤ Our Challenge:

- Determine what drives the price of wine:
  - Country, variety, vintage, reviews, etc.?
- Build a predictive model that can estimate what a wine should cost

## ➤ Our Solution

- Analyze a set of wine data from Wine Enthusiast Magazine
- Examine the raw numbers to extract some broad understanding
- Use regression modeling techniques to:
  - Quantify the effect of various factors on price
  - Create predictive model



# WORKFLOW

Data  
Gathering

Cleaning/  
EDA

Cluster  
Analysis

Modeling

Analysis

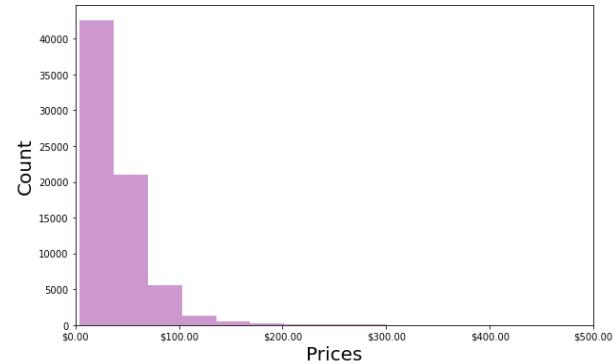
Conclusions

# EXPLORATORY DATA ANALYSIS (EDA)

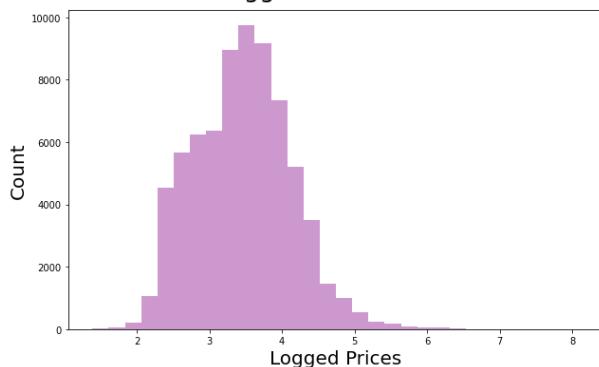
---

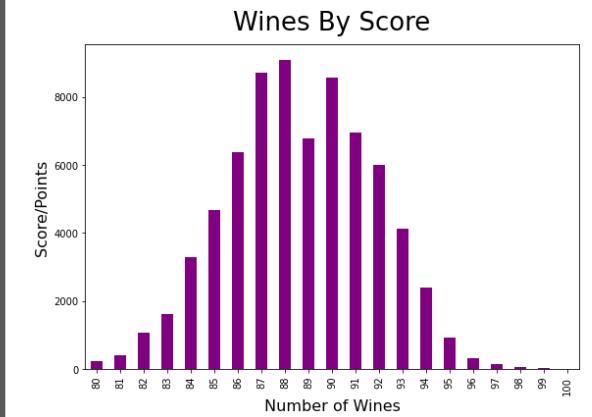
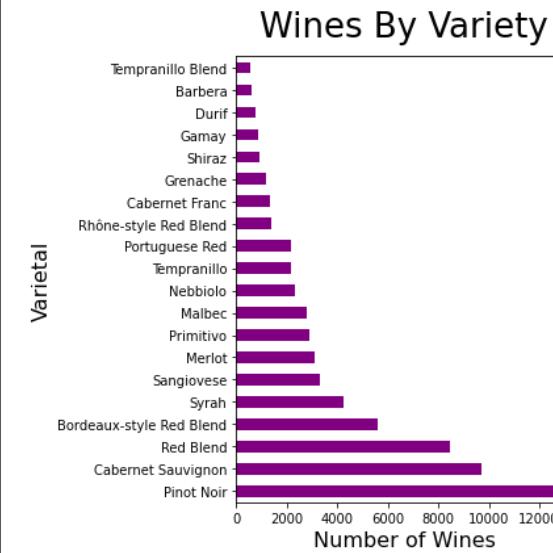
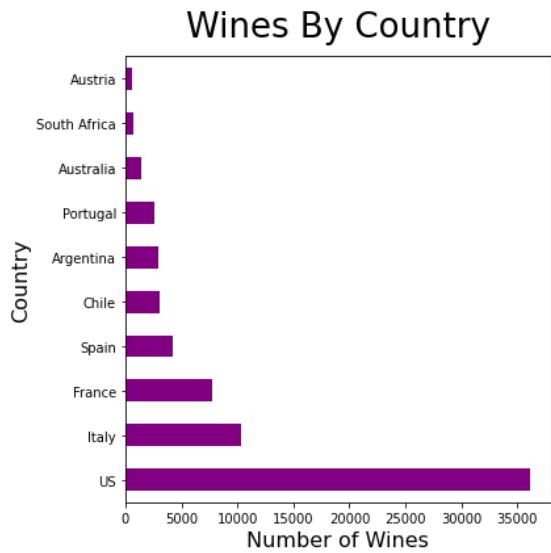
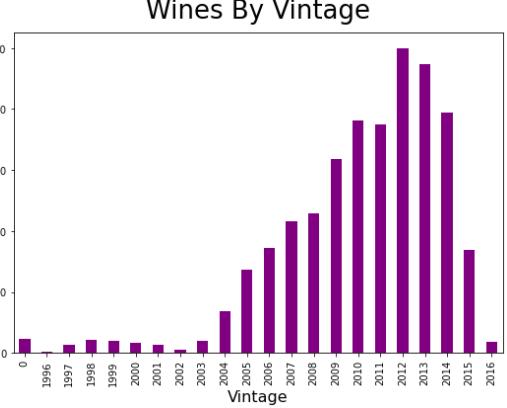
## The Target

Wine Prices



Logged Wine Prices

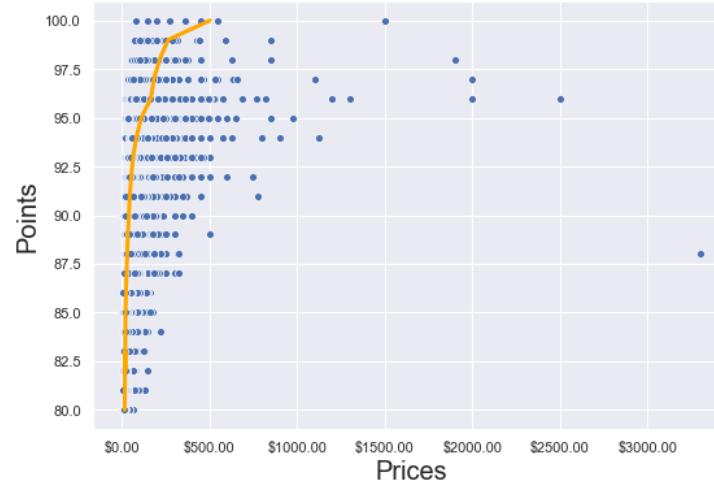




# EDA

## Data Distribution

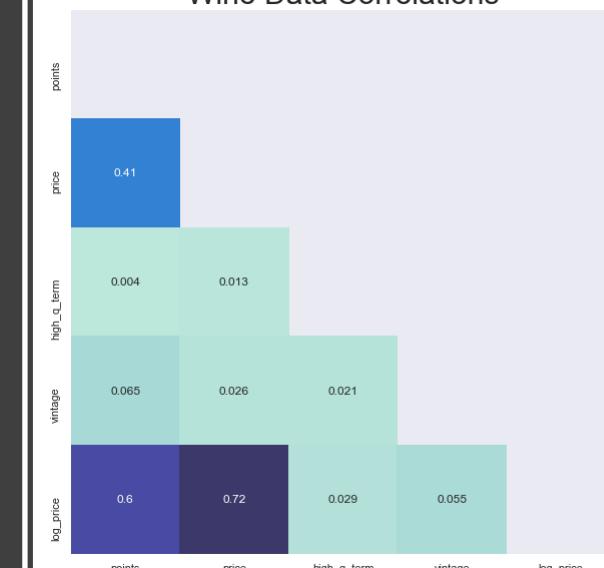
### Wine Prices vs Review Score



### Logged Prices vs Review Score



### Wine Data Correlations



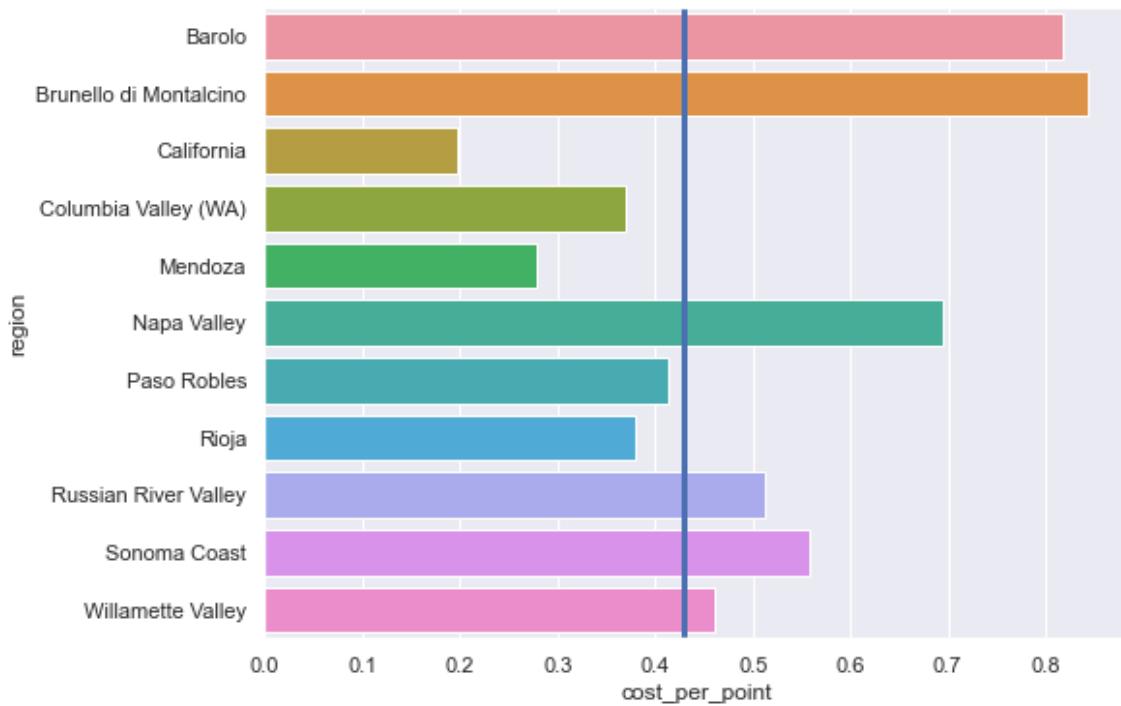
# EDA

## Price vs Review Score

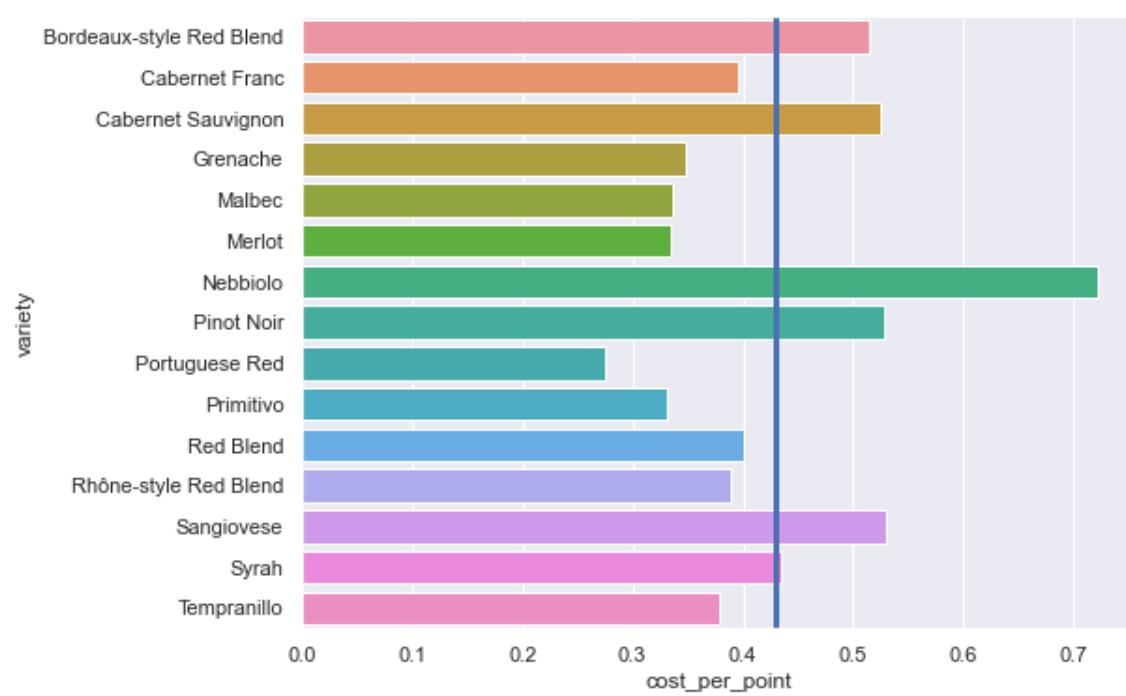
# EDA

## Cost per Point

Cost Per Point by Region



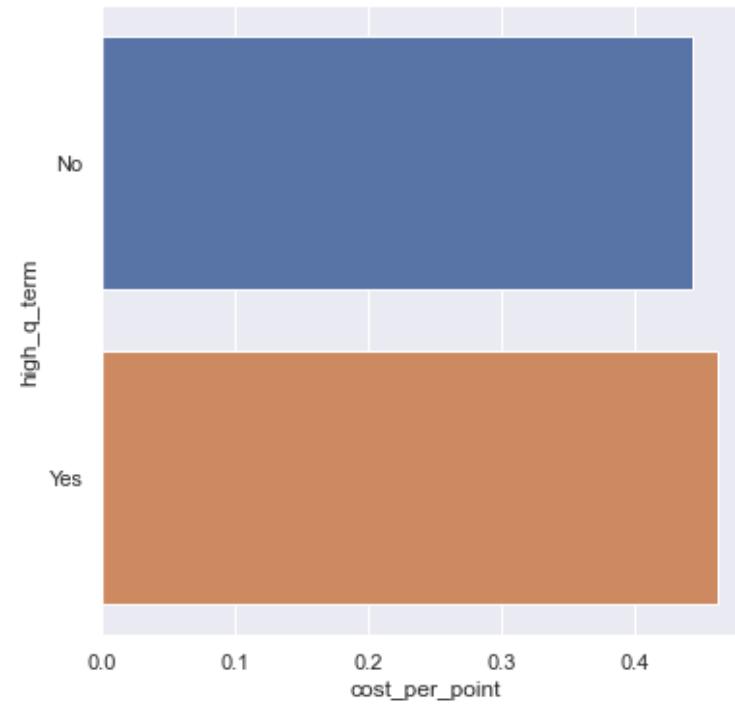
Cost Per Point by Variety



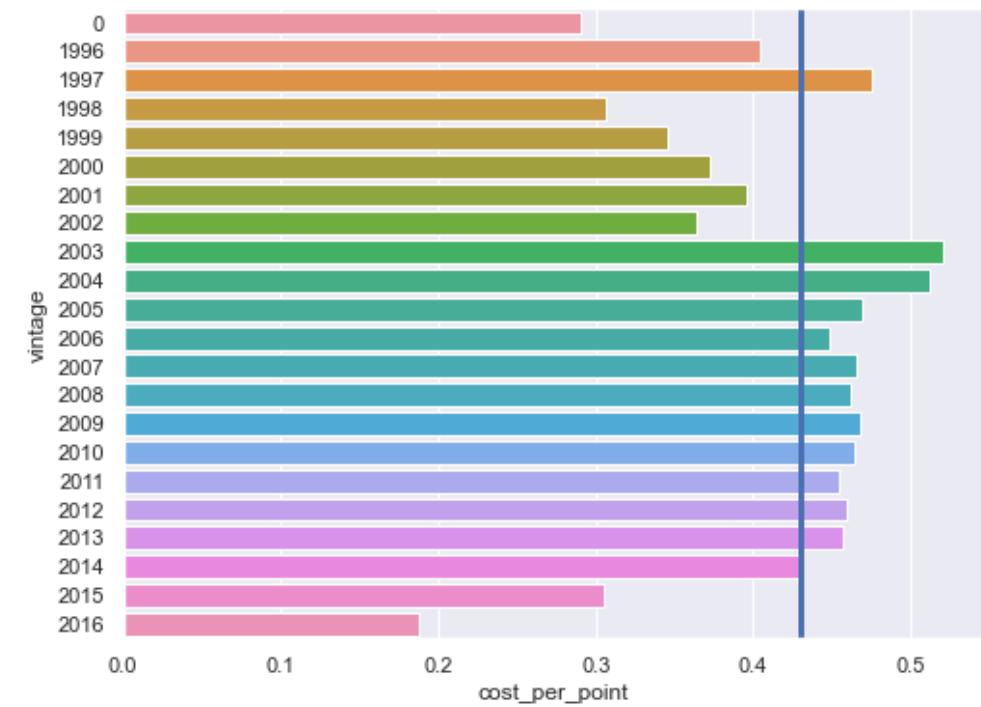
# EDA

## Cost per Point

Cost Per Point by Reserve Designation



Cost Per Point by Vintage



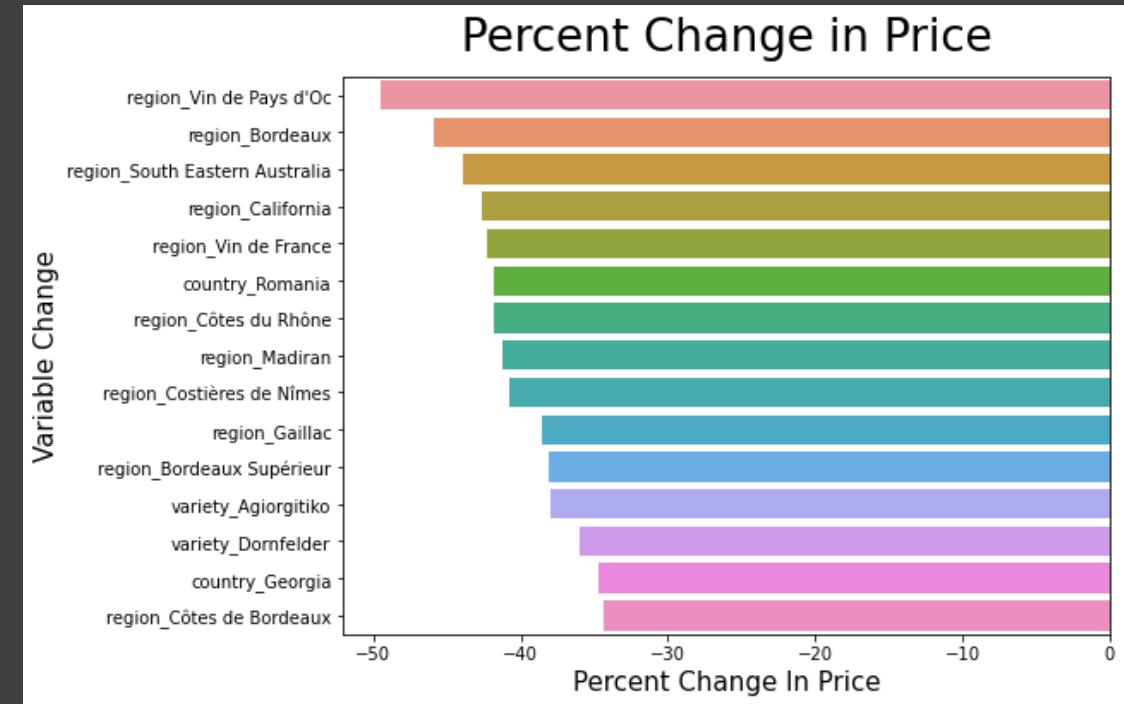
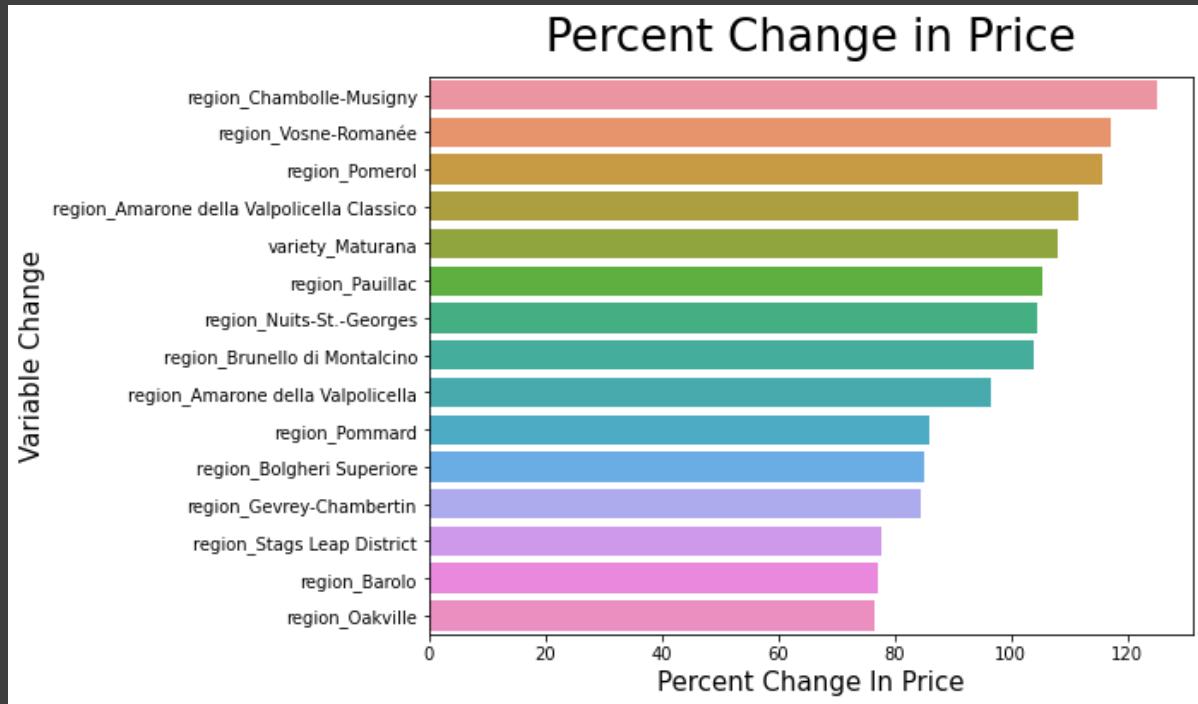


MODELING

- Regression Models
  - Linear Regression
  - Ridge
  - LASSO
  - Decision Trees
  - Random Forest
  - Extra Trees
  - Feed Forward Neural Network

# MODEL SELECTION

model	train_score	test_score
Dummy Regression	43.806031	48.292622
Linear Regression	37.039905	41.422648
Ridge Regression	37.050128	41.433638
LASSO Regression	37.055192	41.445604
Decision Tree Regression	34.318392	38.751863
Random Forest Regression	31.374335	37.837495
Extra Tree Regression	30.551202	37.520836
Feed Forward Neural Network	39.148201	39.130711



# MODEL EVALUATION



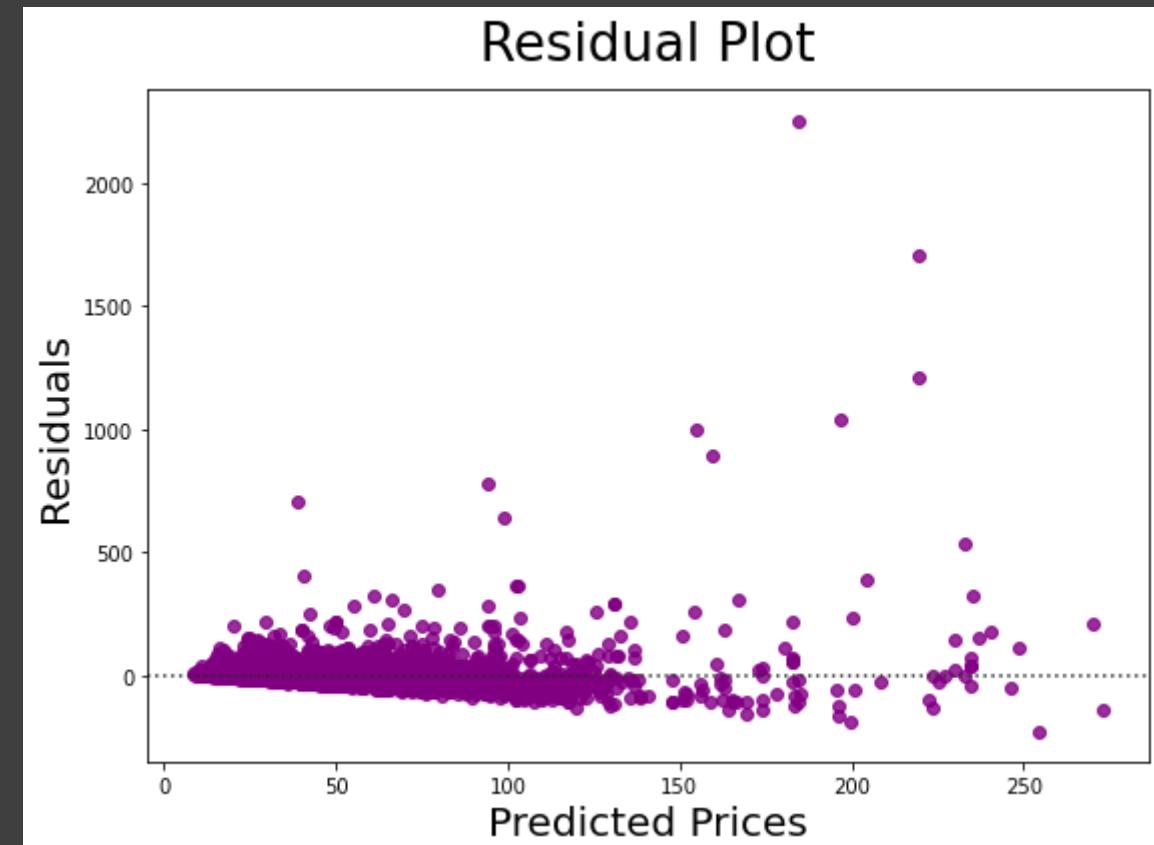
# CONCLUSIONS/ NEXT STEPS

## ➤ CONCLUSIONS:

- From our EDA and our model coefficient analysis, we have started to explore how a wine is priced and how we can make our budgets go further.
- Our predictive model has a nice start but, while it looks pretty good up to about \$150, things go a little off of the rails after that.

## ➤ FUTURE RESEARCH:

- There are some great resources out there that can help us to both expand and narrow our study. Expand with more reviews and narrow our scope of pricing to retailers that we can actually access,
- Include other kinds of wine if there is club interest.



# THANK YOU

➤ TEACHERS:

- Mahdi Shadkam-Farrokh
- Noah Christiansen
- Matt Brems

➤ FELLOW STUDENTS:

- Danielle Medellin
- Julia Kelman

➤ CREDITS

- Original Study – zackthoutt via Kaggle.com
- Wine Information
  - winemag.com
  - wine-searcher.com
  - thewinecellarinsider.com
  - wikipedia.org
  - University of Virginia
- Images
  - wikipedia.org
  - foodandwine.com
  - heart.org
  - lazenne.com
  - medicalnewstoday.com
  - socialvignerons.com

