



Wine Enthusiasts

Using data science to build a winery

Overview

Wine Enthusiast magazine has decided to start a winery.
They know they can make great wine.

We like wine and we know data.

We will investigate wine data to help Wine Enthusiast
make some decisions about the winery.



Meet the Team

Delia

Wordsmith. Favors big bold reds.

Juhi

Data guru. Craves cabernet.

Rebekah

Machine Learner. Does not discriminate.

Roseanne

Wine finder. Likes red blends.



Questions to answer

We will analyze wine data with machine learning to investigate:

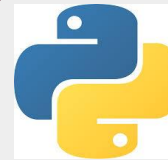
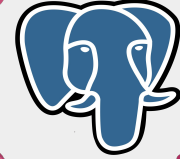
- Price
- Variety
- Category
- Location



Technology Utilized



kaggle



Google APIs



JS



Data Mining and Cleaning

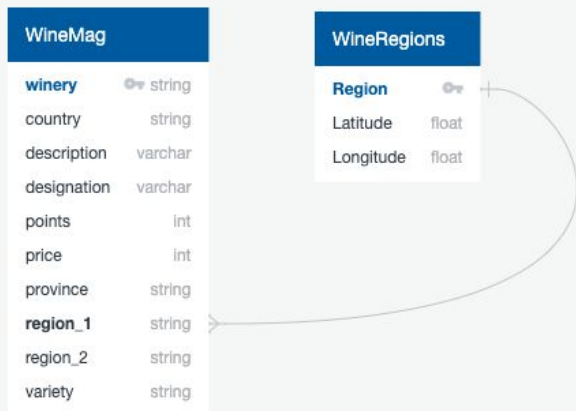
Data:

- Wine Enthusiast Magazine - Kaggle
- Google Places API - Latitude and Longitude

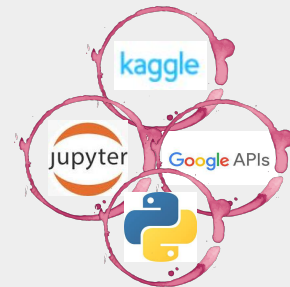
Google APIs

- Region → Place IDs
- Place IDs → Place Search

www.quickdatadiagrams.com



Database: postgres



Data Cleaning:

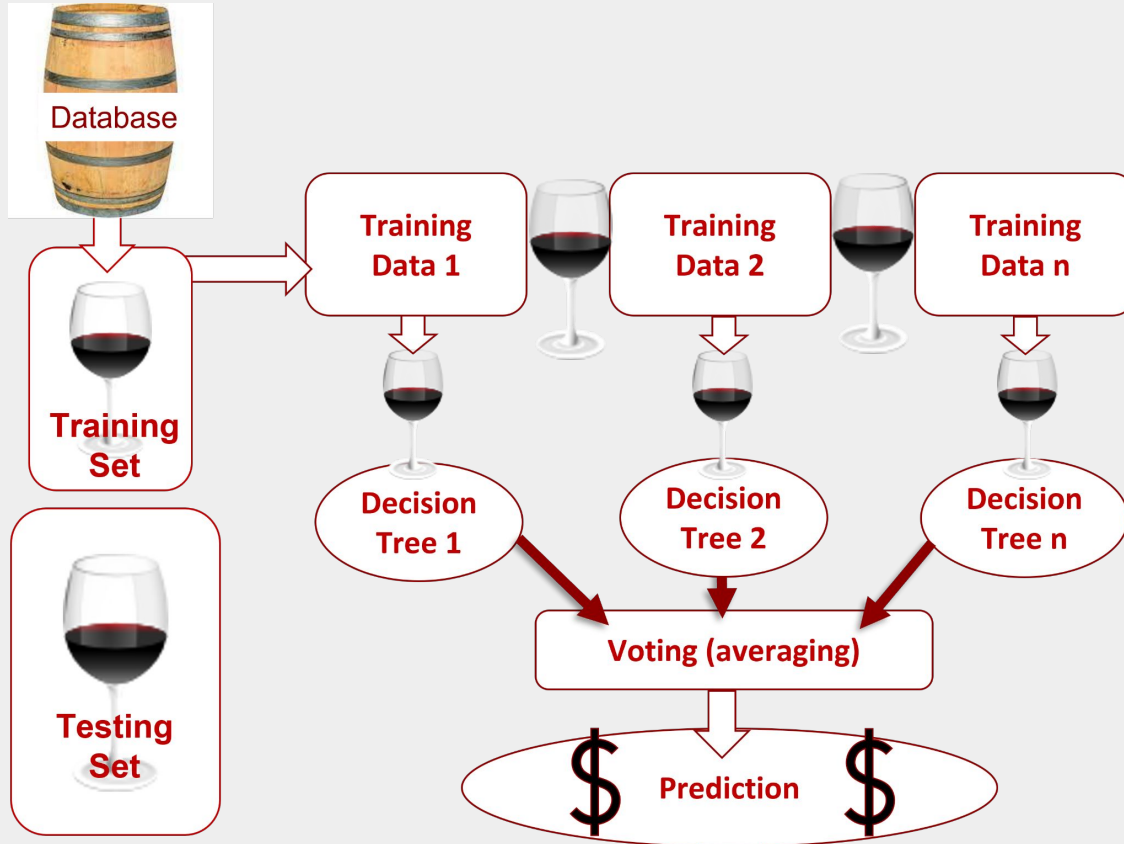
- Only US
- Only Washington, Oregon and California
- Binned Dataset
 - Region
 - Varieties

Global Dataset	
43	Countries
1229	Regions
707	Varieties



Project Dataset	
1	Country
40	Regions
23	Varieties

Random Forest Classifier - Analysis Phase



Random Forest Classifier (RFC) Advantages

- Robust against overfitting as the weak learners are trained on different pieces of the data.
- Runs efficiently on large datasets.

Defined Features

- Points
- Variety

Defined Target set by binning the prices

Data was split into training (75%) and testing sets (25%).

Random Forest Classifier - Results

Observations:

- Model accurately predicts moderately expensive wines (\$60 and under).
- Consistent and reproducible results with accuracy around 0.60
- Precision was the most important success measure but recall and other metrics were also good.

Recommendation to Wine Enthusiast Magazine:

- The moderately expensive category will suit a broader market for Wine Enthusiast to offer for their initial production run

Accuracy Score : 0.6154639935136729

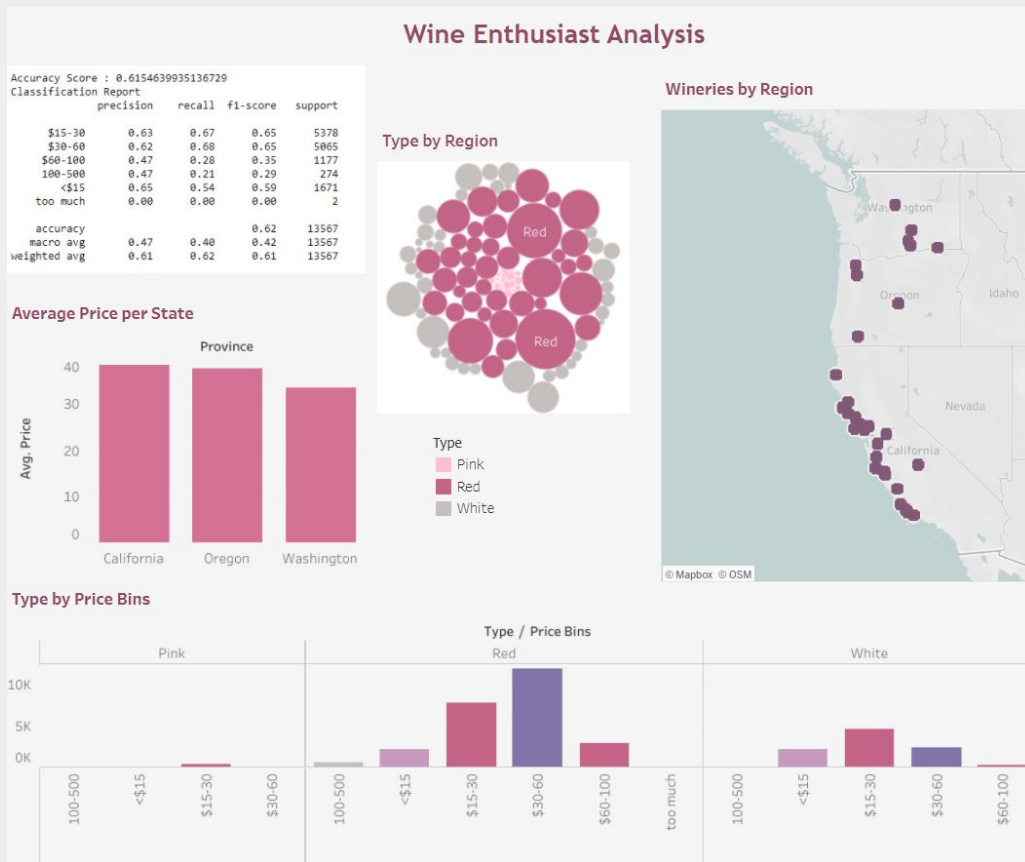
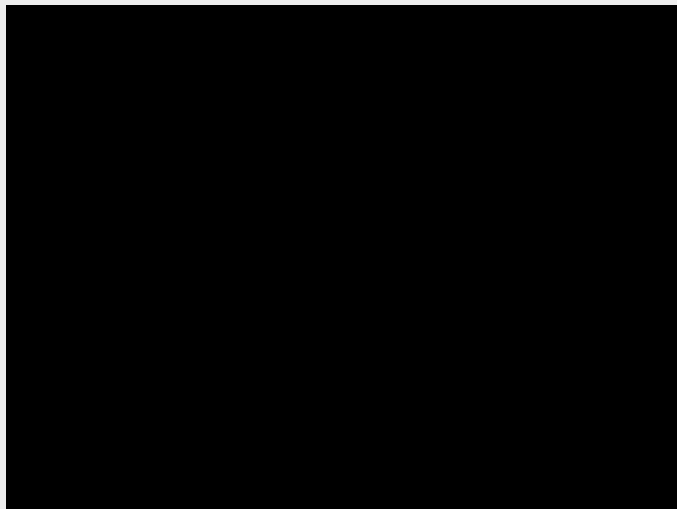
Classification Report

	precision	recall	f1-score	support
\$15-30	0.63	0.67	0.65	5378
\$30-60	0.62	0.68	0.65	5065
\$60-100	0.47	0.28	0.35	1177
100-500	0.47	0.21	0.29	274
<\$15	0.65	0.54	0.59	1671
too much	0.00	0.00	0.00	2
accuracy			0.62	13567
macro avg	0.47	0.40	0.42	13567
weighted avg	0.61	0.62	0.61	13567

Dashboard

We prepared a dashboard for the client so they can review data by:

- Average price by State
- Type by Region
- Wineries by Region
- Type by Price Bin



[Find it on](#)
[Tableau](#)

Find our [Wine Finder](#) online

Reset Filters

Wine Enthusiasts

"In wine, there's truth." — Pliny the Elder, Natural History

Filter Search

Enter Province

e.g., Oregon

Enter Region

e.g., Willamette Valley

Enter Winery

e.g., Sweet Cheeks

Enter Variety

Zinfandel

Enter Type

e.g., Red


Enter Points (min: 80 and max: 100)

90

Enter Price (Returns +/- \$5)

e.g., 18

What's your favorite flavor?



Wine Selection Table

Province	Region	Winery	Variety	Type	Title	Points	Price	Description
California	Alexander Valley	Matrix	Chardonnay	White	Matrix 2007 Stuhlmuller Vineyard Chardonnay	92	36	The vineyard is one of the better Chardonnay sources in Alexander Valley, and in this fine vintage it lends ripe pineapple, pear, Key lime pie and Mandarin orange flavors to the wine. Lots of toasty new oak adds buttered toast and vanilla, but the wood influence is in balance. Classic Burgundian-style California Chardonnay, rich and opulent.
California	Napa Valley	Silverado	Cabernet Sauvignon	Red	Silverado 2006 Cabernet Sauvignon	92	45	This sophisticated wine is mostly Cabernet Sauvignon, which accounts for the firm tannins and dense structure as well as most of the blackberries and blackcurrants. Smaller amounts of Merlot and Cabernet Franc add cherries and violets, while oak contributes smoky complexities. Rich and elegant, the wine provides lovely drinking now.
California	Napa Valley	Clark-Clauden	Cabernet Sauvignon	Red	Clark-Clauden 2007 Cabernet Sauvignon	91	78	Give this young Cab time in the cellar to come around. Right now, it's fresh and tannic, with primary fruit blackberry and cherry flavors, and the sweet smoky oak hasn't been integrated. Should knit together over the next 4-6 years, at least.
California	Santa Lucia Highlands	Banyan	Riesling	White	Banyan 2007 Riesling	90	17	A beautiful wine, sleek, crisp and ultra-dry, with the most refined flavors of petrol, white peach, kaffir lime, apricot skin and minerals. The acidity is very high and palate-cleansing. The wine was half fermented in barrels, half in stainless steel. Sommeliers, buy this by the case, if you can find it.

[Link to wine finder](#)

10

Project outcomes



- Developed a machine learning model to predict the price categories of interest
- Created a dashboard to visualize the data
 - The trends suggest:
 - Red is the most popular / expensive
 - Oregon has lower prices than California but less competition
- Provided a wine finder so the clients could get a detailed look at the data used

Recommendations for further work

In the future, we could:

- Expand number of countries
- Analyze by vintage
- Add competitor information
- Do in-depth analysis of specific varieties
- Create a smaller, leaner database
- Increase search functionality
- Drink more wine!



The End

Go pour yourself a glass of wine.