Wine Enthusiasts

Using data science to build a winery

Overview

Wine Enthusiast magazine has decided to start a winery.
They know wine and wineries and can create great wine.

They know wine. We like wine. We know data.

We will investigate wine data to help Wine Enthusiasts determine what varietals to start with, what they can charge, and even what location to consider for their winery.

Project objective

Use data analysis techniques to analyze wine data and provide visualizations to make the data easy to understand.

Questions to answer

The Wine Enthusiasts know every detail about wine.

We can help them with information about:

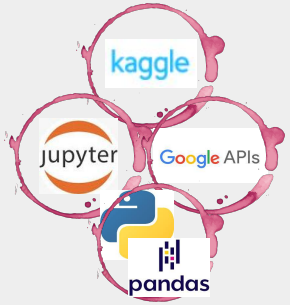
- Price
- Location
- Varieties and Category
- The Competition

Technology Utilized

Data Mining and Cleaning

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

Database: postgres



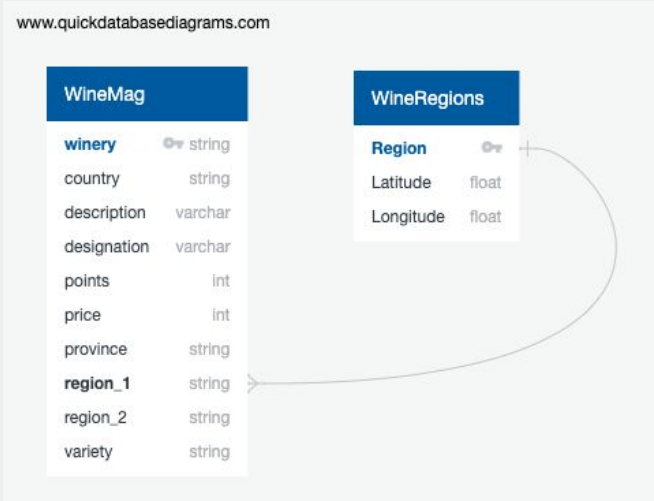
Data Cleaning:
- Only US
- Only Washington, Oregon and California
- Binned Dataset
  - Region
  - Varieties

Google APIs
- Region → Place IDs
- Place IDs → Place Search

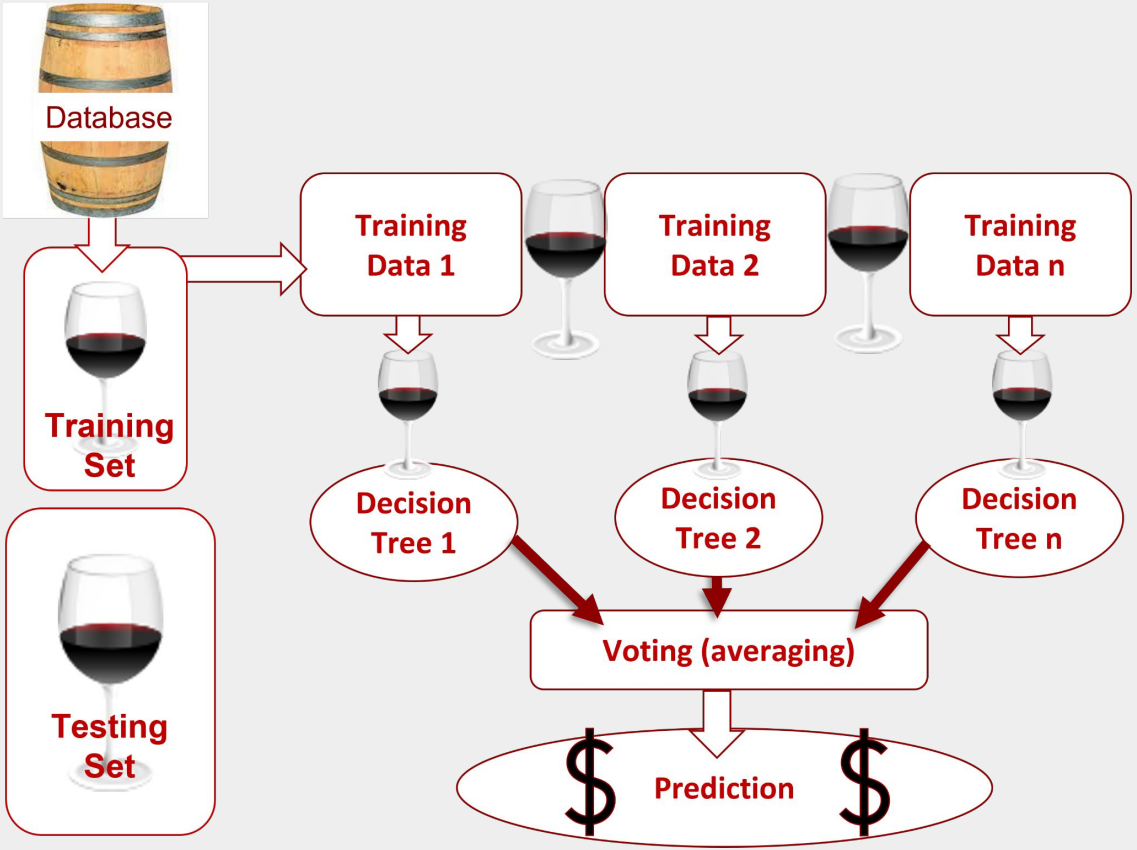
Global Dataset	
43	Countries
1229	Regions
707	Varieties

→

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 main goal but other measures of accuracy 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

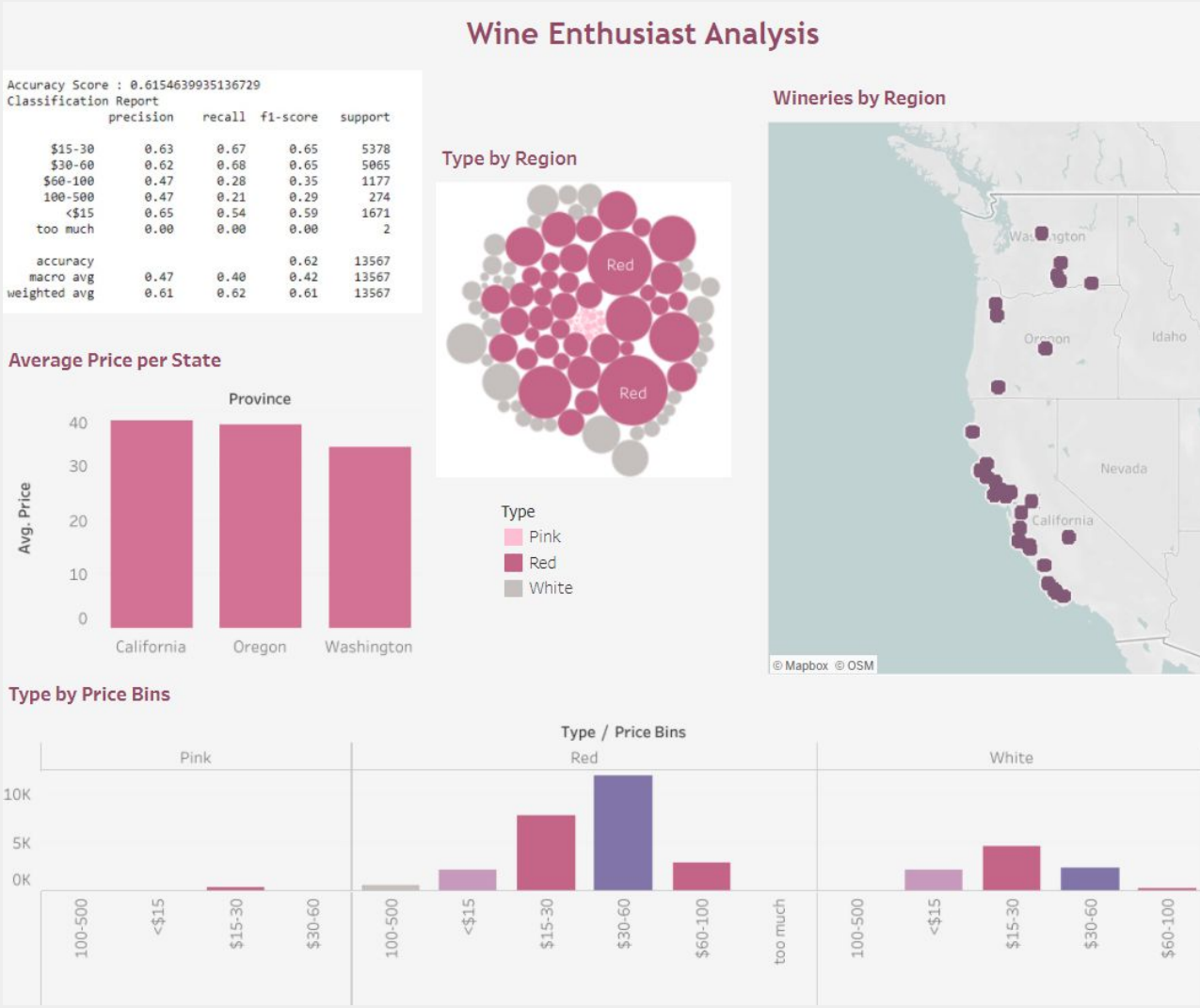
	precision	recall	f1-score	support
$15-30	0.63	0.67	0.65	5378
$30-60	0.62	0.68	0.65	5065
$60-100	0.47	0.28	0.35	1177
100-500	0.47	0.21	0.29	274
<$15	0.65	0.54	0.59	1671
too much	0.00	0.00	0.00	2
accuracy			0.62	13567
macro avg	0.47	0.40	0.42	13567
weighted avg	0.61	0.62	0.61	13567

Dashboard

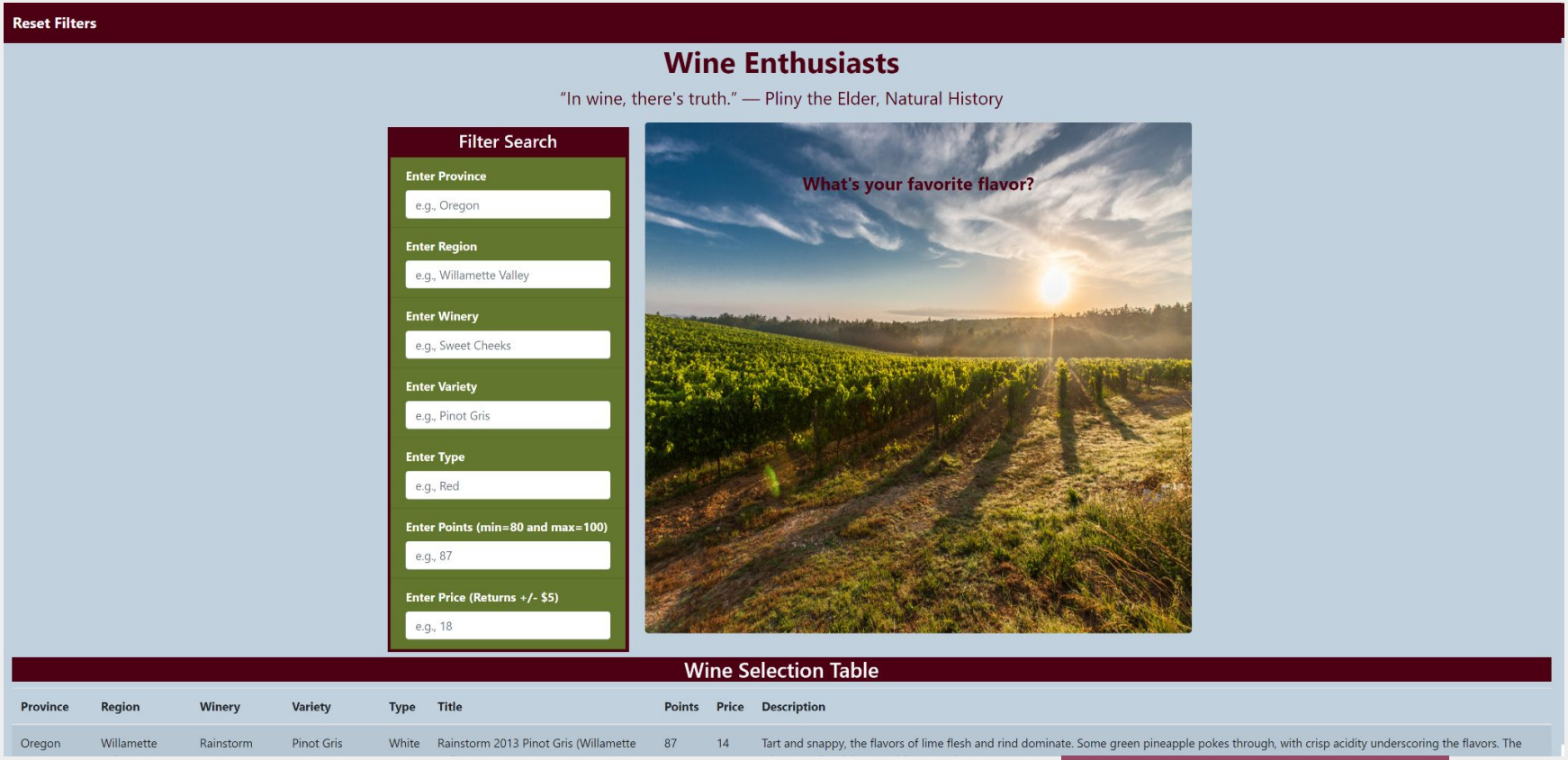
We prepared a dashboard for the Wine Enthusiasts so they can review data by:

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

[Find it on Tableau]



Find our Wine Finder online



Link to wine finder

Results Overview / Project outcomes

- Machine learning model was predictive in the price categories of interest to Wine Enthusiast (expensive but not extreme)
- We gave the clients a dashboard that allows them to easily break down the data
- The breakdown might suggest some options:
  - Start with red
  - California has higher prices but more competition
- We provided a wine picker so the Wine Enthusiast could get a detailed look at the data used

Future Recommendations

In the future, we suggest additional work on this project:

- Expand number of countries
- Analysis by vintage
- Competitor information added
- In-depth analysis on specific varieties
- Create a lean database to allow for a smaller database

Also, we might:

- Drink more wine
- Make the search descriptions by flavor
- Add partial search on varieties

The End

Go pour yourself a glass of wine.