



WINE QUALITY

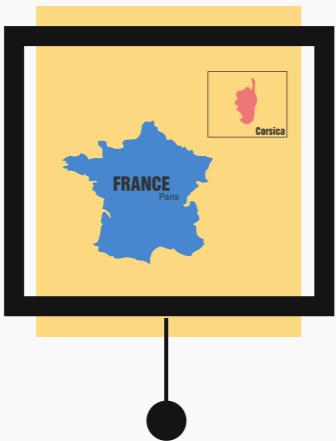
ANALYSIS & PREDICTIONS

PRESENTED BY:
THE SOMMELIERS

PROJECT OVERVIEW

01

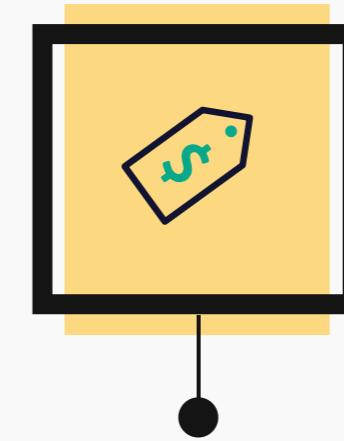
Utilize a machine learning classification model to predict wine quality category based on various features including...



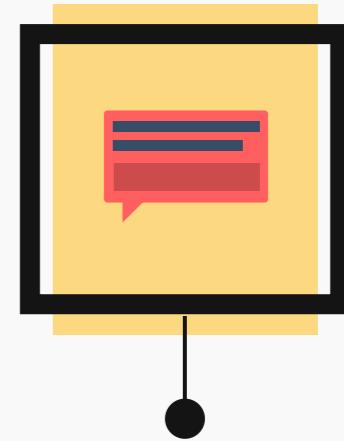
COUNTRY



VARIETY



PRICE



DESCRIPTIVE WORDS

02

Develop interactive dashboard to inform business owners in developing a comprehensive wine inventory.



EXPLORATORY

DATA ANALYSIS



01

Original dataset included 130,000 entries

02

Data cleaning included: [A] Clean-up specialty characters, [B] Review for null values and removal of columns with many null values (ie: "region_1"), [C] Remove unnecessary columns (ie: "taster_twitter_handle")

03

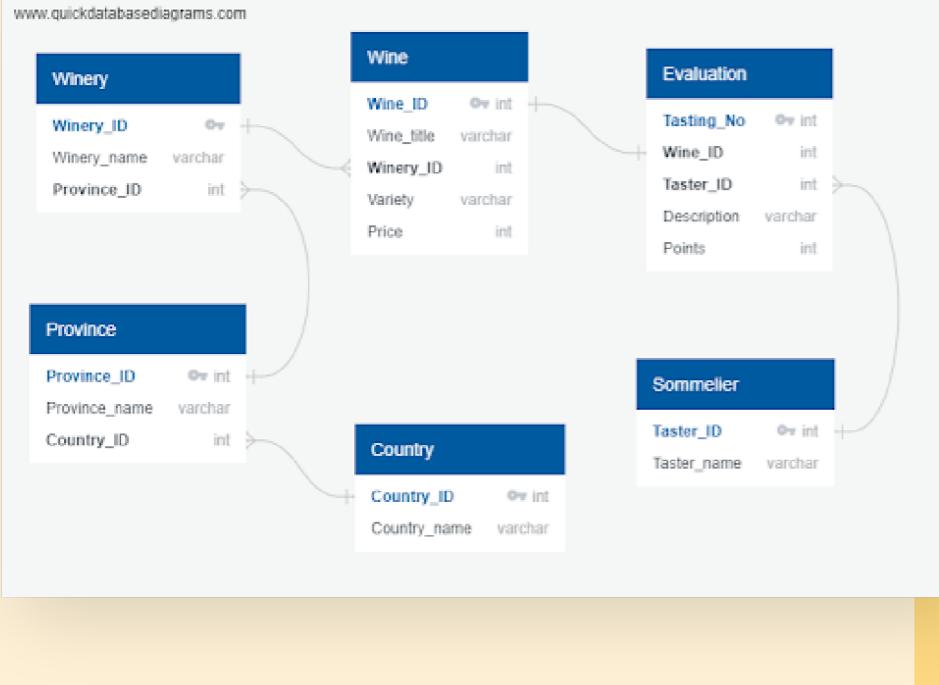
Cleaned dataset includes: [A] 96,240 entries, [B] 10 columns

04

NLP Parsing Description Field: Identified most common descriptors and assigned as additional features

01

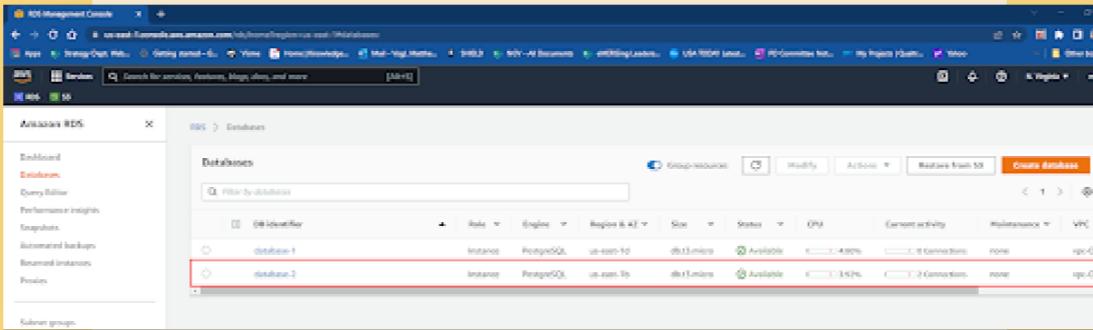
We began by visually mapping out our database tables connections in an Entity Relationship Diagram (ERD).



DATABASE

03

Finally, we developed our database schema in pgAdmin and joined all of the tables into a master .CSV file.



Next, we created our RDS database in AWS to host our database in the cloud...then ended up deleting it after it was hacked.

02

A screenshot of pgAdmin 4. The left sidebar shows 'Servers' with 'PostgreSQL 11' selected. Under 'Databases', there are several tables listed. A red box highlights the 'ml_master_table' table in the 'Tables' section. The 'Query Editor' tab shows a script of SQL code used to create the tables.

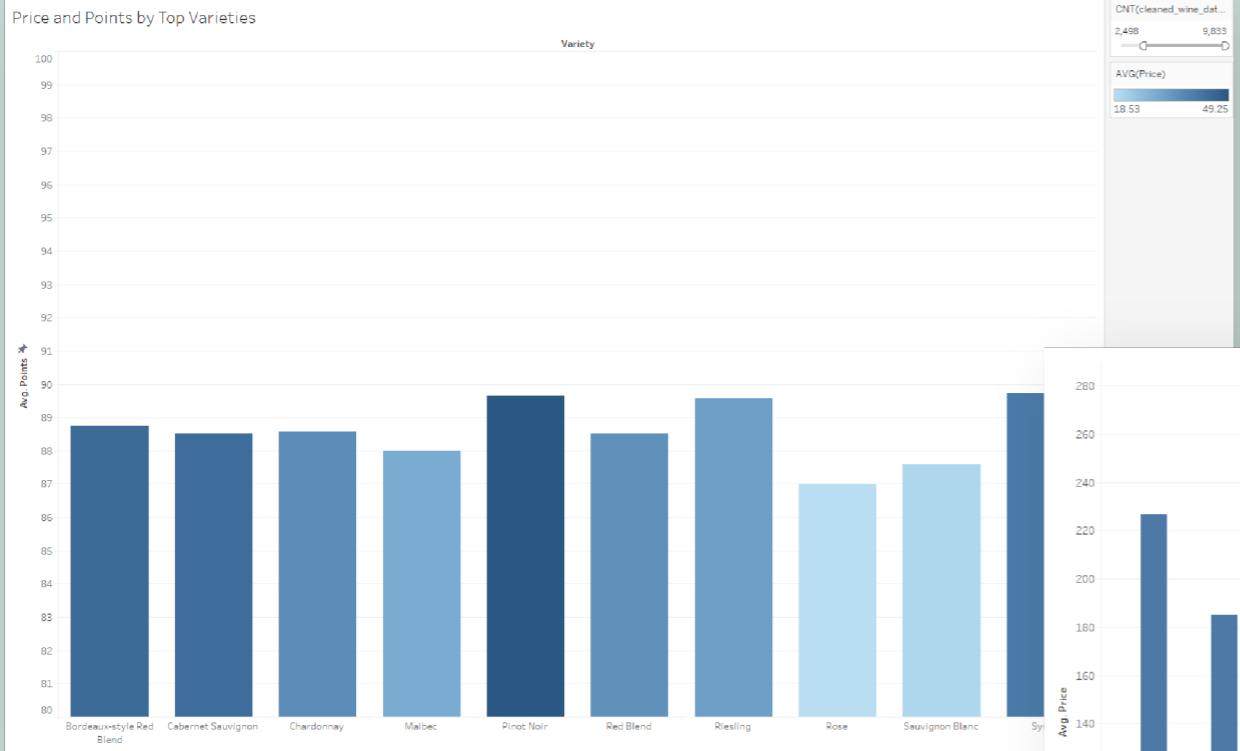
```
33 -- Creating Province table
34 CREATE TABLE province (
35   Province_ID INT,
36   Province_name VARCHAR,
37   PRIMARY KEY (Province_ID)
38 );
39 -- Creating Country table
40 CREATE TABLE country (
41   Country_ID INT,
42   Country_name VARCHAR,
43   PRIMARY KEY (Country_ID)
44 );
45 -- Creating Sommeller table
46 CREATE TABLE Sommeller (
47   Taster_ID INT,
48   Taster_name VARCHAR,
49   PRIMARY KEY (Taster_ID)
50 );
51 -- Joining the tables
52 SELECT e.description,
53   e.points,
54   w.wine_ID,
55   w.variety,
56   w.price,
57   pr.province_ID,
58   co.country_ID,
59   co.country_name,
60   w.winery_ID
61   INTO ml_master_table
62   FROM Evaluation AS e
63   JOIN Wine AS w
64   ON e.wine_ID = w.wine_ID
65   JOIN Winery AS wy
66   ON w.winery_ID = wy.winery_ID;
```

The 'Tables' section also includes 'country', 'evaluation', 'ml_master_table', 'province', 'sommeller', 'wine', 'winery', and 'wines'.

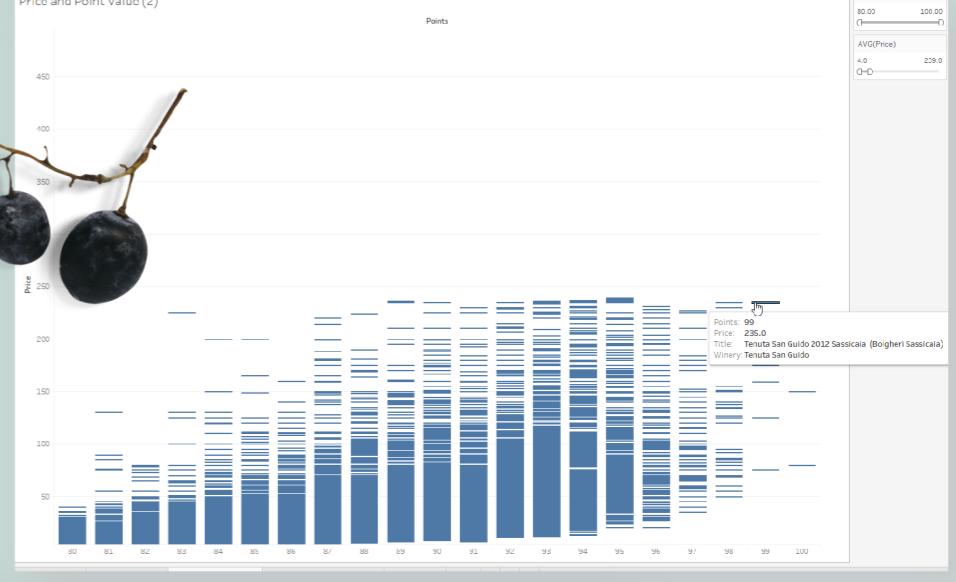
DASHBOARD

DEVELOPMENT

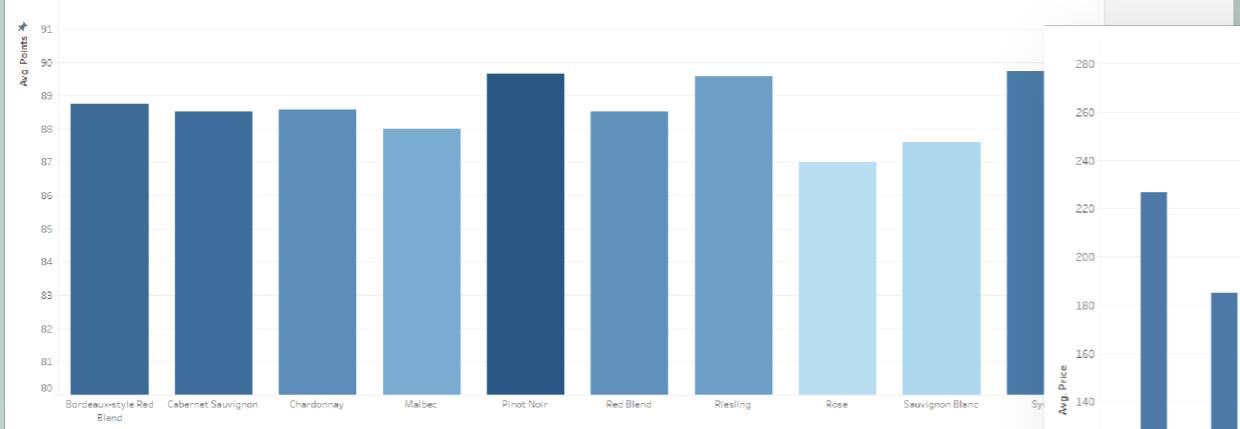
Price and Points by Top Varieties



Price and Point Value (2)



Avg Points ↗



Winery

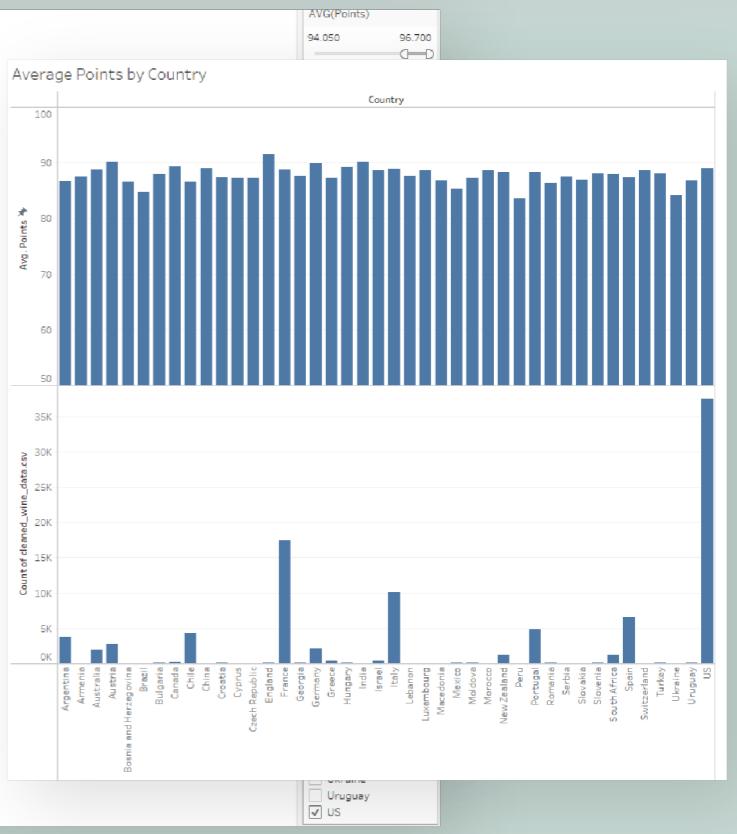


Avg Price ↗



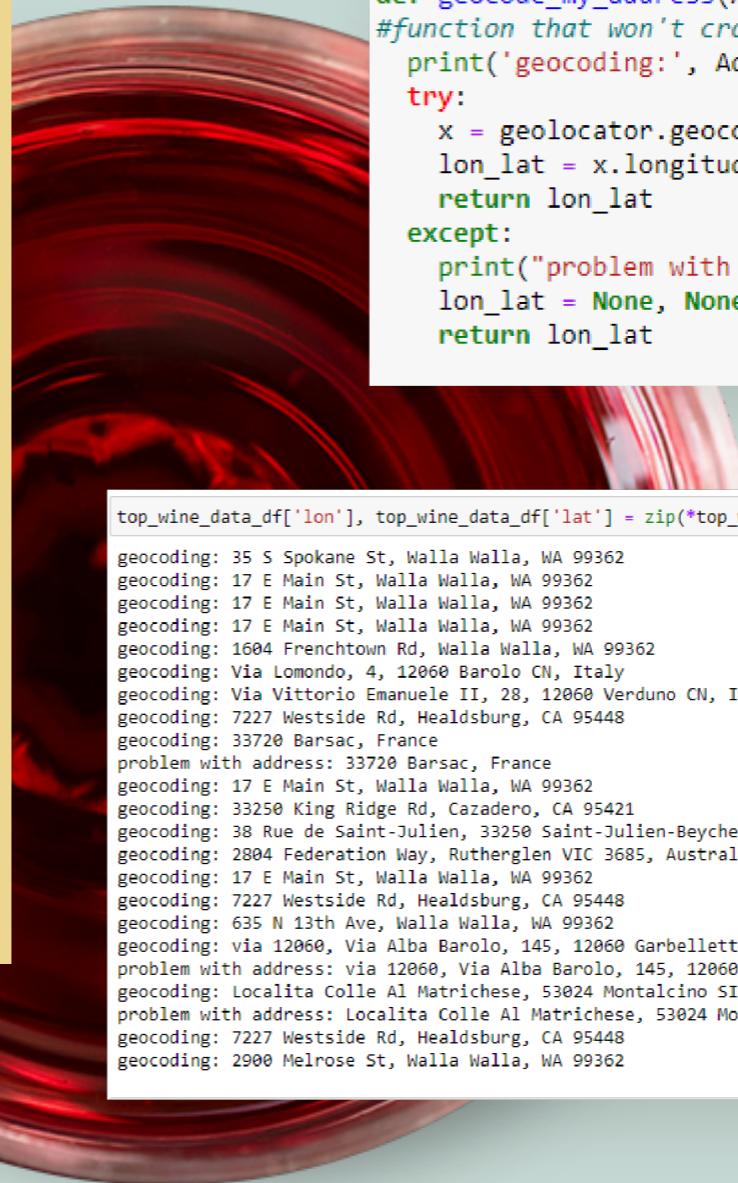
Winery

Average Points by Country



MAPPING

- Manual collection of addresses for top 20 reasonably priced (<\$100) wines
- Added applicable columns (address, lat, long)
- Lat/Long API call
- Export dataframe as csv and json



```
def geocode_my_address(Address):
    #function that won't crash if it runs into a bad address
    print('geocoding:', Address)
    try:
        x = geolocator.geocode(Address) #[‘geometry’].squeeze()
        lon_lat = x.longitude, x.latitude
        return lon_lat
    except:
        print("problem with address:", Address)
        lon_lat = None, None
        return lon_lat
```

```
top_wine_data_df['lon'], top_wine_data_df['lat'] = zip(*top_wine_data_df['Address'].apply(lambda x: geocode_my_address(x)))

geocoding: 35 S Spokane St, Walla Walla, WA 99362
geocoding: 17 E Main St, Walla Walla, WA 99362
geocoding: 17 E Main St, Walla Walla, WA 99362
geocoding: 17 E Main St, Walla Walla, WA 99362
geocoding: 1604 Frenchtown Rd, Walla Walla, WA 99362
geocoding: Via Lomondo, 4, 12060 Barolo CN, Italy
geocoding: Via Vittorio Emanuele II, 28, 12060 Verduno CN, Italy
geocoding: 7227 Westside Rd, Healdsburg, CA 95448
geocoding: 33720 Barsac, France
problem with address: 33720 Barsac, France
geocoding: 17 E Main St, Walla Walla, WA 99362
geocoding: 33250 King Ridge Rd, Cazadero, CA 95421
geocoding: 38 Rue de Saint-Julien, 33250 Saint-Julien-Beychevelle, France
geocoding: 2804 Federation Way, Rutherford VIC 3685, Australia
geocoding: 17 E Main St, Walla Walla, WA 99362
geocoding: 7227 Westside Rd, Healdsburg, CA 95448
geocoding: 635 N 13th Ave, Walla Walla, WA 99362
geocoding: via 12060, Via Alba Barolo, 145, 12060 Garbelletto CN, Italy
problem with address: via 12060, Via Alba Barolo, 145, 12060 Garbelletto CN, Italy
geocoding: Localita Colle Al Matrichese, 53024 Montalcino SI, Italy
problem with address: Localita Colle Al Matrichese, 53024 Montalcino SI, Italy
geocoding: 7227 Westside Rd, Healdsburg, CA 95448
geocoding: 2900 Melrose St, Walla Walla, WA 99362
```



MAPPING

- Utilize Leaflet & Tableau functionality to explore map options
- Leaflet:
 - circle markers for top 20 wineries
 - multiple layer groups & overlays
- Tableau:
 - pop-up markers with AVG price and points
 - filtered by country



Machine Learning Model Development

1

Build multiple ML models to test

- LASSO regression Model
- Neural Network Optimization Model
- Decision Tree Classifier
- Logistic Regression Model
- Random Forest Classifier Model

2

Utilize variety of features to predict target value of points classification, i.e.:

- Price
- Variety
- Country
- Descriptive Words

A wide-angle photograph of a wine cellar. Numerous wooden barrels are stacked in rows on both sides, filling the space. The barrels have various labels, including "Pahlmeyer 2016" and "Pahlmeyer 2010". The ceiling is made of concrete, and the floor is a polished concrete surface. A white rectangular box containing the text is overlaid on the center of the image.

ON TO OUR
INTERACTIVE
WEBSITE