Winos

The Team

How is Climate Change affecting wine?

Kyle Johnson

Square

First Segment: Square
Second Segment: Circle
Third Segment: Triangle
Fourth Segment: X

Marisa Shideler

Triangle

First Segment: Triangle
Second Segment: Triangle
Third Segment: X
Fourth Segment: Triangle

Brenya Skaggs

Circle

First Segment: Circle
Second Segment: X
Third Segment: Circle
Fourth Segment: Circle

Zackary Gheen

X

First Segment: X
Second Segment: Square
Third Segment: Square
Fourth Segment: Square

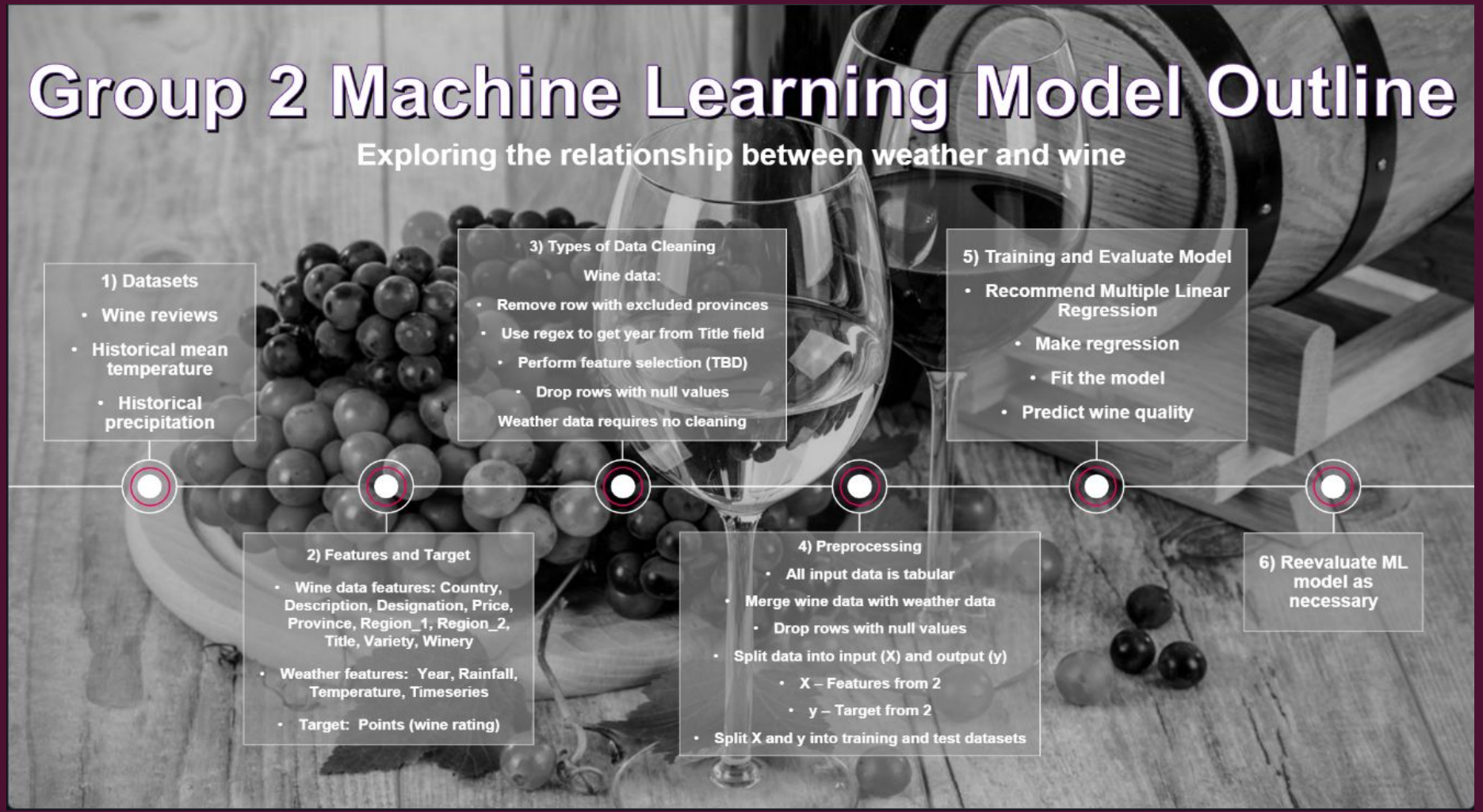
The topic of this project is predicting the quality of wine based on environmental factors such as geography, temperature and rainfall. Climate change is expected to have a dramatic impact on these variables in the future and as enthusiastic wine drinkers, we hope to learn if this will have an impact on our preferred wine growing regions.

Questions to Answer

- Do higher temperatures/rainfall correlate with higher or lower quality wine?
- What effect will future changes in rainfall and temperatures have on wine quality from various regions?
- Are new regions poised to emerge as premiere locations for growing grapes and producing wine?

Group 2 Machine Learning Model Outline

Exploring the relationship between weather and wine



1) Datasets

- Wine reviews
- Historical mean temperature
- Historical precipitation

2) Features and Target

- Wine data features: Country, Description, Designation, Price, Province, Region_1, Region_2, Title, Variety, Winery
- Weather features: Year, Rainfall, Temperature, Timeseries
- Target: Points (wine rating)

3) Types of Data Cleaning

Wine data:

- Remove row with excluded provinces
- Use regex to get year from Title field
- Perform feature selection (TBD)
- Drop rows with null values

Weather data requires no cleaning

4) Preprocessing

- All input data is tabular
- Merge wine data with weather data
- Drop rows with null values
- Split data into input (X) and output (y)
  - X – Features from 2
  - y – Target from 2
- Split X and y into training and test datasets

5) Training and Evaluate Model

- Recommend Multiple Linear Regression
- Make regression
- Fit the model
- Predict wine quality

6) Reevaluate ML model as necessary

Technologies, Languages & Tools

AWS
- -RDS
- -S3

PostgreSQL / PGAdmin

Python version 3.8.8
- -Pandas
- -Sklearn
- -Tensor Flow

Flask

HTML

CSS

Javascript

Jupyter Notebook

Google Colaboratory

Google Slides

VSCode

Data Sources

Environment Dataset

- The world bank provides observed rainfall and temperature data by year for regions within individual countries from 1901-present.
- Future predictions of the weather with the same structure are provided from 2020-2100.
- Home | Climate Change Knowledge Portal (worldbank.org)

Wine Dataset

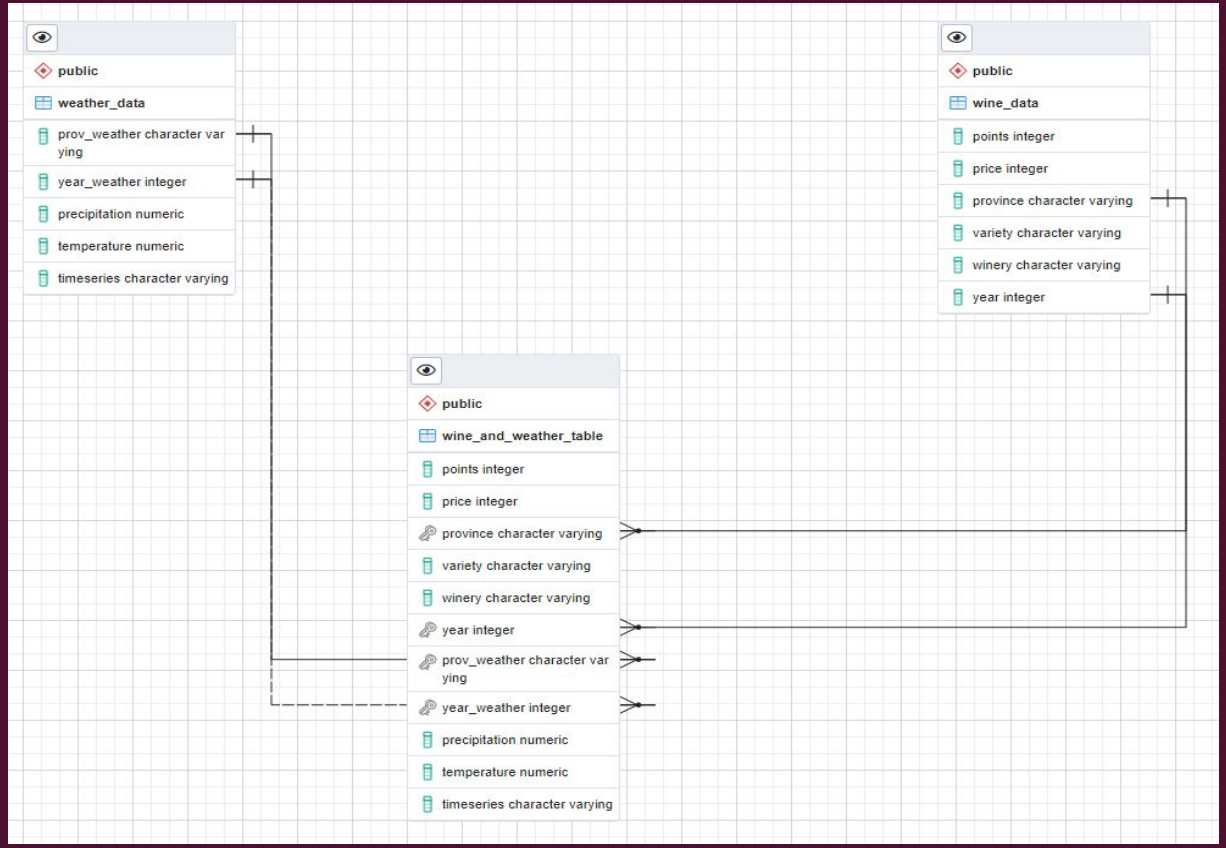
- This dataset includes 130,000 records of wine reviews from 2000-2017.
- Wine Reviews | Kaggle

Regions to be examined

- California, US
- Washington, US
- Aquitaine, France
- Tuscany, Italy
- Oregon, US
- Cantabria, Spain
- Burgundy,France
- Piemonte,Italy
- Veneto, ITaly
- New York, US
- Alsace, France
- Sicily, Italy
- Champagne Ardenne, France
- Sardinia, Italy

Dataset Analysis

Weather & Wine Data



Weather Data

Raw Data

Historical and projected precipitation and temperature were extracted from the country data for the following provinces/states:

- Alsace, France
- Aquitaine, France
- Burgundy, France
- Champagne Ardenne, France
- Piemonte, Italy
- Sardinia, Italy
- Sicily, Italy
- Tuscany, Italy
- Veneto, Italy
- Cantabria, Spain
- California, US
- New York, US
- Oregon, US
- Washington, US

	Unnamed: 0	Spain	Andalucía	Aragón	Principado de Asturias	Illes Balears	Canarias	Cantabria	Castilla y León	Castilla-La Mancha	...
0	1901	660.97	731.37	629.56	813.21	636.41	623.12	953.56	584.90	557.29	...
1	1902	660.49	653.91	633.91	837.45	642.47	292.17	960.72	636.91	552.15	...
2	1903	556.15	452.08	481.78	903.15	627.99	539.16	909.01	588.50	410.55	...
3	1904	582.28	614.13	485.96	767.75	562.99	534.67	834.29	549.63	526.19	...
4	1905	530.70	471.13	522.26	761.79	519.78	390.46	798.84	510.66	417.18	...

5 rows × 21 columns

Wine Data
Raw Data

The raw wine data, which contain over 80,000 wine reviews, came from Kaggle. This dataset contains varieties of wine along with other information such where and when they came from, what wineries were used, how much they cost and what the review rating is.

	Unnamed: 0	country	description	designation	points	price	province	region_1	region_2	title	variety	winery
0	0	Italy	Aromas include tropical fruit, broom, brimston...	Vulkà Bianco	87	NaN	Sicily & Sardinia	Etna	NaN	Nicosia 2013 Vulkà Bianco (Etna)	White Blend	Nicosia
1	1	Portugal	This is ripe and fruity, a wine that is smooth...	Avidagos	87	15.0	Douro	NaN	NaN	Quinta dos Avidagos 2011 Avidagos Red (Douro)	Portuguese Red	Quinta dos Avidagos
2	2	US	Tart and snappy, the flavors of lime flesh and...	NaN	87	14.0	Oregon	Willamette Valley	Willamette Valley	Rainstorm 2013 Pinot Gris (Willamette Valley)	Pinot Gris	Rainstorm
3	3	US	Pineapple rind, lemon pith and orange blossom ...	Reserve Late Harvest	87	13.0	Michigan	Lake Michigan Shore	NaN	St. Julian 2013 Reserve Late Harvest Riesling ...	Riesling	St. Julian
4	4	US	Much like the regular bottling from 2012, this...	Vintner's Reserve Wild Child Block	87	65.0	Oregon	Willamette Valley	Willamette Valley	Sweet Cheeks 2012 Vintner's Reserve Wild Child...	Pinot Noir	Sweet Cheeks

Combined Weather and Wine Data

Google Colaboratory was used to write the weather and wine data to an Amazon RDS instance. Using SQL we were able to create a new table with the joined data. The following table ill be used for our Machine Learning Model.

```
CREATE TABLE wine_weather_table
AS (SELECT * FROM wine_data
        LEFT JOIN weather_data ON wine_data.province=weather_data.Prov_Weather AND wine_data.year=weather_data.Year_Weather
        WHERE weather_data.Timeseries = 'Historical');
```

	points	price	province	variety	winery	year	prov_weather	year_weather	precipitation	temperature	timeseries
0	87	14	Oregon	Pinot Gris	Rainstorm	2013	Oregon	2013	440.10	8.68	Historical
1	87	65	Oregon	Pinot Noir	Sweet Cheeks	2012	Oregon	2012	808.24	9.04	Historical
2	87	24	Alsace	GewÃ¼rztraminer	Trimbach	2012	Alsace	2012	938.63	10.39	Historical
3	87	27	Alsace	Pinot Gris	Jean-Baptiste Adam	2012	Alsace	2012	938.63	10.39	Historical
4	87	19	California	Cabernet Sauvignon	Kirkland Signature	2011	California	2011	489.67	13.43	Historical
...	...	...	...	...	...	...	...	...	...	...	...
58846	90	28	Alsace	Pinot Gris	Domaine RieflÃ©-Landmann	2013	Alsace	2013	949.40	9.94	Historical
58847	90	75	Oregon	Pinot Noir	Citation	2004	Oregon	2004	596.22	9.28	Historical
58848	90	30	Alsace	GewÃ¼rztraminer	Domaine Gresser	2013	Alsace	2013	949.40	9.94	Historical
58849	90	32	Alsace	Pinot Gris	Domaine Marcel Deiss	2012	Alsace	2012	938.63	10.39	Historical
58850	90	21	Alsace	GewÃ¼rztraminer	Domaine Schoffit	2012	Alsace	2012	938.63	10.39	Historical

58851 rows × 11 columns

Machine Learning

Challenges

Results

Predictions

Dashboard

Recommendations