Group 2 Machine Learning 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