Project 1 Outline

Team: Don’t Touch My Script

Members: Tom Manh, Nabil Qureshi, Ashvin Rajput & Jordan Mullin

WHO HAS THE BEST WINE?

**Outline**

The kind of data we will be working with will pertain to the production of wine and any subset of information that we find or expect to contribute to our findings. The data we expect to analyze includes but is not limited to:

**Data**

* For any wine:
  + Price of Wine
  + Preference scores (and the taster who tasted and rated it)
  + Location of where the wine was produced (country and region)
  + The kind of wine (which indicates the grapes used to produce it)
  + The winery it is from
* Growing conditions of the various regions
* Ideal temperatures to grow wine (expecting multiple for various kinds of grapes and wine)
* Output (export) metrics to determine the size and scope of the industry
* Production metrics will also be included (this is in contrast to export because some countries or areas consume a vast majority of the product they produce and therefore have less than expected exports)
* How popular wine is in various areas of the world, whether it be country or region.

The kinds of questions we will be asking will be around the central theme of determining which countries/regions/companies (or vineyards) around the world are best at various aspects of wine. We will seek to understand:

* which country produces the most wine?
* which country has wine that is most popular among consumers?
* Which country has the best climate or conditions for growing wine?
* Which country or region has the largest growing market for wine?
* What kind of wine is the most expensive?
* What kind of wine has the best value (quality to price ratio)?
* Is more expensive wine always better? (price / value ratio)
* Which country consumes the most of their countries production (as a percent)?
* What is a reasonable amount to spend on good quality wine?
* What is the range of prices for wine of any particular kind?

The above is simply a preface to the kind of questions we aim to ask and we expect these questions to change or grow as we see more data and have a better idea of what is best to analyze.

**Data Sources**

The sources of data will vary. We intend to use previously compiled datasets from other sources such as Kaggle.com or other similar websites. In addition, we will need to gather weather related information to determine what the growing conditions are in certain areas of the world, likely by utilizing an API as we have covered in this class. The sources of data will be vetted by the team to ensure that there isn’t inherent flaws within the dataset itself, but for the purpose of this project we will have to assume that the data represented in any third-party sources is correct and accurate.