Chase Johnson CTJ527

Term Project Proposal

The goal of my project is to generate wine recommendations from analyzing wine ratings I have been collecting from me and my friends over the past year.

There are two datasets I will be using for my project. The first is a spreadsheet containing information on the wines we have rated. This includes the name, type, region, and alcohol content of each wine, as well as the date we reviewed it, who selected the wine, and the individual ratings from each person. There is also a column for the rating of the wine found on the wine rating website vivino.com. I will be referring to this dataset as the ratings dataset. The second dataset is a large collection of wine data that contains as many of the same characteristics recorded in the ratings dataset. I will be referring to this dataset as the wine dataset.

To generate wine recommendations, the first step is to analyze the ratings dataset. I will generate regressions between each person’s ratings and the wine characteristics. This will allow me to have a prediction of someone’s wine rating given its characteristics. The second step is to predict each person’s rating for each of the wines in the wine dataset. From there, the wines with the highest predicted ratings will be recommended.

I will be using python to do my data cleaning and analysis. I will use pandas for my data cleaning and the analysis tools we learned in class for the regressions. I will include functionality in my code to either read in data from the google sheet where the ratings data is stored, or a downloaded excel file. This requires the gspread python package and setting up google API access to my google drive.

My main expected result is a new dataset containing the various wines from the wine dataset and the predicted ratings. I will also have functionality in my code to read this dataset to give recommendations with an option for filters, such as a maximum price or type of wine. I also have various side goals for this project. One goal is to track my project on GitHub. I also want to do additional analysis unrelated to predicting wine recommendations, such as analyzing who brings the best wine, which people have the most similar tastes, how similar or different our reviews are compared to the online reviews, and generating a set of data containing information about each person’s taste in wine and their highest ratings.