Project Report

Purpose: Purpose of project is to examine aspects of different type of coffee beans (auroma, flavor, taste and etc).

Data Sources: Two type of Coffee bean (Arabica, Robusta) data was gathered from Kaggle.com in the format of csv. Arabica and Robusta each has one individual CSV file. Both have listed coffee beans by different country of origin, processing method, region and both have scores for different aspects: aroma, flavor, sweetness, body, balance and overall score. Two files are similar in content but has slightly different column headers. Therefore, in order to do analysis on two coffee beans, transformation on column headers are needed in order to consolidate two files into one.

Data Transformation:

1. Aligning Column Headers: As mentioned above, two files have different column names for same content. Renaming column names to be consistent is needed
2. Removing useless columns: The original file came in 44 columns which includes additional and useless information. I am only interested in scoring information and basic information about coffee bean. Therefore, extra columns were removed to keep data streamlined

Load:

I created a database (Coffedd\_db) in PGAdmin and load transformed data into two tables. Since data came in a very structured format, relational database is chosen. Since we want to analyze two type of coffee beans togethers, two tables were merged into one table in PGAdmin.