CAPSTONE PROJECT README

Information

This project encompasses several Jupyter notebooks and .csv files that will be outlined in chronological order to understand the project flow.

The project was to create a recommender system based on beer reviews found in a dataset.

Inspiration for the project is drawn from this paper:

https://nycdatascience.com/blog/student-works/ninkasi-beer-recommender-system/

Dataset

Dataset location: https://data.world/socialmediadata/beeradvocate

Dataset information:
This dataset consists of beer reviews from beeradvocate. The data span a period of more than 10 years, including all ~1.5 million reviews up to November 2011. Each review includes ratings in terms of five "aspects": appearance, aroma, palate, taste, and overall impression. Reviews include product and user information, followed by each of these five ratings, and a plaintext review. We also have reviews from ratebeer.

Dataset statistics:
- Number of reviews: 1,586,259
- Number of users: 33,387
- Number of beers: 66,051
- Users with > 50 reviews: 4,787
- Median no. of words per review: 126 <- no reviews were included in the dataset
- Timespan: Jan 1998 - Nov 2011

Jupyter Notebooks

All Jupyter notebooks were composed and run in the base environment, all packages imported are denoted at the beginning of each notebook for their intended purpose and function to run the code and produce the outputs wanted/needed for the cleaning, analysis and/or modeling conducted.

Notebook order:
1) Capstone EDA
2) Capstone KNN
3) Capstone Logistic Regression
4) Capstone Recommender System

Data folder:
Within the 'data' folder there are five (5) .csv files that were used for the project.

1) beer_reviews.csv: The original dataset taken from the link shared at the beginning of this README.
2) brewery_location.csv: The file containing brewery name, city and country that was created to merge into 'beer_reviews.csv'.
3) beer_taste.csv: The file containing beer styles and their requisite tasting notes. This file was merged with the original 'beer_reviews.csv' file
4) capstone.csv: This file is our cleaned dataframe that is used with the recommender system notebook, exported from the Capstone EDA notebook.
5) logreg.csv: This file was created in the Capstone kNN notebook and exported for use with the Logistic Regression notebook.

Executive Summary & Presentation folder:

Contains the executive summary and final presentation, both in .pdf format.