**Global Beer Trends**

****

**ETL Project**

**Cameron Kilpatrick, Jacob Stegal, Yash Patel, & Yuli Likhonina**

**E**xtraction:

The two main sources for global beer data extraction include Kaggle and the Open Beer database. These data sources contributed two CSV files in our extraction process. The first CSV data set explores beer reviews in aroma, appearance, palate, and taste across globally located breweries. The data set correlates reviews to each type of beer style and beer name across breweries. The main areas of interest from this data set are the top signature beers and breweries across the globe based on the provided cirtiera reviews. The second CSV file furthers the scope and includes the brewery locations to join on the brewery names. This allows an insight into the leading cities, states, and countries for top breweries based on global beer reviews.

**T**ransformation:

Jupyter Notebook was used in this transformation process with a data model using the primary key as brewery. As the CSV data files are from two different sources the main key in the transformation process was to join the two files and eliminate breweries not on the brewery location list associated with beer reviews. The data set required a removal of all duplicate values created by multiple reviews for the same brewery beers and a drop all of the NA values to produce a clean count as well. The original brewery location CSV file also had additional blank columns included that were not relevant or useful to the overall data set that were dropped. All of the beer review criteria were averaged for one final review output per beer. From there, column names were renamed to be uniform across the dataset. The final transformed data was then extracted as a single CSV file to SQL pgAdmin for final database analysis.

**L**oad:

The transformed CSV file was loaded into SQL pgAdmin for the final database table creation. This was chosen due to SQL’s relational and analytical database benefits to create an overall table that included the brewery, beer names, styles, aroma, appearance, palate, and taste reviews with the locations of breweries. The final data was sorted by reviews over 4 on a scale from 1 - 5 in the dataset to capture the leading breweries and beers across the globe. A further analysis was conducted on the data to narrow down the results and find answers to specific questions on global beer trends for the final data set exploration.