**PROJECT 2**

# Project proposal:

To visualize Beer Data Sets, market values, Ratings, and to find the Top 50 Beer in market.

**Data Sources:**

Data.world

Beeradvocate

brewerydb

# Analysis:

**1-Extract:**

BreweryDB.com API - <https://www.brewerydb.com/>

* The API program and data transfer are located in file: brewerydbAPI.ipynb
* This was done twice, we merged the data, and eliminated the duplicates in Excel.
* The data was stored in a CSV after cleaning it using Jupiter. Check our Resources folder.

Beer Advocate Data - <https://data.world/socialmediadata/beeradvocate>

* We directly downloaded a CSV

Beeradvocate- <https://www.beeradvocate.com/>

* We webscraped this website to find the top 50 beers.
* We built the code to extract all information needed

**2-Transform:**

The datasets were loaded into a MySQL database. We created the database for each one. Then we used the data import wizard to load the CSVs in. We have relational data.

**3-Load:**

We developed a Flask app to load the data into a locally hosted website. You can call all the beer names for each set and you can also call all the data. The data is presented as a list of dictionaries.

# Accountability:

Areej Edaibat

Isaac Asare

Justin Dooley

Sandra Regan

## Data Cleanup & Analysis

Once you have identified your datasets, perform ETL on the data. Make sure to plan and document the following:

\* The sources of data that you will extract from.

\* The type of transformation needed for this data (cleaning, joining, filtering, aggregating, etc).

\* The type of final production database to load the data into (relational or non-relational).

\* The final tables or collections that will be used in the production database.

You will be required to submit a final technical report with the above information and steps required to reproduce your ETL process.

## Project Report

At the end of the week, your team will submit a Final Report that describes the following:

\* \*\*E\*\*xtract: your original data sources and how the data was formatted (CSV, JSON, MySQL, etc).

\* \*\*T\*\*ransform: what data cleaning or transformation was required.

\* \*\*L\*\*oad: the final database, tables/collections, and why this was chosen.