Extract

Our group’s ETL project is based on the Craft Beer project done earlier. We extracted data from 2 csv files. The first is breweries.csv which was cleaned in previous project and had clean data. The second csv is final\_cleaned\_beer\_4.csv which cleaned up the previous file: cleaned\_beer.csv

Transform

Although cleaned\_beer.csv had clean data, it needed to be cleaned further. We were able to get the data into Pandas as a dataframe, and it displayed correctly there. When we tried to put this in the MySQL database, we were running into errors. The reason for the errors was that there were characters in German language, Spanish / French. We did encode the data to utf-8 to allow special characters like slashes, underscore, Pound, but it would still not take the foreign language characters.

To solve this we identified few rows with issues, and modified the characters. There were 3 rows that were not working after modification as well, so we removed them and got the total number of rows from 1403 to 1400 rows.

Load

When creating table in database, we chose 2 tables: beer and breweries.

Breweries table has brewery\_id as the primary key.

Beer table has id as the primary key and brewery\_id as the foreign key referencing the breweries table. The other columns in 2 tables are separate.

We modified the 2 columns with special characters as following to allow loading them in MySQL:

name VARCHAR(100) character set utf8

style VARCHAR(100) character set utf8

We are including the Create Table scripts in our submission.

API

We created a flask app as the backbone of the website using template rendering. Along with template rendering we have a basic API to access the data using basic SqlAlchemy queries.