1. ***Background:***

Beer is one of the oldest drinks humans have produced. The first chemically confirmed barley beer dates back to the 5th millennium BC in Iran, and was recorded in the written history of ancient Egypt and Mesopotamia and spread throughout the world. The word beer comes from old Germanic languages, and is with variations used in continental Germanic languages, bier in German and Dutch, but not in Nordic languages. The word was imported into the British Isles by tribes such as the Saxons. It is disputed where the word originally comes from.

1. ***Rational:***

Beer is an incredibly complex beverage containing more than 3000 different compounds, including carbohydrates, proteins, ions, microbes, organic acids, and polyphenols, among others. Beer becomes even more complex during storage, for over time it may undergo chemical changes that negatively affect the flavor, aroma, and appearance. Thus, it can be expected that maintaining the quality of beer throughout its lifetime is a difficult task. Since it is such a popular drink throughout the world, and based on the fact that consumers everyday require beers as a common beverage with specific enhanced flavor, aroma, appearance or even ABV% profiles, we got motivated to make ourselves familiar with proper data science analytic techniques for beer evaluation using the existence dataset of 1,048,575 that has discussed in Oreilly conference in 2017 to understand what are the factors that make a brand one of the top five highest-ranking beer styles in beer industry based on existed reviews.

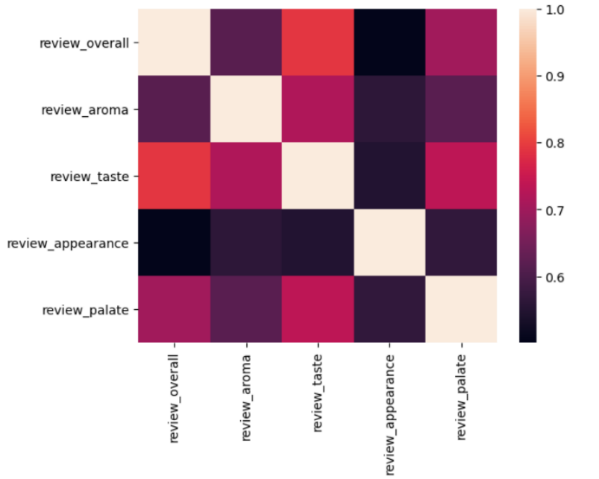
1. ***Data Set:***

This is the dataset originally has discussed in the talk " How to hire and test for data skills: A one-size-fits-all interview kit" from “<https://conferences.oreilly.com/strata/strata-ny-2017/public/schedule/detail/59542>”.

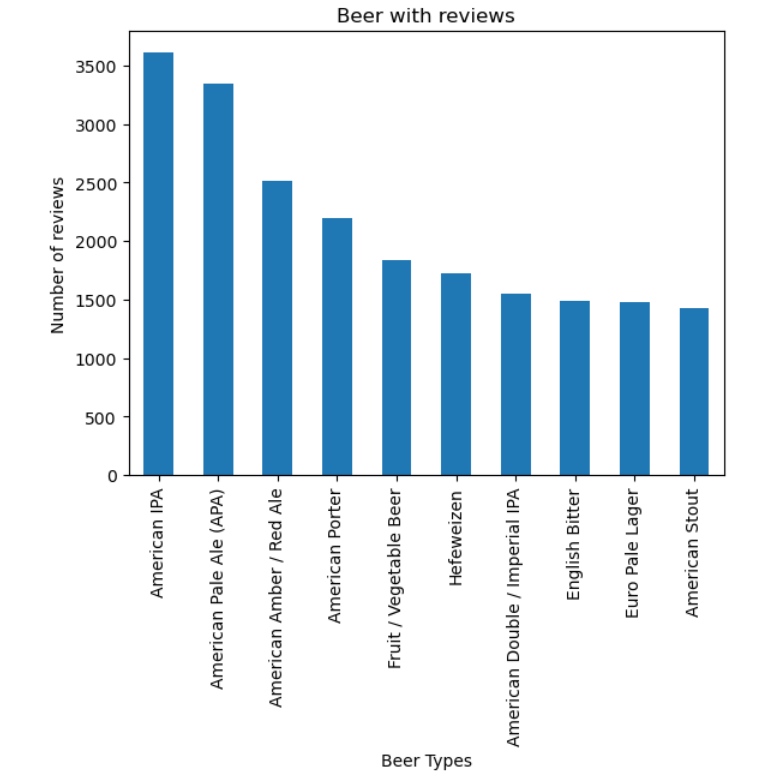
However, we have gotten the data set from Kaggle at “<https://www.kaggle.com/rdoume/beerreviews>”.

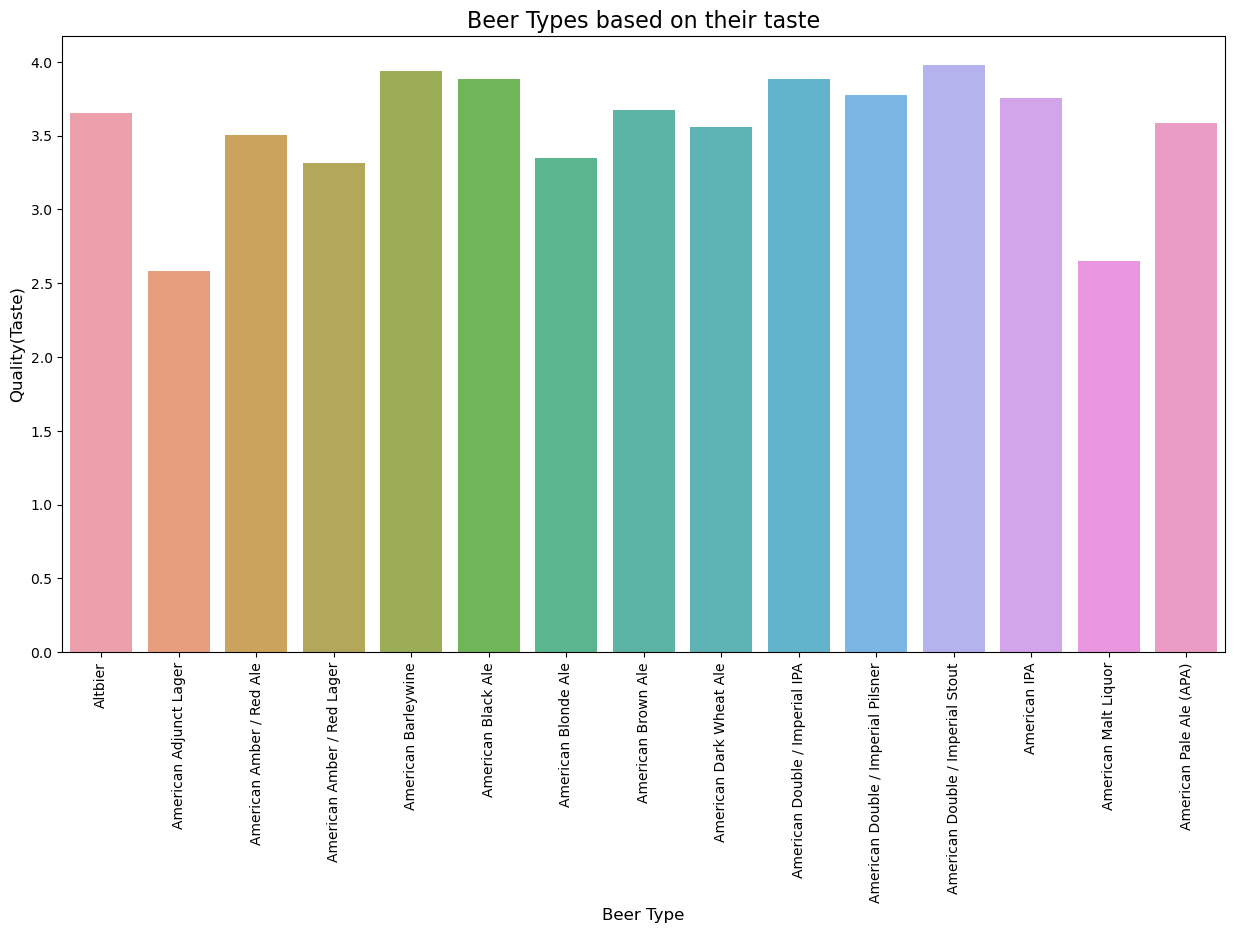
1. ***Visualization Inspiration:***

Seaborn heatmap has used in order to understand the role of each factors such as aroma, taste, appearance, palette and many other factors in determining the overall quality of a beer as follow:

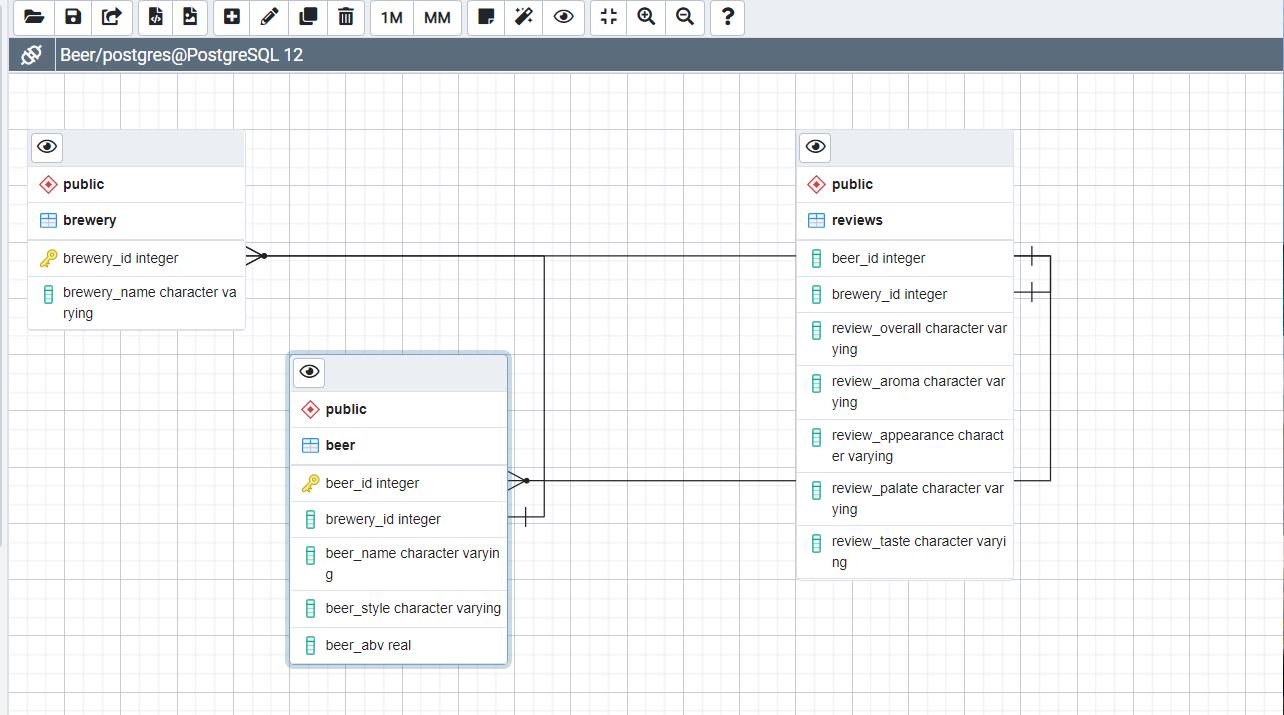


#### There are thousands of different beer style exists in this data set. For the ease of analyzing the dataset, we considered the top 10 beer brand that have more than 30K reviews to find the beers with top ratings in this dataset as follow:





## ***Architecture Diagram:***



1. ***ETL Process***:

* **Extract:** Beer Review data were extracted from [Kaggel Open Data](https://data.torontopolice.on.ca/pages/open-data) source using APIs.
* **Transform**: Using Jupyter Notebook and pandas, we cleaned and reorganized the data according to our needs.
* **Load**: Looking at the unpredictable structure of our data from our sources, PLpgSQL was our database of choice.

1. ***Data Science Techniques:***

Our project dashboard consists of the following technologies:

* Database: SQL
* Data Visualization: Python ( Seaborn, Matplotlib)
* Programming Language: Python (scikit-learn, pandas, numpy)
* IDE tools: Jupyter Notebook

1. ***Conclusion:***

The five most popular beer styles were evaluated by grouping by the data frame by beer style and taking the mean of the aroma and appearance as well as the number of available reviews and maximum value in each category. This data is then organized by the number of ratings for aroma and appearance followed by average value and maximum value for both categories. In doing so, it was found that the most popular beer styles were, in descending order, American IPA, American Double / Imperial IPA, American Pale Ale (APA), Russian Imperial Stout, and American Double / Imperial Stout.