**1.**

**Problem Statement:**

Which brewery produces the strongest beers by abv?

**Problem Analysis Approach:**

I thought of solving this by directly looking at the abv values and just giving the brewery with highest abv containing beer.

But I wanted my solution to also go over all different kinds of breweries and provide one with the highest mean of abv content.

**Solution Option 1:**

As mentioned above, I just used a straight-forward solution to find the beer with highest abv using pandas and finding the max value and the corresponding brewery details.

**Solution Option 2:**

I iterated over the different breweries and calculated the mean abv value. Finally, I outputted the brewery with highest abv.

**2.**

**Problem Statement:**

If you had to pick 3 beers to recommend to someone, how would you approach the problem?

**Problem Analysis Approach:**

I wanted to solve this problem by taking review\_overall scores into account. But as it is a subjective field, I had to iterate over all reviewers and then look at their best recommendation using the overall field.

**Solution Option 1:**

Looped over all reviewers and took their best reviewed beer. Calculated mode of all those beers and provided top 3.

**Solution Option 2:**

Should have also included columns which were less correlated to review\_overall.

**3.**

**Problem Statement:**

What are the factors that impacts the quality of beer the most ?

**Problem Analysis Approach:**

I thought of providing a straight-forward answer by calculating the correlation coefficient between different columns to review\_overall to show which impacts the quality of beer the most.

**Solution Option 1:**

Displayed correlation graph and matrix

**Solution Option 2:**

Could’ve used Tableau for visualization of the correlation as well.

**4.**

**Problem Statement:**

I enjoy a beer which aroma and appearance matches the beer style. What beer should I buy?

**Problem Analysis Approach:**

I thought of going over the beers with best appearance and aroma from all types of beer styles and then providing a separate table with just beer style and beer name.

**Solution Option 1:**

I used pyspark (Python) for this question as the query was taking a lot of time to run using Pandas.

**Solution Option 2:**

Could use Pandas as well but script would take more than 4 hours to run.