**Introduction:**

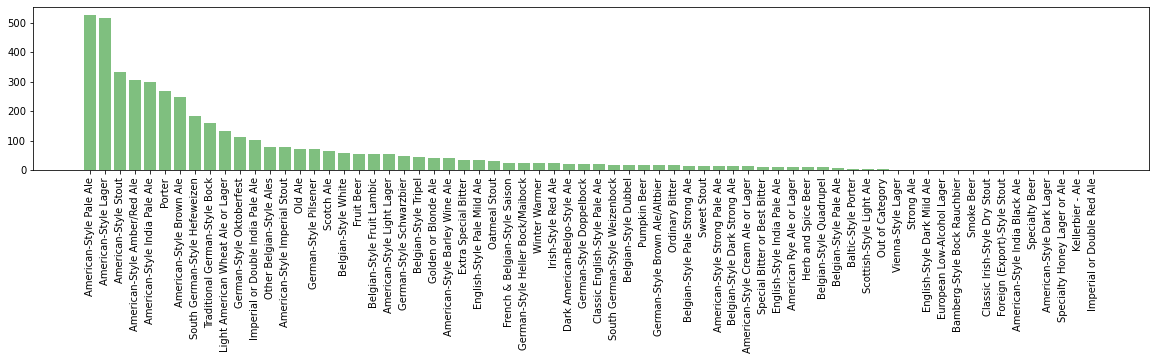
Strange Brew! The basis of the project was interest in homebrew and craft beer. As a group, everyone was interested in the music, so decided to base the project around music venues. Three initial research questions were posed:

1. What are the most popular beer styles nationwide and by state?
2. What is the number of breweries in each state?
3. What are the top cities for breweries?
4. What are the cities with the most breweries per person?
5. What style of beer has the most alcohol content by style?

Research was focused on national breweries and their selection of craft beer. Further, the team looked at the highly consumed beer styles. Brewery, beer, and census data was drawn from 2017. These data sources were collected from published csv and excel files.

**What are the most popular beer styles nationwide and by state?**

The top beer brewed in the United States is the American-Style Pale Ale. The second most brewed beer is the American-Style Lager. Finally, the third most popular beer style is the American-Style Stout. These are types of beers are generic, which may explain the amount of beers within these categories. Further, these beers share a relatively standard alcohol level, hovering in the 4.0 – 6.0% ABV (alcohol by volume) range. Additionally, these styles of beers share a lower bitterness rating, hovering in the 5-50 IBU range.



Each state had its own most popular style of beer. The states with the highest populations—California, Texas, and Florida—generally followed the same trend as on the national level. However, Florida’s most popular style of beer was the American-Style India Pale. This type of beer is less generic than the previous styles mentioned. Additionally, this style of beer has a comparatively higher ABV and IBU.

**What are the number top cities and states for breweries?**

At first, I was using 2 different data frames, Kaggle & Data World, to pull from and found that the different information listed in both data frames were conflicting so in order to get the correct analysis I had to omit the data from Data World. I cleaned up the data by pulling the columns that were necessary and dropped any columns that were not. I then had to count how many breweries were in each State and I output them into a bar chart. Colorado had the most breweries in the US but also not surprising. Seeing that there were 9 states with 2 or less breweries was interesting to me. I thought beer drinking was an American tradition?

After finding the state with the most breweries I then broke down the cities with the most breweries. I ran into an issue with Portland, seeing that there was a Portland in Oregon & in Maine. Once I was able to separate the two Portland’s I then was able to put a bar plot with the top 20 cities. Many cities listed did not have more than 2 breweries and we were only looking for the top cities. Even though Colorado had the most breweries I found that they were all spread out in the state because Portland, Oregon had the most breweries. I think I would have liked to explore a little more into chain breweries as I noticed a few. I’m curious to see if the chains only exist states or if they are planted throughout the country.

**Chart, histogram

Description automatically generatedChart, bar chart, histogram

Description automatically generated**

**What are the cities with the most breweries per person?**

In order to answer this question, I used a Breweries csv data frame from Kaggle as well a population csv data frame from the Census Bureau. The Breweries data was created in 2017 so that was the year I used for the population data as well. There was some difficulty in getting the data frames to merge once they were cleaned but I was able to use the values from the Census data frame to create another population data frame that I could use to merge with my breweries data frame. I narrowed my breweries data frame to just cities that had more than 3 breweries as I decided this would be the most useful values for finding per capita values as there were 29 cities with 3 or more breweries, 75 cities with 2 or more, and 401 cities with at least 1 brewery in the data frame.

I was surprised to find that the city with the most breweries per person is Edwards, CO which is a small town 14 miles west of Vail, CO. It was interesting that Portland, ME had more breweries per person than Portland, OR even though Portland, OR had the most total breweries and 5 more than Portland Maine. Another local city you could visit that had a per capita value of 1 brewery per just under 12,000 residents is Boulder, CO whereas in Denver, CO you are looking at 1 brewery per 88,000 residents. I was very surprised that Fort Collins, CO was not on this list as I thought it was one of the most popular microbrew cities in the country.

Chart, histogram

Description automatically generated

**What style of beer has the most alcohol content by style?**

For this question I used a Beers csv data frame from Kaggle. As this wasn’t a question based on any type of location or population data, I did not need to merge this data frame with any other data frame that we used in this project. I chose to first bin the beers based on alcohol content to see how many beers had low, average, strong, or very strong (One and Done) alcohol content by volume(abv). I then grouped the beers by style and found the average abv for each style. There were 99 different styles of beer but those with the highest average abv were English Barleywine at 10.77%, Quadrupel at 10.4%, American Barleywine and, everyone’s favorite, American Malt Liquor both at 9.9%. For a reference, Bud Light is at 5% abv. The most popular of the types of beers, American IPA, finished at about the average alcohol content of 6.45%. If you are looking for a popular beer with one of the higher abv values of 8.7%, I would suggest looking for an American Double / Imperial IPA. One thing I would have liked to do with this data was dig a little deeper into beers data frame to see range in each of the beer styles in regard to alcohol content versus just finding the average.

**Conclusion:**

After extracting the raw datasets, the data needed to be cleaned. Once the data was cleaned, Strange Brew was able to evaluate the data based on our own assumptions. Many of the answered to the initial research questions were counter than our own-held hypothesis. Although the data is dated a few years, the information gained from the ETL process may be important for key stakeholders in the craft brew industry. The data also revealed information that can be important for determining consumer tastes.