Jared Egerer

CMP 262 - Final Project Summary

9 May 2023

**Tools Used:** Python withthePandas and MatPlotLib PyPlot libraries, Jupyter Notebooks

**Data Source:** <https://www.kaggle.com/datasets/stephenpolozoff/top-beer-information>

**Introduction**

Given the significant gross profit bars and restaurants see on sales of alcohol, it is in their best interest to consider what types of alcohol is most preferred by their target customer base. To that end, it is important that these businesses consider using market research to gauge their customers’ interests and preferences. Data analytics on existing data sources may be used to perform this investigation. This project seeks to analyze information distilled from the beer review website BeerAdvocate.com (via a dataset listed on Kaggle.com) to answer a couple questions:

1. Based on average ratings for breweries, which breweries performed the best and worst overall?
2. Using text analysis, what are the two or three most frequently associated words for the top 10 styles of beer (by user rating)?

**Methodology**

To perform this analysis, the BeerAdvocate data set was imported into Python. This data set was largely clean in its initial state. The cleaning that needed to be completed was removing NaN values. In the two identified instances, the NaNs were beer names where the brewery name was the same. For the first question, the data grouped by brewery and all instances of breweries having fewer than five beers in the data set were removed. The average user ratings for all beers from each brewery were averaged and sorted, and the top and bottom five were identified. For the second question, styles were filtered by those with 20 or more reviewed beers. The average rating was taken for each style, the styles were sorted, and the top 10 were taken. Taste profile descriptors were counted for each style and the top three from each were identified.

**Insights, Recommendations, and Future Work**

When looking at the top and bottom five breweries, it was not surprising that the most highly rated breweries were craft brewers that may not have much exposure outside of the region served by distributors for those breweries. On the other hand, the lowest rated breweries were run-of-the-mill, corporate breweries, such as Anheuser Busch and Miller. Given the audience that BeerAdvocate.com has—that is, beer afficionados—it is not a shock that lower rated brewers are large batch producers. Bars and restaurants should consider having a variety of beers from these breweries to best accommodate all tastes. For a selection of beer from craft brewers, further research can be done to identify regional breweries that are held in high regard as well as smaller producers that may not be well known but are eager to partner with bars and restaurants for better exposure.

In terms of taste profiles, this analysis identifies several beers that are currently in vogue, including IPAs and stouts. The taste profiles identified here for the top beer styles can be used to narrow down beers that show these characteristics. For future studies, beers from certain breweries can be identified and a sentiment analysis can be performed to gain better insight into which beers may be well received.