

A Guide to Selecting Your Next Craft Beer

(Comparing Individual vs Mass User Recommender System Results)

Objective

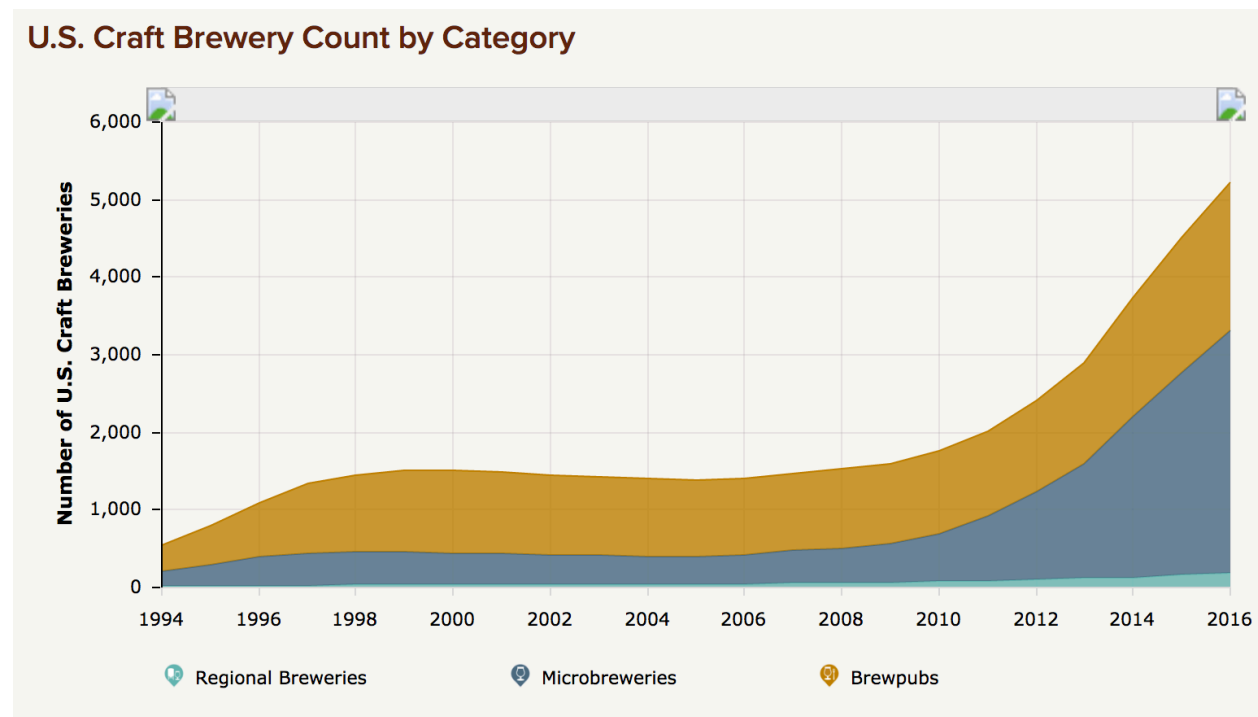
An application like Pandora is great at leveraging mass user data to provide recommendations for your next song, but what if you wanted to have tailored recommendations based on all of the music that you have *personally listened to and rated*? This is the question I will set out to answer in my initial capstone project.

I will create two different recommender systems based on 1) a dataset of ~200 personally rated beers 2) the same dataset but leveraging the average (mass user) ratings. I am looking to compare recommendation and prediction results to understand how my personal tastes differ from the average beer drinker.

Target Audience and Why

My target audience will initially be the Untappd user base, however, believe that any craft beer drinker will find the results interesting.

Since the early 1990s the U.S. craft beer scene has been on the rise. We now have over 5,000 unique craft breweries in the U.S. which may cause newfound confusion for the average consumer when shopping for beer.



Source: <https://www.brewersassociation.org/statistics/number-of-breweries/>

Given the huge rise in the craft brewery scene in the United States, it can be overwhelming trying to decide which new brewery or beer to try. My app will help them decide which beer to try next given their unique tastes and historical ratings.

The Data

All of the data will be pulled via the Untappd API. My personal dataset consists of ~200 distinct beers and ratings. The total beer set to predict will be roughly 10,000+ new beers. Each beer has a number of factors based on brewery, style, my personal rating, mass user rating, location (city, state, country), IBU (integer, bitterness rating), ABV (alcohol content), and potentially other factors that I still need to explore.

Approach

1. Leverage Python's rich libraries to pull the data via API (in JSON format) calls for analysis.
2. Since I am limited to 100 API calls per hour and 50 records per call, I will write a script to pull as much data as I can at one time and repeat until I have a complete dataset.
3. Create my personal beer list (~200 distinct beers) and analyze/clean data.
4. Create untasted beer list (~10,000+ distinct beers) and analyze/clean data.
5. Analysis and factor engineering: understand key drivers to beer ratings.
6. Leverage machine learning methodologies, test multiple algorithms for prediction approach. Select best algorithm.
7. Build prediction ratings from ~200 distinct beers and extract upon 10,000+ beers *based on a unique user's ratings data*.
8. Build prediction ratings from ~200 distinct beers and extract upon 10,000+ beers *based on mass user ratings data*.
9. Compare results from #7 and #8, produce insight and recommendations.
10. Identify challenges with approach / results and provide recommendations on how to improve for future analysis.

Deliverables

Associated code, paper, and presentation. In addition, I plan to publish a blog post as my final presentation.