Identifying the Best Breweries in Seattle, WA using Combined Rating Systems (Decision Matrix Analysis) and Word Cloud Visualizations

IBM Data Science Capstone Project

Prepared by: Todd Hayes

February 2021

Introduction

Background

If you are serious craft beer fan, you know that the craft beer scene has been growing by leaps and bounds over the last decade or so. In fact, the craft beer industry nationwide has increased 357% between 2000 and 2017, making it one of the fastest growing industries in the United States. Seattle is no snoozer when it comes to the craft beer scene. According to a recent report from SmartAsset 'Best Beer Cities for Drinkers - 2019 Edition', the total number of breweries within the neighborhoods of Seattle rank fifth overall in the United States with 66. With so many choices, here lies the dilemma: how does one decide which breweries to visit for the best craft beer experience? Since choosing your next brewery stop is serious business for some, I've narrowed down the must-try spots for you.

Problem

Data sets that may contribute to determining the best breweries in Seattle will include a review of rating systems such as 10-point scales, 5-star ratings, likes/dislikes, tips, and word cloud visualizations that may help to describe each breweries specialties, unique characteristics, and vibe. Details on the data sources and how the information was put together for the top brewery rankings are provided in the Data and Methodology sections below.

Interest

The findings of this study will be helpful for anyone living, working, or vacationing in the area who would like to enjoy delicious and unique craft beer from one of the best breweries in Seattle. This study may also be helpful to investors, sponsors, or business developers seeking to collaborate (partnership) or to consider future business development in the neighborhood.

Data Acquisition and Cleaning

Data Sources

To obtain a list of breweries in the Seattle area I first imported a CSV file containing a list of neighborhoods in Seattle from the Madrona Group website here. A total of 78 neighborhoods were identified. Next, using Foursquare API, I queried the term ‘brewery’ to capture all the venues categorized as breweries within the neighborhoods. I also obtained venue ratings, likes, and tips for each brewery where available using Foursquare API. To complement these datasets, I obtained additional datasets from the BeerAdvocate website here. These include user ratings for 'beers' and user ratings for 'places'. BeerAdvocate breaks the 'beers' score down into five ratable attributes:

- Appearance (look) = 6%
- Smell = 24%
- Taste = 40%
- Mouthfeel (feel) = 10%
- Overall Impression = 20%

Similar to the beer score, BeerAdvocate breaks the 'places' score down to five ratable attributes:

- Atmosphere (vibe) = 10%
- Quality = 30%
- Service = 25%
- Selection = 25%
- Food = 10%

If a place does not offer food, or the rater does not include a food rating, the 10% is spread across the other attributes.

Data Preparation and Cleaning

Since the list of neighborhoods from the source website were not structured as tables, I transformed the list to a CSV file which could be easily read using the Pandas library and converted to a dataframe. I then used the geocoder class of the Geopy client to generate coordinates for each neighborhood using the code snippet below:

```
from geopy.geocoders import Nominatim # convert an address into latitude and longitude values
geolocator = Nominatim(user_agent="tor_explorer")
geocodes = [geolocator.geocode(df['Neighborhood'][i] + ', Seattle, WA')for i in range(len(df))]
```

Next, I wrote a function that contained a URL to fetch data from the Foursquare API that included a search query for breweries within a radius of 1,000 meters of each neighborhood. The brewery information gathered includes the venue name and corresponding coordinates (latitude and longitude), venue category, and venue id. This is what it looks like:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category	Venue ID
0	Belltown	47.613231	-122.345361	Cloudburst Brewing	47.611565	-122.345212	Brewery	54f8d7cd498e011cb93ae3c9
1	Belltown	47.613231	-122.345361	Old Stove Brewing Co - Marketfront	47.609591	-122.343041	Brewery	595c0cb5123a195de691c5ea
2	Belltown	47.613231	-122.345361	Pike Brewing Company	47.608161	-122.339923	Brewery	40b13b00f964a520e1f61ee3
3	Belltown	47.613231	-122.345361	No Anchor	47.615612	-122.348440	Beer Bar	57def285498ea530e654b0ef
4	Belltown	47.613231	-122.345361	Downtown Spirits	47.617892	-122.341815	Liquor Store	50679a14e4b0890847c12a3f
5	Belltown	47.613231	-122.345361	Tankard & Tun	47.608116	-122.339903	Gastropub	59555e7995a72203b0dcefc1
6	Belltown	47.613231	-122.345361	Elephant & Castle	47.610067	-122.335582	Pub	40b13b00f964a52073f61ee3
7	Belltown	47.613231	-122.345361	Tap House Grill	47.611505	-122.334858	Beer Bar	4a15be09f964a520c2781fe3
8	Capitol Hill	47.623831	-122.318369	Outer Planet Craft Brewing	47.618050	-122.316821	Brewery	53c1db29498eabb50501e673

The dataframe contained 113 rows, many of which were duplicates or categorized as something other than a brewery (see red frame above). I removed the duplicate rows with the help of **`df.drop_duplicates()`** and limited the results to breweries by utilizing `series.str.contains('brewery')`. Once the code was added and executed the dataframe was reduced to 65 rows containing all breweries and no duplicates! Using the Folium mapping library, I was able to render the breweries (see figure below). The brewery count seems reasonable considering that the SmartAsset article reported a count of 66.

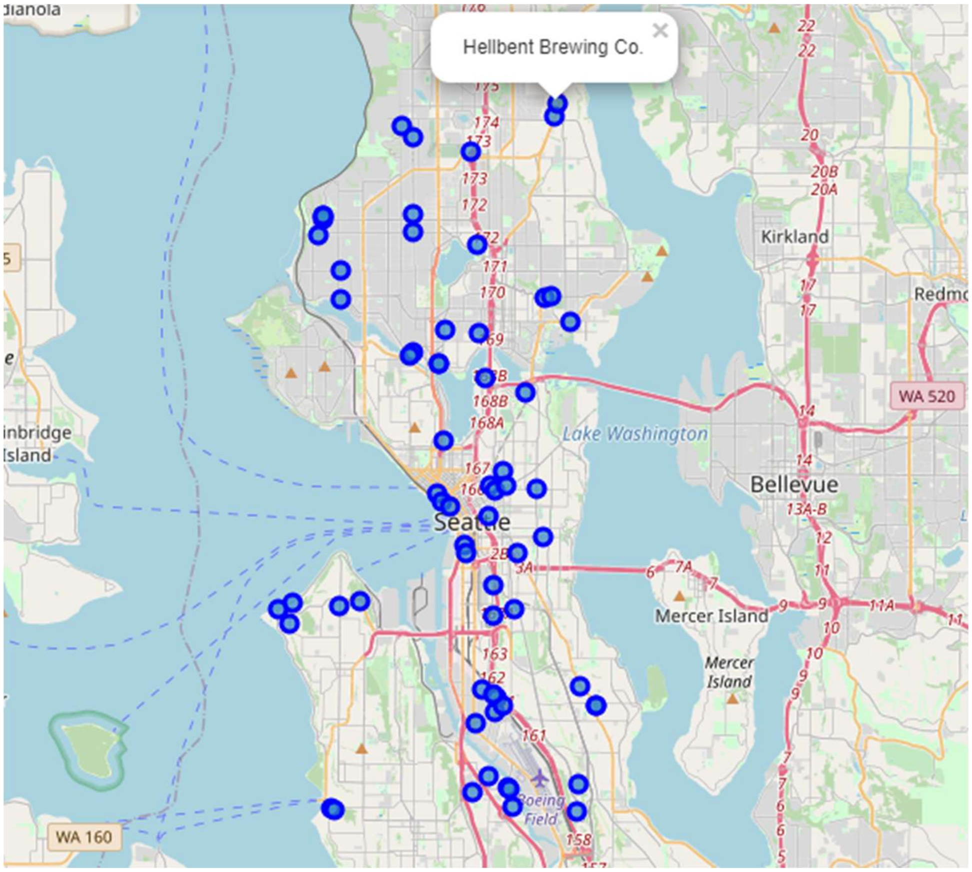


Figure 1 - Map of all breweries identified in Seattle

I then wrote another function that contained a URL to fetch venue details for each brewery using the Foursquare API. The details requested include venue likes, average ratings, and venue tips. Unfortunately, Foursquare did not have rating stats for over half of the breweries on our list (no data available for 34 venues). Therefore, further data analysis for these breweries were not possible.

Methodology

Decision Matrix Analysis

To help evaluate our brewery options with greater clarity, a decision matrix analysis was performed using customer rating system data. A decision matrix is a useful technique to help one to decide among many choices when you have many factors to consider. It is a great technique to use where there isn't a clear winner.

To begin, I examined five features for each brewery:

- Beer rating (5-point scale, source BeerAdvocate)
- Place rating (5-point scale, source BeerAdvocate)

- Overall rating (10-point scale, Foursquare)
- Total ‘like’ counts (5 to 710 range, Foursquare)
- Total ‘tip’ counts (0 to 236 range, Foursquare)

Since these features have varying degrees of magnitude and range, it’s likely that higher weightage would be given to the features with a higher magnitude. This would make it difficult to compare the features with each other. Therefore, to make these comparable, I transformed the data by normalizing the values to a 10-point scale so that all the features contribute equally to the result. Next, I weighted each feature based on relative importance. For instance, I feel that beer quality is the most important feature of a brewery, so I assigned the beer rating feature a multiplier of 1.5. Brewery place rating was assigned a multiplier of 1.0, and both ‘likes’ and ‘tips’ were multiplied by 0.5. The ‘likes’ were weighted less because I could not retrieve dislike counts from Foursquare API and therefore the ‘likes’ number is biased. Tips could be either positive or negative, but I used them to estimate foot traffic and brewery engagement. Since this is just an assumption, I assigned ‘tips’ a reduced multiplier of just 0.5.

Once the features were weighted by relative importance, they were added up for a final weighted score for each brewery. The results of my findings which include brewery rankings are presented in the *Results and Discussions* section below.

Word Cloud Visualization

Once the breweries are ranked using the decision matrix technique, I wanted to try to visualize a summary of the conversations for the top-rated breweries. I did this by evaluating tips (customer reviews and comments) from the top three breweries in the form of word clouds to help gain deeper insights into each brewery’s specialties, unique characteristics, and vibe. A word cloud is a visual representation of text that’s based on the frequency with which a particular word appears. The larger the word in the visual the more common the word was in the customer reviews. However, the words alone do not provide context, so if the word ‘service’ is prominent in your word cloud, it could be referring to good or bad service. The sentiment is unknown. Because of this uncertainty, I was going to parse the tip data into positive and negative data sets. The comments however were overwhelmingly positive, and the negative comments were sparse to non-existent. Therefore, the word cloud visualizations presented in the *Results and Discussion* section below are reflective of positive sentiment for each brewery.

Since the Foursquare API personal account only allowed the retrieval of two tips per brewery, I assembled a list of up to 20 tips for the top three breweries and saved them to an excel spreadsheet. I then read the excel file to a Pandas dataframe.

I then used the WordCloud libraries’ build-in stop word list to omit words that are commonly used in the English language such as ‘the’, ‘a’, ‘in’. I also appended my own stop words to the list that were not helpful. For example, I omitted words like ‘beer’, ‘brewery’, and ‘brew’ since these words are analogous with breweries. Next, I generated a word cloud using the WordCloud function for the breweries. Here is a snippet of the first one:

```
fremont_brewing_wc = WordCloud(
    background_color='white',
    max_words=75,
    max_font_size=50,
    stopwords=stopwords,
    ).generate(text_fremont)
```

The final step is to display the word cloud that was generated using the code below. The matplotlib.pyplot library was used to display the word cloud:

```
# Displaying the WordCloud
fig = plt.figure()
fig.set_figwidth(10) # set width
fig.set_figheight(10) # set height

plt.axis('off')
plt.title('Fremont Brewing and Beer Garden, Fremont\n',fontsize=25)
plt.imshow(fremont_brewing_wc, interpolation='bilinear')
```

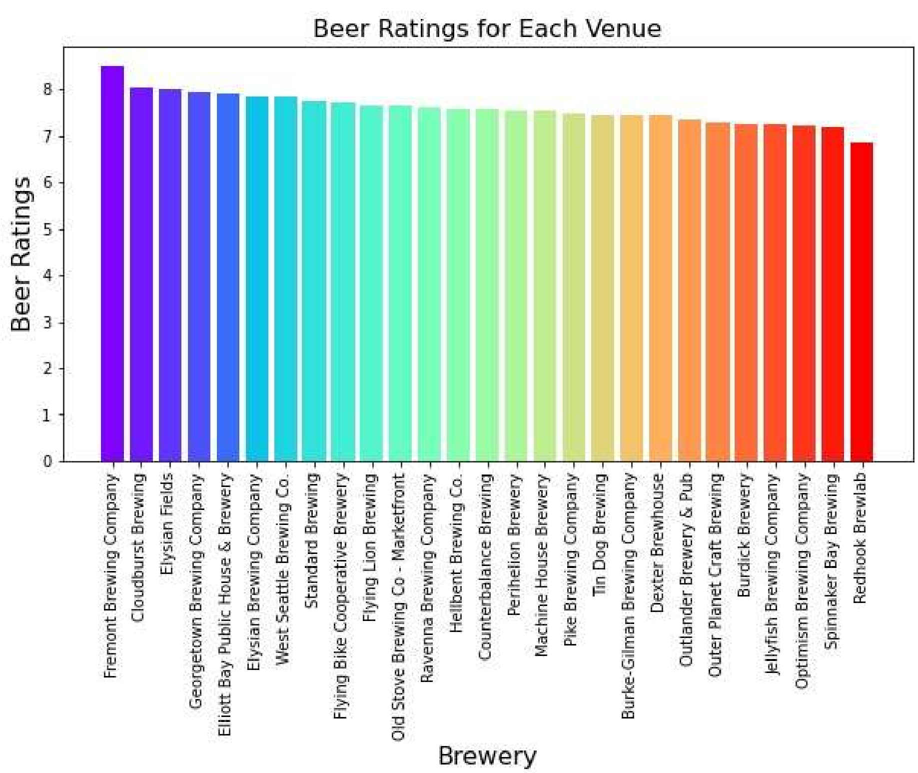
Results and Discussion

In this section we will present bar graphs to visualize the individual features that make up our decision matrix. These includes beer rating, overall/place rating, like counts, and tip counts. Then we will create a final bar graph that includes the sum of the weighted features that will reveal the best (and worst) breweries overall, based on our decision matrix model. Below I utilize the matplotlib bar() function and feed it the vertical and horizontal data, along with a rainbow color designation as follows:

```
all_venue_dist = seattle_brewery_venues_only['Neighborhood'].value_counts()
colors = cm.rainbow(np.linspace(0, 1, len(all_venue_dist.index)))
plt.figure(figsize = (10, 5))
plt.xticks(rotation = 90)
plt.xlabel('Seattle Neighborhoods', fontsize = 16)
plt.ylabel('Brewery Count', fontsize = 16)
plt.title('Number of Breweries in each Neighborhood', fontsize = 16)
plt.bar(all_venue_dist.index, all_venue_dist.values, color = colors)
```

Brewery Ratings by Beer

The beer ratings from BeerAdvocate for each brewery are plotted below:

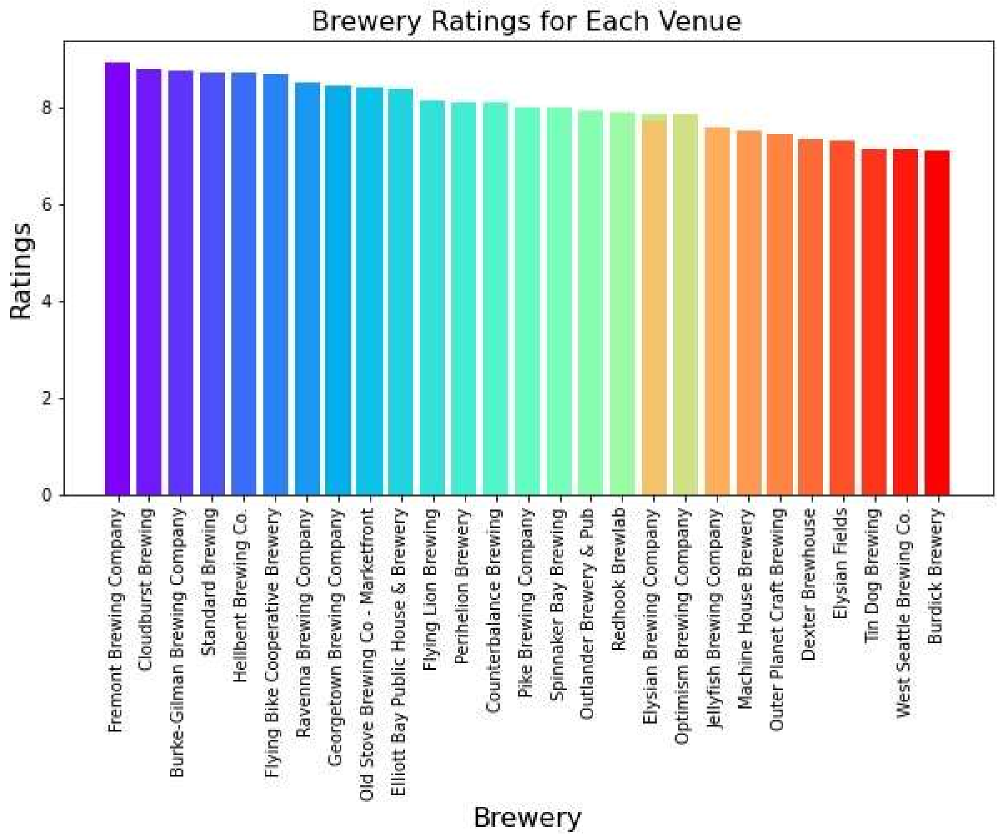


The top five breweries with the highest beer ratings are shown above and listed as follows:

1. Fremont Brewing Company
2. Cloudburst Brewing
3. Elysian Fields
4. Georgetown Brewing Company
5. Elliot Bay Public House & Brewery

Brewery Ratings by Venue

The brewery venue ratings (overall & place) from the Foursquare API and BeerAdvocate were combined, averaged, and then plotted with respect to each brewery. The resulting graph displays the breweries with the highest ratings in terms of place (atmosphere, quality, service, selection, and food):

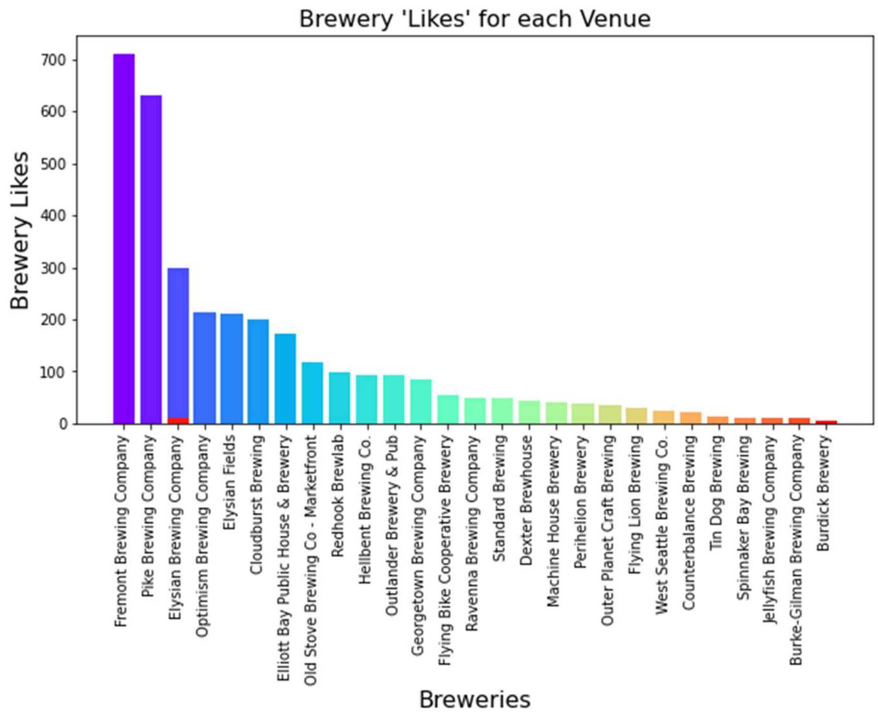


As you can see from the graph above, the top five breweries are:

1. Fremont Brewing Company
2. Cloudburst Brewing
3. Burke-Gilman Brewing Company
4. Standard Brewing
5. Hellbent Brewing Company

Brewery Ratings by 'Likes'

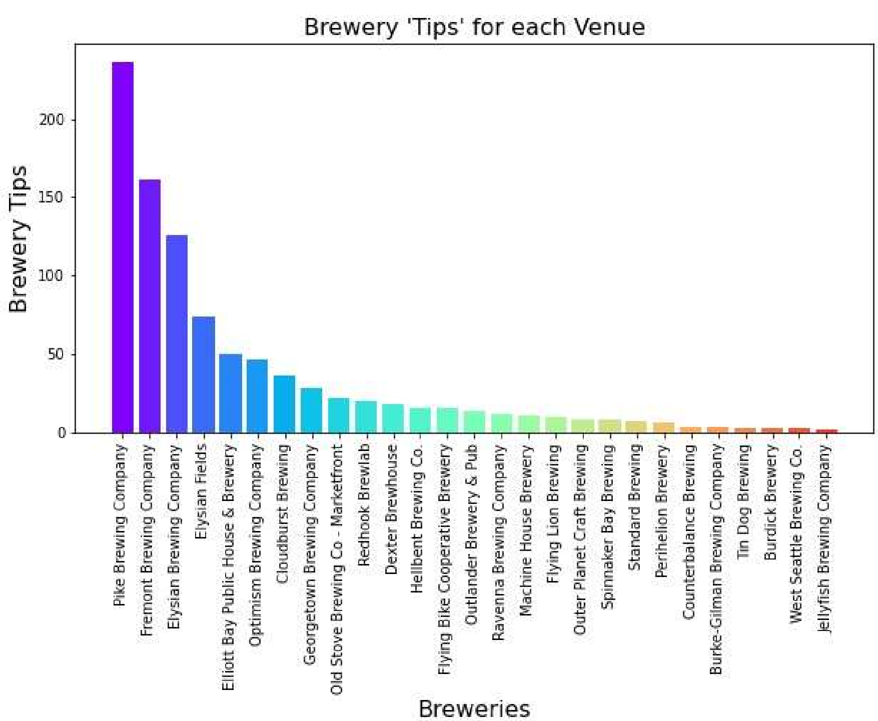
The brewery like counts from the Foursquare API was plotted for each brewery. The resulting graph is shown below:



The top five breweries in terms of likes are as follows: Fremont Brewing Company, Pike Brewing Company, Elysian Brewing Company, Optimism Brewing Company, and Elysian Fields.

Brewery Ratings by 'Tips'

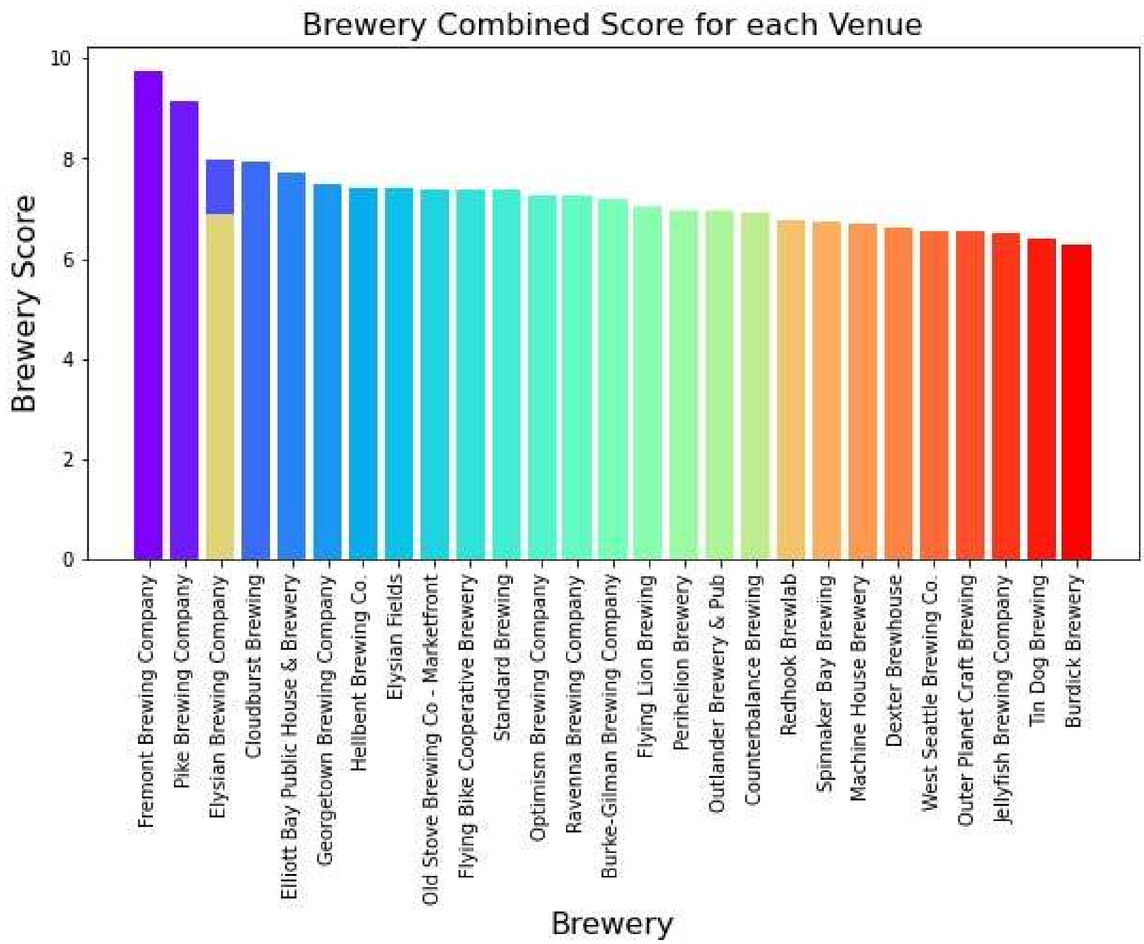
The brewery tip counts from the Foursquare API was plotted for each brewery. The resulting graph is shown below:



The top five breweries in terms of tip counts are as follows: Pike Brewing Company, Fremont Brewing Company, Elysian Brewing Company, Elysian Fields, and Elliot Bay Public House & Brewery.

Total Combined Score by Venue

Here we are! The bar graph below illustrates the decision matrix score that includes the sum of the weighted features. Recall that the multipliers I used for each feature is a follows: beer stats x 1.5, place/overall rating stats x1, like counts/tip counts x1.



The top five brewery rankings overall according to the decision matrix score:

1. Fremont Brewing Company
2. Pike Brewing Company
3. Elysian Brewing Company
4. Cloudburst Brewing
5. Elliot Bay Public House & Brewery

There you have it! These are the top five. Note that Fremont Brewing Company was #1 in all the feature sets except for brewery tips (Fremont came in second after Pike Brewing). Needless to say, Fremont is a standout from the brewery crowd!

To visualize the locations of the breweries analyzed in this study, I've created a leaflet map using Folium to see the distribution of the best breweries in the area.

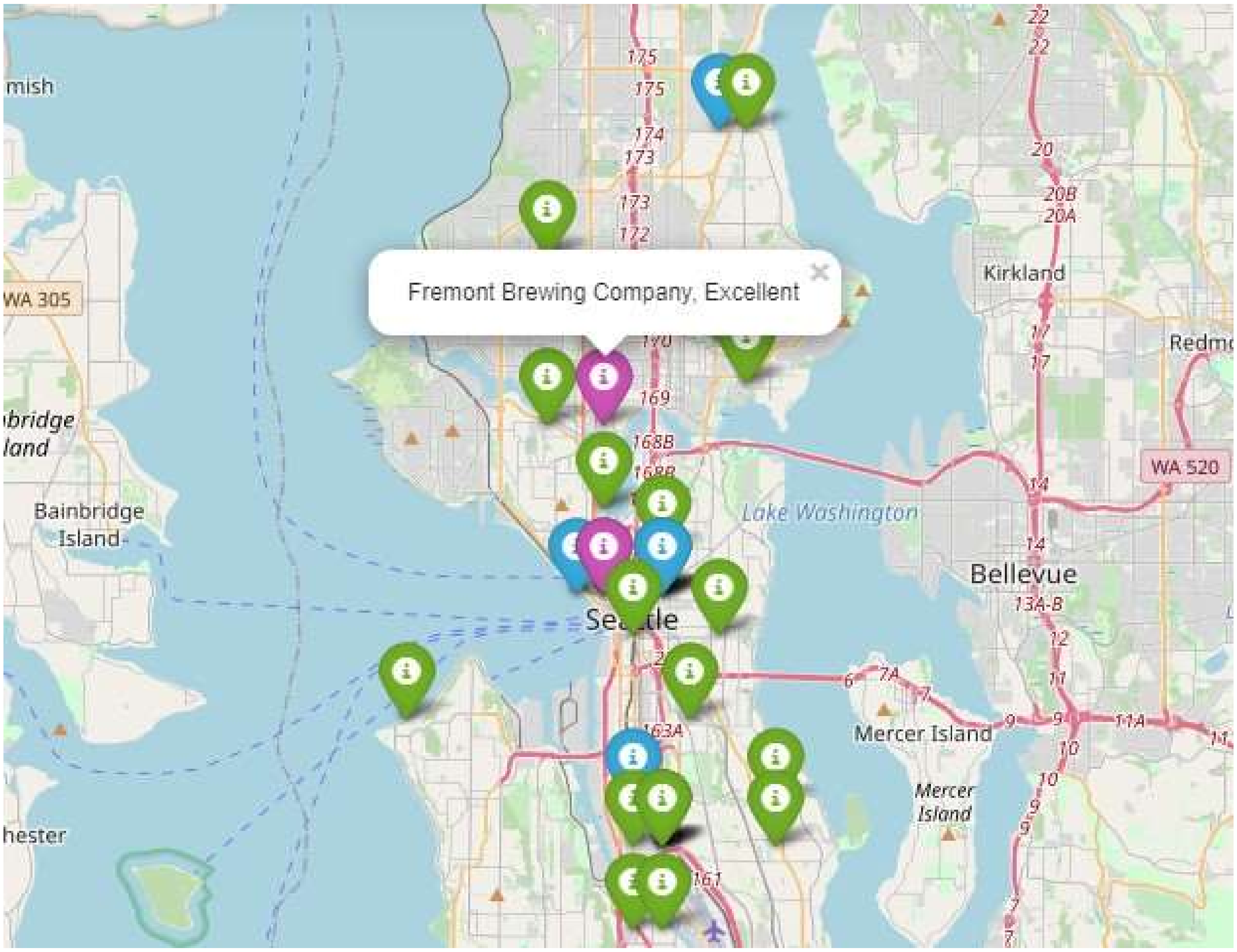


Figure 2 - Map of rated breweries identified in Seattle. The brewery markers are color coded as follows: purple = excellent, blue = very good, green = good to average. The top five breweries we identified are shown in purple and blue.

Now, let's take a deeper look into a few of the top breweries using word cloud visualizations.

Word Cloud Visualization Results

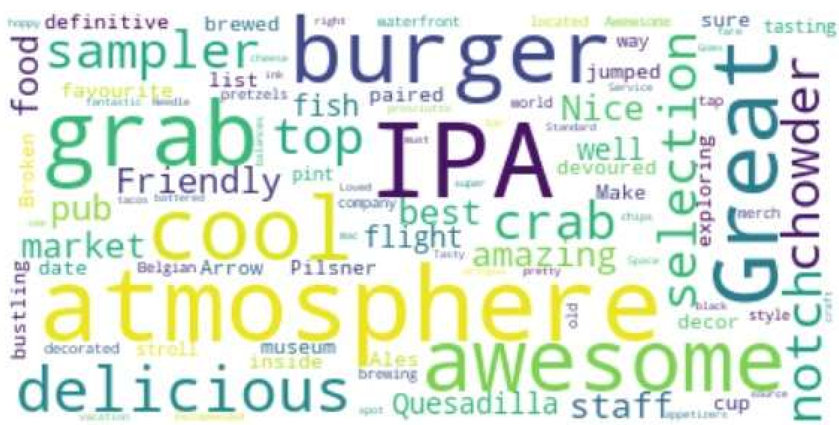
Following the quantitative analysis of each brewery, the customer tips from the top three breweries were used to create a word cloud. The word clouds are reflective of positive sentiment for each brewery. Alongside of each word cloud is a table showing the top 10 reoccurring words ordered by frequency in each cloud. This will show what aspects the customers visiting each venue talk about most.

[Fremont Brewing Company](#)

Fremont Brewing:

outdoor	1.000
great	1.000
friendly	0.875
dog	0.750
free pretzels	0.625
kid	0.500
seating	0.500
best	0.500
outside	0.500
family	0.500

[Pike Brewing Company](#)



Pike Brewing:

IPA	1.00
atmosphere	1.00
grab	0.75
Great	0.75
burger	0.75
awesome	0.75
cool	0.75
delicious	0.75
sampler	0.50
crab	0.50

[Elysian Brewing Company](#)

Elysian Brewing:

food	1.000
great	0.750
amazing	0.375
hummus	0.375
seasonal	0.375
atmosphere	0.375
craft	0.250
best	0.250
meatloaf	0.250
every	0.250

As you can see, the word clouds give us a glimpse into what customers are talking about. Interestingly, the frequent words used to describe each brewery are quite different. For instance, popular words to describe Pike and Elysian Brewing revolve around food such as ‘burger’, ‘crab’, ‘hummus’, and ‘meatloaf’, whereas Fremont Brewing contains popular words such as ‘outdoor’, ‘family’, ‘friendly, ‘free pretzels’ to name a few. In addition, I am really one to enjoy an India Pale

Ale (IPA) style of beer, and I see that IPA is the top word at Pike Brewing along with atmosphere. So, if I want to enjoy a burger with an IPA, I may take a closer look at Pike Brewing.

Conclusions

In this study, I used a decision-making tool called a decision matrix to evaluate breweries by weighing their feature sets based on perceived importance and generating an overall score to compare against other breweries in the Seattle area. The brewery features that were used include customer beer ratings, place ratings, overall ratings, total likes, and total tips. The data was acquired from Foursquare and BeerAdvocate. The magnitude of the weight applied to each feature was subjective and not based any external studies. For example, I used a multiplier of 1.5 for beer ratings, 1.0 for place ratings, and 0.5 for total likes and tips. Others may feel differently, and believe that place ratings take precedence, for example. However, based on the decision matrix analysis performed and the brewery rankings observed, I conclude that the decision matrix model performed well in identifying the top breweries, which was my goal of this study. The word clouds prepared also gave me interesting insights on what customers were saying about each venue and gave me impressions about the brewery's atmosphere, unique characteristics, and vibe, whether real or perceived.

Limitations and Future Work

Future work might include a trip to these breweries with a friend or group to see for yourself how the breweries stack up to one another using the feature sets described in this report. The breweries were ranked solely on the data provided by Foursquare and BeerAdvocate. Data from other reliable sources would likely enhance the study.