SOUR GLASS ALE WORKS

Using Data Science to Assist in Branding Strategies

Manny Navarro

Contents

1. Introduction

1.1 Background

Continuing the traditions of American creativity, entrepreneurship and innovation, the Brewer's Association reports 4,522 microbreweries were operating in the United States in 2018, a 15% increase from 2017. With the craft beer industry still going strong, many microbreweries look for every opportunity to build and retain loyal customers through various branding strategies. Sour Glass Ale Works is one such microbrewery located in Southern California where the company specializes in locally sourced, small-batch artisan beer.

1.2 Problem

The Strategic Planning team of Sour Glass Ale Works understands that in a highly competitive craft beer industry, building brand awareness through pleasurable experiences is vital toward a lasting relationship with its customers. Therefore, to successfully share its craft directly to consumers, a major strategic initiative is to increase brand awareness in the Southern California region by opening a tasting room in San Diego, California. The Strategic Planning team has requested an analysis on consumer expenditure, nearby competition, and location recommendations around a five kilometer radius from the downtown area.

1.3 Stakeholders

The Strategic Planning team of Sour Glass Ale Works is comprised of four members. Each member, along with a short description of their role, is summarized below:

- Chief Executive Officer: Sets, drives, and enforces the strategic vision as well as directing the execution of operating plans and company goals.
- Director of Operations: Oversees all areas of production, packaging and distribution.
- Director of Marketing: Oversees the design and implementation of brand consistency, presence, and strategic alignment to business objectives.
- Regional Sales Manager: Responsible for the development and execution of growth sales strategies and building customer relationships.

2. Data

2.1 Data Sources

The United States Bureau of Labor Statistics provides consumer expenditure information for major metropolitan areas including San Diego, California. Updated on a yearly basis, the latest available data on consumer expenditures were averaged over the 2016-17 time period. The data will be used to examine the percentage of average annual expenditures allocated to alcoholic beverages in the San Diego area.

Foursquare is a popular social networking application which enables users to discover and share information about venues (e.g., restaurants, bars, parks, etc.) in a given geographical area. The company offers an API which I will leverage in order to retrieve venue names, venue categories, and latitude/longitude coordinates to conduct an analysis on competition and possible tasting room locations and within the San Diego area.

The San Diego Regional Data Library provides information containing the metro statistical areas (MSA) and sub-regional areas (SRA) in the SANGIS Region Dataset. For the purposes of this analysis, we will define sub-regional areas as neighborhoods and utilize their latitude/longitude coordinates to analyze relationships between neighborhoods given the Foursquare venue data.

2.2 Data Wrangling

Consumer Expenditure data for the San Diego area was scraped from the United States Bureau of Labor Statistics website and read into two Pandas DataFrames. For the purposes of this analysis, only the expenditure data was needed for this analysis, the average annual expenditures in dollars, as well as the percent distribution of expenditures by type were loaded into Pandas DataFrames. Other data such as Consumer Unit Characteristics (e.g., average income, average age) were excluded from the data sets:

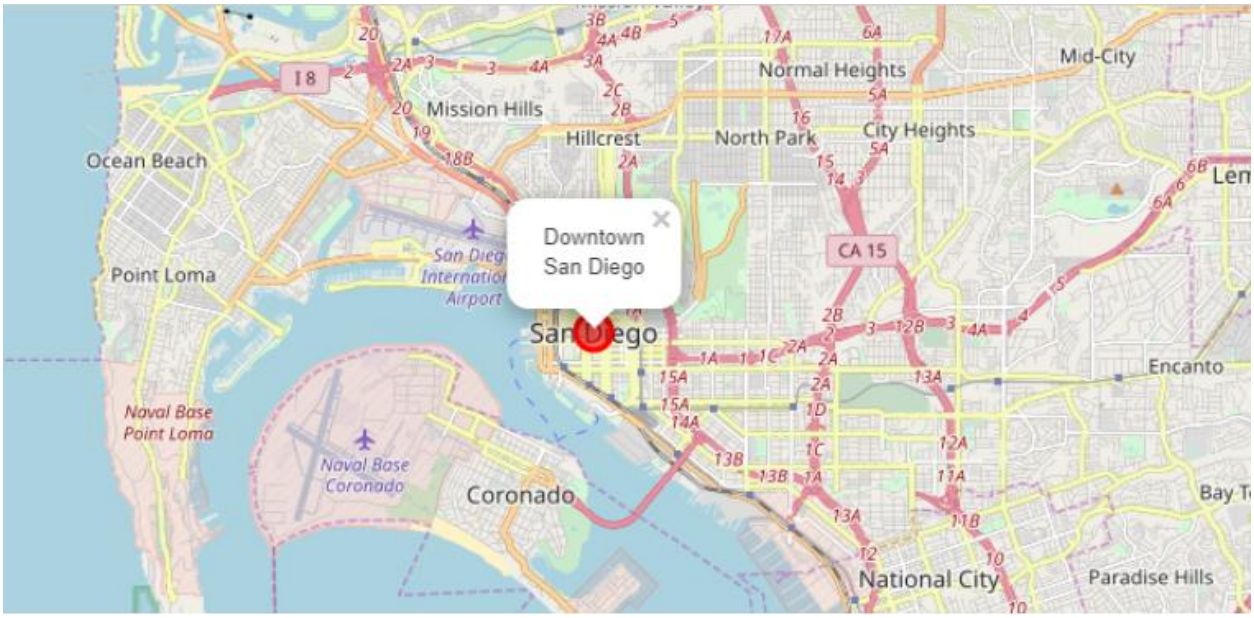
	Area	Average Annual Expenditures
0	United States	58681
1	San Diego	79585

	Category	United States	San Diego
0	Food	12.7	12.5
1	Alcoholic beverages	0.9	1.4
2	Housing	33.0	35.9
3	Apparel and services	3.1	3.2
4	Transportation	15.9	13.3
5	Healthcare	8.1	6.6
6	Entertainment	5.2	5.7
7	Personal care products and services	1.3	1.4
8	Reading	0.2	0.2
9	Education	2.4	3.1
10	Tobacco products and smoking supplies	0.6	0.2
11	Miscellaneous	1.7	1.7
12	Cash contributions	3.4	2.6
13	Personal insurance and pensions	11.6	12.2

Preparing the San Diego brewery data involved four steps. First, the Geopy library was used to obtain the latitude and longitude coordinates for San Diego:

```
The geograpical coordinates of San Diego are 32.7174209, -117.1627714.
```

Once these coordinates were obtained, the second step applied the coordinates as inputs to generate a map of San Diego using the Folium library. These coordinates are represented as the red circle in the map below:



The third step involved using the Foursquare API to search for venues containing 'brewery' as the keyword. I limited the results to a five kilometer radius, then converted the venue data into a Pandas DataFrame:

	name	categories	address	cc	city	country	crossStreet	distance
0	Karl Strauss Brewery & Restaurant	Brewery	1157 Columbia St	US	San Diego	United States	at W B St	414
1	Bolt Brewery	Brewery	1971 India St	US	San Diego	United States	NaN	1050
2	Ko Underground Brewery	Brewery	1320 5th Ave	US	San Diego	United States	1320 5th Ave	277
3	Belching Beaver Brewery Tasting Room	Beer Bar	4223 30th St	US	San Diego	United States	El Cajon Blvd	5149
4	San Diego Brewery Tours	Tour Provider	241 14th St	US	San Diego	United States	NaN	1367

The fourth and final step involved a review of the Brewery data. Here, I discovered 'noise' during a review of the DataFrame results. For example, the fifth record below displays 'San Diego Brewery Tours' as the name of the venue and is classified as a 'Tour Provider' which is technically not a 'brewery' venue I wanted to capture. Furthermore, 'Real Estate Offices' and 'Residential Buildings' I also wanted to exclude from the Brewery data set:

categories	
Beer Bar	2
Beer Garden	1
Brewery	14
Gastropub	1
Real Estate Office	1
Residential Building (Apartment / Condo)	1
Tour Provider	2

The DataFrame was then cleansed by keeping the following categories only: 'Beer Bar', 'Beer Garden', 'Brewery', and 'Gastropub':

categories	
Beer Bar	2
Beer Garden	1
Brewery	14
Gastropub	1

Here is an example of the revised DataFrame which contained 18 total brewery related venues within a five kilometer radius of downtown San Diego:

	name	categories	address	cc	city	country	crossStreet	distance	formattedAddress	labeledLatLngs	lat	lng
0	Karl Strauss Brewery & Restaurant	Brewery	1157 Columbia St	US	San Diego	United States	at W B St	414	[1157 Columbia St (at W B St), San Diego, CA 9...	[{'label': 'display', 'lat': 32.71752458907822...	32.717525	-117.167194
1	Bolt Brewery	Brewery	1971 India St	US	San Diego	United States	NaN	1050	[1971 India St, San Diego, CA 92101, United St...	[{'label': 'display', 'lat': 32.72534560154953...	32.725346	-117.168871
2	Ko Underground Brewery	Brewery	1320 5th Ave	US	San Diego	United States	1320 5th Ave	277	[1320 5th Ave (1320 5th Ave), San Diego, CA 92...	[{'label': 'display', 'lat': 32.719048, 'lng':...	32.719048	-117.160530
3	Belching Beaver Brewery Tasting Room	Beer Bar	4223 30th St	US	San Diego	United States	El Cajon Blvd	5149	[4223 30th St (El Cajon Blvd), San Diego, CA 9...	[{'label': 'display', 'lat': 32.75460899954223...	32.754609	-117.130057
5	Thorn Street Brewery	Brewery	3176 Thorn St	US	San Diego	United States	at 32nd St	4263	[3176 Thorn St (at 32nd St), San Diego, CA 921...	[{'label': 'display', 'lat': 32.739407, 'lng':...	32.739407	-117.125496

The San Diego neighborhood data was easily accessible via download. Formatted as a CSV file, the data was loaded into a Pandas DataFrame:

	tract	sra	sra_name	msa	msa_name	lon	lat
0	15.0	1	CENTRAL SAN DIEGO	0	CENTRAL	-117.123317	32.743750
1	16.0	6	MID-CITY	0	CENTRAL	-117.121322	32.752161
2	17.0	6	MID-CITY	0	CENTRAL	-117.120640	32.758246
3	18.0	6	MID-CITY	0	CENTRAL	-117.118525	32.762886
4	19.0	6	MID-CITY	0	CENTRAL	-117.119155	32.769539

Viewing the DataFrame, it was discovered that a MSA can have many neighborhoods (SRA), and a neighborhood can have many census tracts. I then summarized the count of tracts by MSA:

```
msa_name
CENTRAL              148
EAST COUNTY            6
EAST SUBURBAN        102
NORTH CITY           151
NORTH COUNTY EAST     81
NORTH COUNTY WEST     69
SOUTH SUBURBAN        70
```

As the scope of this analysis places a five kilometer radius from downtown San Diego, I filtered the neighborhood data to include the CENTRAL MSA only. This revealed that CENTRAL contains six distinct neighborhoods:

```
CENTRAL SAN DIEGO
MID-CITY
SOUTHEASTERN SAN DIEGO
PENINSULA
CORONADO
NATIONAL CITY
```

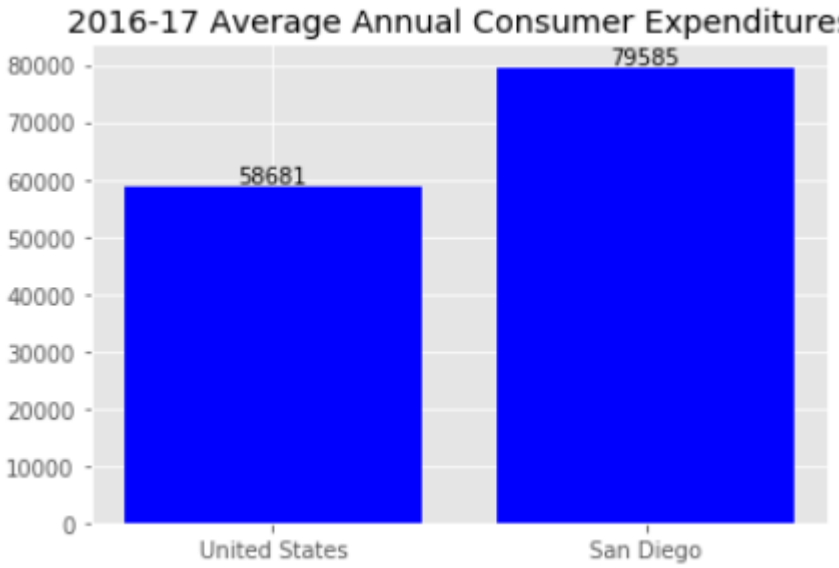
These neighborhoods, along with the corresponding census tract coordinates where then used to search Foursquare for nearby venues and loaded into a Pandas DataFrame containing 2,269 rows and seven columns:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	CENTRAL SAN DIEGO	32.743750	-117.123317	Panchita's Kitchen & Bakery	32.747576	-117.124877	Bakery
1	CENTRAL SAN DIEGO	32.743750	-117.123317	Chris's Market	32.741860	-117.120750	Liquor Store
2	CENTRAL SAN DIEGO	32.743750	-117.123317	Vintage Religion	32.747923	-117.124844	Art Gallery
3	CENTRAL SAN DIEGO	32.743750	-117.123317	North Park Nursery by Mooch	32.739580	-117.125186	Garden
4	MID-CITY	32.752161	-117.121322	Streetcar Merchants Of Fried Chicken, Doughnut...	32.749813	-117.120060	Fried Chicken Joint

3. Methodology

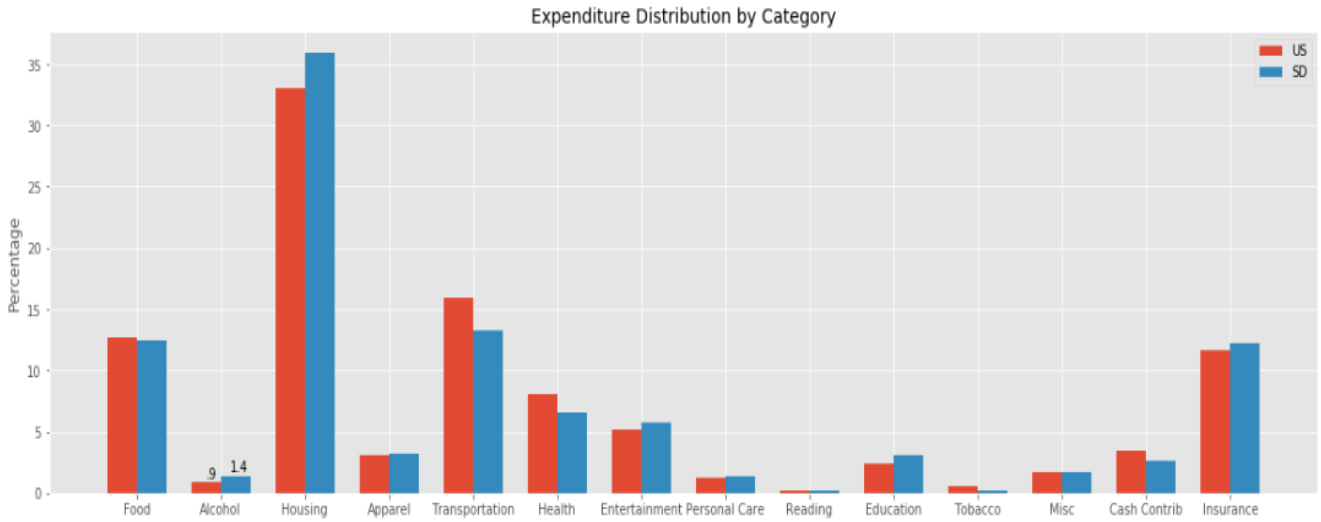
3.1 Exploratory Data Analysis

To understand the San Diego market in terms of the voluntary consumption of goods and services, I plotted the average Consumer Expenditure data to compare San Diego versus the United States:



The chart explains that San Diegans spent an average of $79,585 per year between 2016 and 2017. This average expenditure was $20,904 more than the United States average.

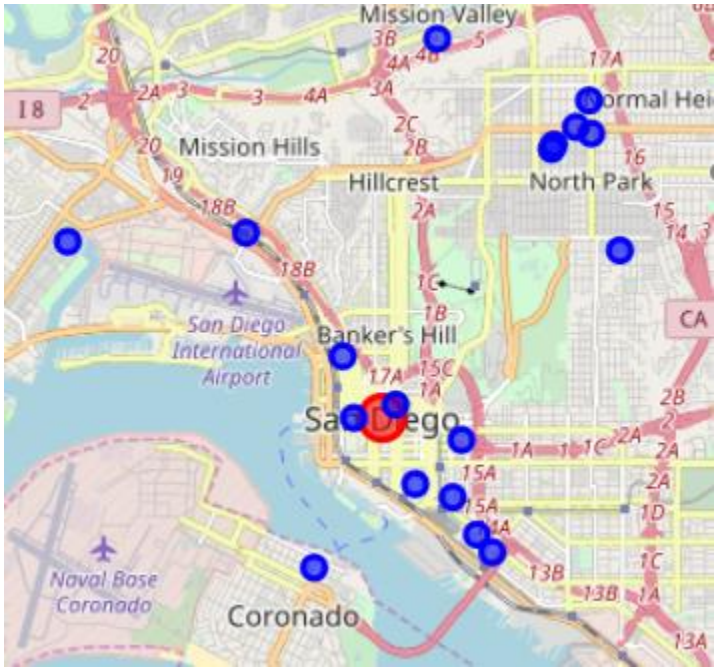
I examined the average expenditure further by visualizing the percentage distribution by category:



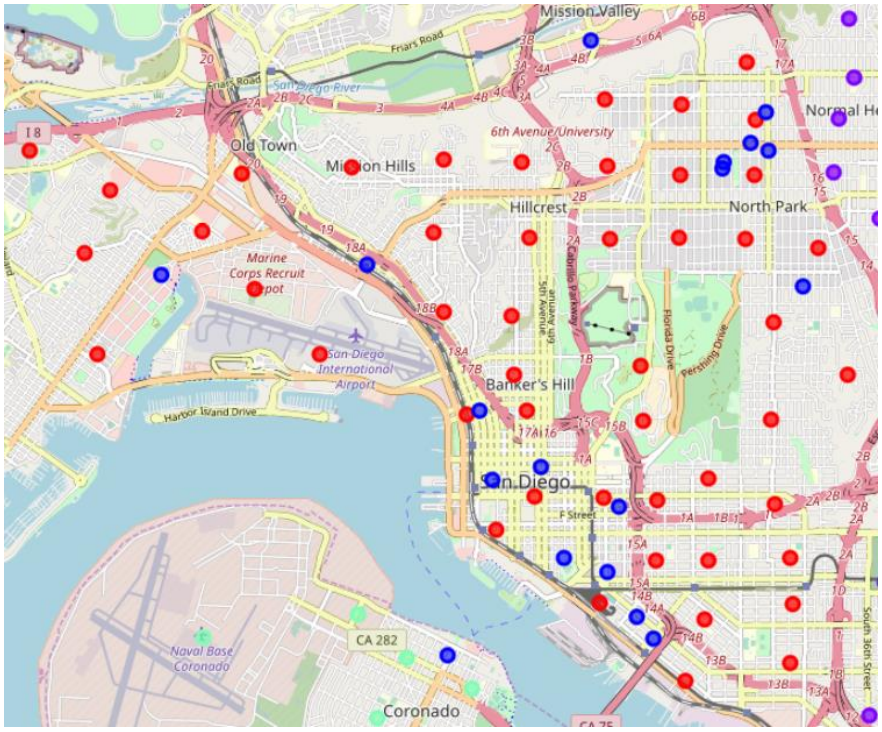
The side by side comparison of average expenditures on alcohol reveals that on average, San Diegans spend more than the United States average, 1.4% ($1,114.19), .9% ($528.13) respectively.

A key item to note regarding the consumer expenditure data is it averaged at the San Diego metropolitan level and may not be truly indicative to expenditures of downtown San Diego specifically. Since it is likely that downtown San Diego is both a tourist attraction and entertainment district, expenditures for this area may be underestimated.

From a competitive analysis perspective, I first looked at the brewery data gathered from Foursquare. As highlighted in the map below, we see the 18 breweries grouped into two clusters across different neighborhoods of San Diego.



I then applied the additional venue data gathered from Foursquare and overlaid them on the map. To further gain some insight into underlying patterns of neighborhoods and their venues, I applied a k-means algorithm (where k=3) from the scikit-learn library to cluster similar neighborhoods together:



The results of the k-means clustering allowed for a few interesting findings. First, the breweries grouped in the downtown area all fell within Neighborhood Cluster 0 (red color) where coffee shops, Mexican restaurants and American restaurants are most common. The breweries grouped closer to mid-city fell between Neighborhood Cluster 0 and Neighborhood Cluster 1 (purple color) where Vietnamese, Mexican, and Pizza restaurants are most common. Finally, the brewery located on Coronado Island fell in Cluster 2 (light green color) where coffee shops, grocery stores and American restaurants are most common.

4. Results

The analysis on San Diego consumers and venues revealed a variety of insights into whether an opportunity for a tasting room from Sour Glass Ale Works exists. First, it was discovered that on average, San Diego consumers spend more on alcohol than the United States average. Secondly, within a five kilometer radius, there are 18 brewery-related venues which tend to cluster near downtown and the mid-city neighborhoods. Finally, clustering neighborhoods by common venues revealed the majority of competition (outside of other breweries) were restaurants and coffee shops.

5. Discussion

The data gathered for this analysis was sufficient for a high-level understanding the San Diego brewery market, however more data may be necessary to allow a more in-depth understanding of consumer expenditures, such as demographic and expenditure data at the regional/neighborhood at the regional/neighborhood level. Furthermore, the information gathered from this analysis offers Sour Glass Ale Works with a few options for a tap room location. For example, if the company's looking for a location furthest from competing breweries and other venues, the Coronado neighborhood may be the best option. The disadvantage to this is that this is an Island with limited entry/exit points, which may

limit exposure. Choosing a location inland may be a better option, especially of the company looks at restaurants as more of a partner than competitor. Sour Glass Ale Works may consider a 'bring-your-own-food policy' and partnership with nearby restaurants to gain exposure. Future research is needed to test these assumptions with more venue data and data from services such as brewery tour stops, food trucks, beer festivals which may aid in brand awareness.

6. Conclusion

Fostering a relationship directly with your customer is a key component to branding and positioning in a marketplace. This analysis was aimed to provide the key stakeholders of Sour Glass Ale Works with a general overview of the San Diego brewery market. The company's strategic initiative to generate brand awareness by expanding into San Diego is supported by consumer interest and current competition near the downtown area, and looks to be a promising venture for continued recognition and growth.