



# APPLIED DATA SCIENCE CAPSTONE

---

## The Battle of the Neighborhoods

**Gabriel Allen**

Denver, CO



**Table of Contents**

<b>Big yEAST Exploratory Market Research.....</b>	<b>3</b>
<b>Introduction .....</b>	<b>3</b>
<b>Data Requirements.....</b>	<b>3</b>
<b>Methodology .....</b>	<b>4</b>
<b>Results .....</b>	<b>6</b>
<b>Discussion.....</b>	<b>10</b>
<b>Conclusion.....</b>	<b>10</b>
<b>References .....</b>	<b>11</b>
<b>Appendices .....</b>	<b>11</b>

# Big yEAST Exploratory Market Research

## Introduction

One of the most favored consumables the world over is beer, and that is no different in the United States. According to [usatoday.com](http://usatoday.com), the average American consumes 26.2 gallons of beer per year. Consumption across the US varies however, and that presents dynamic challenges to many individuals and companies looking to break into or strengthen a strong foothold in the market.

### BUSINESS OBJECTIVE

Our fictitious brewing company “Big yEAST” has been a staple on the eastern seaboard for several decades. A recent evaluation of regional market trends has shown a decline in total consumption of beer, especially with the key demographic of (25-34-year-old individuals). Bill Barley, the President & CEO of Big yEAST, has commissioned a high-level exploratory market review of several locations in the western United States in the hope that Big yEAST can establish a position in either a developed or emerging market with the presence of a local brewery. The focus of this report will be Denver, CO. Colorado as a state ranks among the top 20 states in average consumption and makes the top 10 in terms of excessive consumption according to [usatoday.com](http://usatoday.com).

## Data Requirements

For such a high-level analysis we will be focused on two fields, basic socioeconomic data, and competition. Our socioeconomic dataset, will be mined from [denvermetrodata.org](http://denvermetrodata.org). This site contains minable 2010 census data for the Denver Metro area, and in several instances has been extrapolated forward close to present day. For this analysis, we will be using a projected 2017 dataset containing the following:

- Population 25-34 years-old
- Number of Low-Income Households (<\$60k/year annual income)
- Number of Middle-Income Households (between \$60k and \$125k)
- Number of High-Income Households (>\$125k/year annual income)

In addition to the socioeconomic dataset, we will need to acquire some baseline information on our competition relative to the subject geography and demography. This analysis will be using the [Places API](#) offered by [FOURSQUARE](#). We will submit a search query through to the Places API endpoint to produce a list of established breweries by latitude and longitude which we will then tie back into our neighborhood data.

Table 1: Socioeconomic Data

	Neighborhood	Population	Low Income	Medium Income	High Income	Latitude	Longitude
0	Alder-Three Sisters Park	381	544	614	452	39.740010	-104.992020
1	Allendale	1718	1672	1158	457	39.693065	-104.940542
2	Alta Vista and Memorial Park	1417	1393	1015	404	37.171692	-104.520280
3	Apel Bacher Park, Koch Sub and Coulehan Grange	467	600	359	65	39.740010	-104.992020
4	Applewood, Echo Hill, Rolling Hills and Meadows	433	456	549	531	39.740010	-104.992020

Table 2: Brewery Data

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude
0	Alder-Three Sisters Park	39.74001	-104.99202	Iron Hill Brewery Nation	39.741615	-104.997566
1	Apel Bacher Park, Koch Sub and Coulehan Grange	39.74001	-104.99202	Iron Hill Brewery Nation	39.741615	-104.997566
2	Applewood, Echo Hill, Rolling Hills and Meadows	39.74001	-104.99202	Iron Hill Brewery Nation	39.741615	-104.997566
3	Athmar Park	39.70396	-105.01039	Chain Reaction Brewery	39.699577	-105.001335
4	Auraria	39.74575	-105.00997	Breckenridge Brewery Mountain House	39.748078	-105.006897

## Methodology

The analysis begins with an examination of brewery data mined from the FOURSQUARE Place API to determine the number of breweries or brewery density within the designated radius of the search query. This will help to determine market penetration strengths versus the scoped geography.

Table 3: Brewery Count by Neighborhood

	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude
Neighborhood					
Alder-Three Sisters Park	1	1	1	1	1
Apel Bacher Park, Koch Sub and Coulehan Grange	1	1	1	1	1
Applewood, Echo Hill, Rolling Hills and Meadows	1	1	1	1	1
Athmar Park	1	1	1	1	1
Auraria	3	3	3	3	3
Baker	3	3	3	3	3
Bell Park	2	2	2	2	2

Once the relative density of our competition has been determined we pivot to evaluate competitor popularity to determine which breweries have a strong customer base already established that we may want to either avoid or utilize depending on what business path Bill Barely decides to take. We start our popularity analysis by “One-Hot Coding” each of the breweries. This process simply assigns a numerical value to our categorical variable “brewery name” so that mathematical operations may be applied. Upon completion of the “One-Hot Coding”, brewery frequency in terms of geospatial relevance will be compiled and displayed.

**Table 4: One-Hot Coding**

	Neighborhood	Birch Street Brewery	Black Shirt Brewery	Black Sky Brewery	Blue Moon Brewery at The Sandlot	Boggy Draw Brewery	Breckenridge Brewery Mountain House	Brewery Bar	Brewery Bar III	Briar Common	...	Strange Craft Beer Company	The Brewery	The Grateful Gnome Sandwich Shoppe & Brewery
0	Alder-Three Sisters Park	0	0	0	0	0	0	0	0	0	...	0	0	0
1	Apel Bacher Park, Koch Sub and Coulehan Grange	0	0	0	0	0	0	0	0	0	...	0	0	0
2	Applewood, Echo Hill, Rolling Hills and Meadows	0	0	0	0	0	0	0	0	0	...	0	0	0
3	Athmar Park	0	0	0	0	0	0	0	0	0	...	0	0	0
4	Auraria	0	0	0	0	0	1	0	0	0	...	0	0	0

**Table 5: Top Breweries by Neighborhood**

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Alder-Three Sisters Park	Iron Hill Brewery Nation	next day brewery	Hops Grill & Brewery	Hogshead Brewery	Heavenly Daze Brewery	Grossen Bart Brewery	Great Divide Brewing Co.	Gaijin 24886 Brewery	Freshwerks Brewery	Empourium Brewing Company
1	Apel Bacher Park, Koch Sub and Coulehan Grange	Iron Hill Brewery Nation	next day brewery	Hops Grill & Brewery	Hogshead Brewery	Heavenly Daze Brewery	Grossen Bart Brewery	Great Divide Brewing Co.	Gaijin 24886 Brewery	Freshwerks Brewery	Empourium Brewing Company
2	Applewood, Echo Hill, Rolling Hills and Meadows	Iron Hill Brewery Nation	next day brewery	Hops Grill & Brewery	Hogshead Brewery	Heavenly Daze Brewery	Grossen Bart Brewery	Great Divide Brewing Co.	Gaijin 24886 Brewery	Freshwerks Brewery	Empourium Brewing Company
3	Athmar Park	Chain Reaction Brewery	next day brewery	Counter Culture Brewery + Grille	Hops Grill & Brewery	Hogshead Brewery	Heavenly Daze Brewery	Grossen Bart Brewery	Great Divide Brewing Co.	Gaijin 24886 Brewery	Freshwerks Brewery
4	Auraria	Breckenridge Brewery Mountain House	Strange Craft Beer Company	Briar Common	next day brewery	Counter Culture Brewery + Grille	Hops Grill & Brewery	Hogshead Brewery	Heavenly Daze Brewery	Grossen Bart Brewery	Great Divide Brewing Co.

After exploring the brewery data mined from the FOURSQUARE Place API, we will need to go back to our socioeconomic data set. The goal is to use [K-Means](#) clustering to categorize our scoped sub geographies (neighborhoods), so that our stakeholder can make high-level judgment calls on the relative market viability of Denver, CO. First, we will need to normalize our socioeconomic data to a net mean of 0, using min-max scaling for the purpose of this analysis.

#### Equation 1: Min-Max Normalization

$$z = \frac{x - \min(x)}{\max(x) - \min(x)}$$

Table 6: Normalized Data

	Population	Low Income	Medium Income	High Income	Venue
0	0.044101	0.066254	0.127810	0.111138	0.0
1	0.055190	0.073334	0.074729	0.015982	0.0
2	0.050806	0.055127	0.114280	0.130563	0.0
3	0.232237	0.237704	0.199001	0.066142	0.0
4	0.000000	0.000000	0.000833	0.003688	0.5

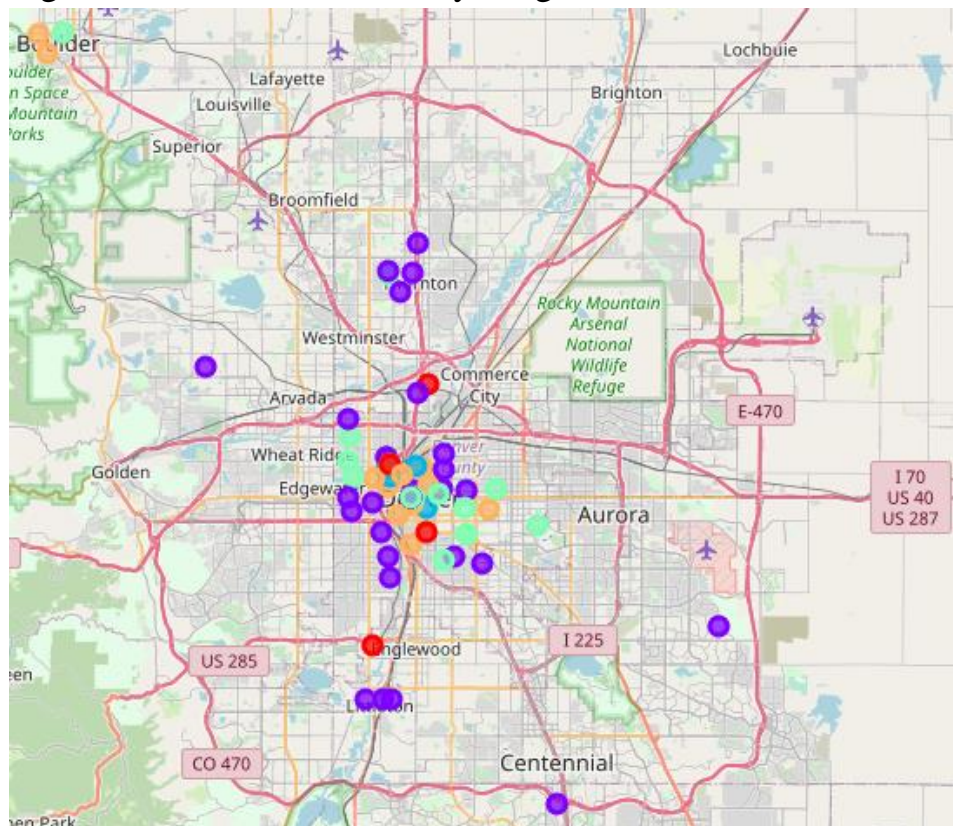
Now that our data has been normalized to a net mean of 0, we can begin to cluster our data using a supervised machine learning technique, K-Means clustering. K-Means is a popular clustering technique in data mining and machine learning that minimizes the within-cluster variance (squared Euclidian distance) ([https://en.wikipedia.org/wiki/K-means\\_clustering](https://en.wikipedia.org/wiki/K-means_clustering)).

## Results

The K-Means analysis produced 5 clusters using our socioeconomic dataset and the queried brewery data set from the FOURSQUARE Places API. The resulting operation may be viewed on the map below, and a discussion of the cluster features will follow.



**Figure 1: K-Means Clusters by Neighborhood**



Cluster #1, shown in red on the cluster map has a left skewed income distribution, but still a relatively high number of medium income households. This cluster represents our second highest concentration of 25-34-year-olds, and our second lowest concentration of existing breweries.

**Table 7: Cluster #1 Data**

Population	Low Income	Medium Income	High Income	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue
5182	6662	3881	1677	39.650391	-105.021878	0	Boggy Draw Brewery	next day brewery	Counter Culture Brewery + Grille	Hops Grill & Brewery	Hogshead Brewery	Heavenly Daze Brewery	Grossen Bart Brewery
3368	2260	4071	2589	39.740010	-104.992020	0	Iron Hill Brewery Nation	next day brewery	Hops Grill & Brewery	Hogshead Brewery	Heavenly Daze Brewery	Grossen Bart Brewery	Great Divide Brewing Co.
3437	2675	4300	2461	39.740010	-104.992020	0	Iron Hill Brewery Nation	next day brewery	Hops Grill & Brewery	Hogshead Brewery	Heavenly Daze Brewery	Grossen Bart Brewery	Great Divide Brewing Co.
2366	3029	3282	4067	39.760114	-105.009369	0	Prost Brewing	next day brewery	Counter Culture Brewery + Grille	Hops Grill & Brewery	Hogshead Brewery	Heavenly Daze Brewery	Grossen Bart Brewery
3710	2467	2409	2789	39.740010	-104.992020	0	Iron Hill Brewery Nation	next day brewery	Hops Grill & Brewery	Hogshead Brewery	Heavenly Daze Brewery	Grossen Bart Brewery	Great Divide Brewing Co.
3001	4813	2153	1549	39.760114	-105.009369	0	Prost Brewing	next day brewery	Counter Culture Brewery + Grille	Hops Grill & Brewery	Hogshead Brewery	Heavenly Daze Brewery	Grossen Bart Brewery

Cluster #2, shown in purple on the cluster map has a relatively normal income distribution, which indicates most households are categorized as medium income. This cluster represents the largest cluster category; however, each individual cluster has a relatively low population of 25-34-year-olds compared to the other clusters. This likely indicates more established communities. The most breweries by neighborhood are found in Cluster #2.

**Table 8: Cluster #2 Data**

Population	Low Income	Medium Income	High Income	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue
381	544	614	452	39.740010	-104.992020	1	Iron Hill Brewery Nation	next day brewery	Hops Grill & Brewery	Hogshead Brewery	Heavenly Daze Brewery	Grossen Bart Brewery
467	600	359	65	39.740010	-104.992020	1	Iron Hill Brewery Nation	next day brewery	Hops Grill & Brewery	Hogshead Brewery	Heavenly Daze Brewery	Grossen Bart Brewery
433	456	549	531	39.740010	-104.992020	1	Iron Hill Brewery Nation	next day brewery	Hops Grill & Brewery	Hogshead Brewery	Heavenly Daze Brewery	Grossen Bart Brewery
1840	1900	956	269	39.703960	-105.010390	1	Chain Reaction Brewery	next day brewery	Counter Culture Brewery + Grille	Hops Grill & Brewery	Hogshead Brewery	Heavenly Daze Brewery
653	881	818	971	39.875605	-104.991202	1	Hops Restaurant	Mangy Dog Home Brewery	Counter Culture Brewery + Grille	Hops Grill & Brewery	Hogshead Brewery	Heavenly Daze Brewery
621	550	737	524	39.816930	-105.028040	1	Jackass Hill Brewing Co.	Saint Patrick's Brewery	Black Shirt Brewery	Counter Culture Brewery + Grille	Hops Grill & Brewery	Hogshead Brewery

Cluster #3, shown in cyan on the cluster map has a strong left skewed income distribution, which indicates most households are categorized as low income or less than \$60k/year annual. This cluster represents the smallest cluster category; however, each individual cluster has the highest population of 25-34-year-olds compared to the other clusters. The fewest breweries are found in this cluster.

**Table 9: Cluster #3 Data**

Population	Low Income	Medium Income	High Income	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue
7794	7192	2960	946	39.733700	-104.979290	2	Counter Culture Brewery + Grille	Lowdown Brewery+Kitchen	next day brewery	Hops Grill & Brewery	Hogshead Brewery	Heavenly Daze Brewery
4015	1607	3034	1323	39.756989	-105.011058	2	Wyntkoop Brewing Co.	Breckenridge Brewery Mountain House	Briar Common	Prost Brewing	Crazy Mountain Brewery Taproom & Beer Garden	Hops Grill & Brewery
5246	7929	4804	1649	39.752550	-105.008770	2	Wyntkoop Brewing Co.	Breckenridge Brewery Mountain House	Briar Common	Prost Brewing	Crazy Mountain Brewery Taproom & Beer Garden	Hops Grill & Brewery
6694	3532	2387	2044	39.759200	-104.987600	2	Jagged Mountain Brewery	Blue Moon Brewery at The Sandlot	Liberati Restaurant & Brewery	Great Divide Brewing Co.	Spangalang Brewery	next day brewery



Cluster #4, shown in sky blue on the cluster map has a normal income distribution, which indicates most households are categorized as medium income. This cluster represents the median cluster category. Each individual cluster has the median population of 25-34-year-olds compared to the other clusters, and the second highest value in terms of brewery count.

**Table 10: Cluster #4 Data**

Population	Low Income	Medium Income	High Income	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue
1970	1753	1321	1264	39.776660	-105.039840	3	The Grateful Gnome Sandwich Shoppe & Brewery	Empourium Brewing Company	next day brewery	Cooper Kettle Brewery	Hops Grill & Brewery	Hogshead Brewery	Heavenly Daze Brewery
1127	1017	1219	3509	39.875294	-105.009828	3	Mangy Dog Home Brewery	next day brewery	Counter Culture Brewery + Grille	Hops Grill & Brewery	Hogshead Brewery	Heavenly Daze Brewery	Grossen Bart Brewery
1039	1557	800	1908	39.717820	-104.948830	3	Hops Grill & Brewery	next day brewery	Iron Hill Brewery Nation	Hogshead Brewery	Heavenly Daze Brewery	Grossen Bart Brewery	Great Divide Brewing Co.
1717	1804	1702	2193	39.717820	-104.948830	3	Hops Grill & Brewery	next day brewery	Iron Hill Brewery Nation	Hogshead Brewery	Heavenly Daze Brewery	Grossen Bart Brewery	Great Divide Brewing Co.
3982	3020	1407	1466	39.733590	-104.950010	3	Birch Street Brewery	Freshwerks Brewery	Iron Hill Brewery Nation	Hops Grill & Brewery	Hogshead Brewery	Heavenly Daze Brewery	Grossen Bart Brewery

Cluster #5, shown in peach on the cluster map has a normal income distribution, which indicates most households are categorized as medium income. Each individual cluster has a low total population of 25-34-year-olds compared to the other clusters, and the median value in terms of brewery count.

**Table 11: Cluster #5 Data**

Population	Low Income	Medium Income	High Income	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
39	20	4	15	39.745750	-105.009970	4	Breckenridge Brewery Mountain House	Strange Craft Beer Company	Briar Common	next day brewery	Counter Culture Brewery + Grille
2042	1379	987	553	39.711170	-104.992090	4	Heavenly Daze Brewery	Brewery Bar	next day brewery	Counter Culture Brewery + Grille	Hops Grill & Brewery
516	487	436	376	39.735260	-104.990580	4	Counter Culture Brewery + Grille	Lowdown Brewery+Kitchen	Iron Hill Brewery Nation	next day brewery	Hops Grill & Brewery
554	836	959	631	39.758354	-104.982106	4	Blue Moon Brewery at The Sandlot	Liberati Restaurant & Brewery	Great Divide Brewing Co.	Spangalang Brewery	next day brewery
1900	2240	1082	806	39.732710	-104.930420	4	next day brewery	Freshwerks Brewery	Birch Street Brewery	Wild Mountain Smokehouse & Brewery	Cooper Kettle Brewery

## Discussion

This analysis is a high-level exploratory market analysis of neighborhood in the Denver Metro Area of Colorado. One specific age group and three socioeconomic classes were identified by our stakeholder Bill Barely in our pre analysis meeting. Future analytics on this specific geography should include the following:

- Real Estate Valuation
  - This will be needed to minimize start-up cost if Denver is chosen
- Expanded Age Demographics
  - Data is available and should be reviewed to ensure all aspects are understood
- Evaluate Other Metro Areas in Colorado
  - Colorado Springs, CO
  - Boulder, CO (more in depth)
  - Pueblo, CO

Based on the results of the K-Means Analysis, Cluster #1 appears to be the most ideal fit to enter a new neighborhood with an established area market and reasonable income levels and population of our desired generation.

## Conclusion

It is the analysts' recommendation that Denver, CO is a viable market to open a new brewery location. We can clearly define an established market for Big yEAST's product portfolio, and we have a cursory review of the geographic and socioeconomic data. Big yEAST has the option of entering an underserved market with reasonable income levels such as Cluster #1 shown in red, or pivoting into a more establish neighborhood within the metro area to attempt to steal market share from what might be categorized as weaker competition such as clusters #2 and #4 in purple and sky blue respectively.

## References

- [Denvermetrodata.org](http://Denvermetrodata.org)
- [USAToday.com](http://USAToday.com)
- [FOURSQUARE](https://www.foursquare.com)
- [Thedenverchannel.com](http://Thedenverchannel.com)
- [Wikipedia](http://Wikipedia)

## Appendices

- List of Figures and Tables
  - Table 1: Socioeconomic Data, page 4
  - Table 2: Brewery Data, page 4
  - Table 3: Brewery Count by Neighborhood, page 4
  - Table 4: One-Hot Coding, page 5
  - Table 5: Top Breweries, page 5
  - Table 6: Normalized Data, page 6
  - Table 7: Cluster #1 Data, page 7
  - Table 8: Cluster #2 Data, page 8
  - Table 9: Cluster #3 Data, page 8
  - Table 10: Cluster #4 Data, page 9
  - Table 11: Cluster #5 Data, page 9
  - Figure 1: Map of K-Means Clusters, page 7
  - Equation 1: Min-Max Normalization, page 6