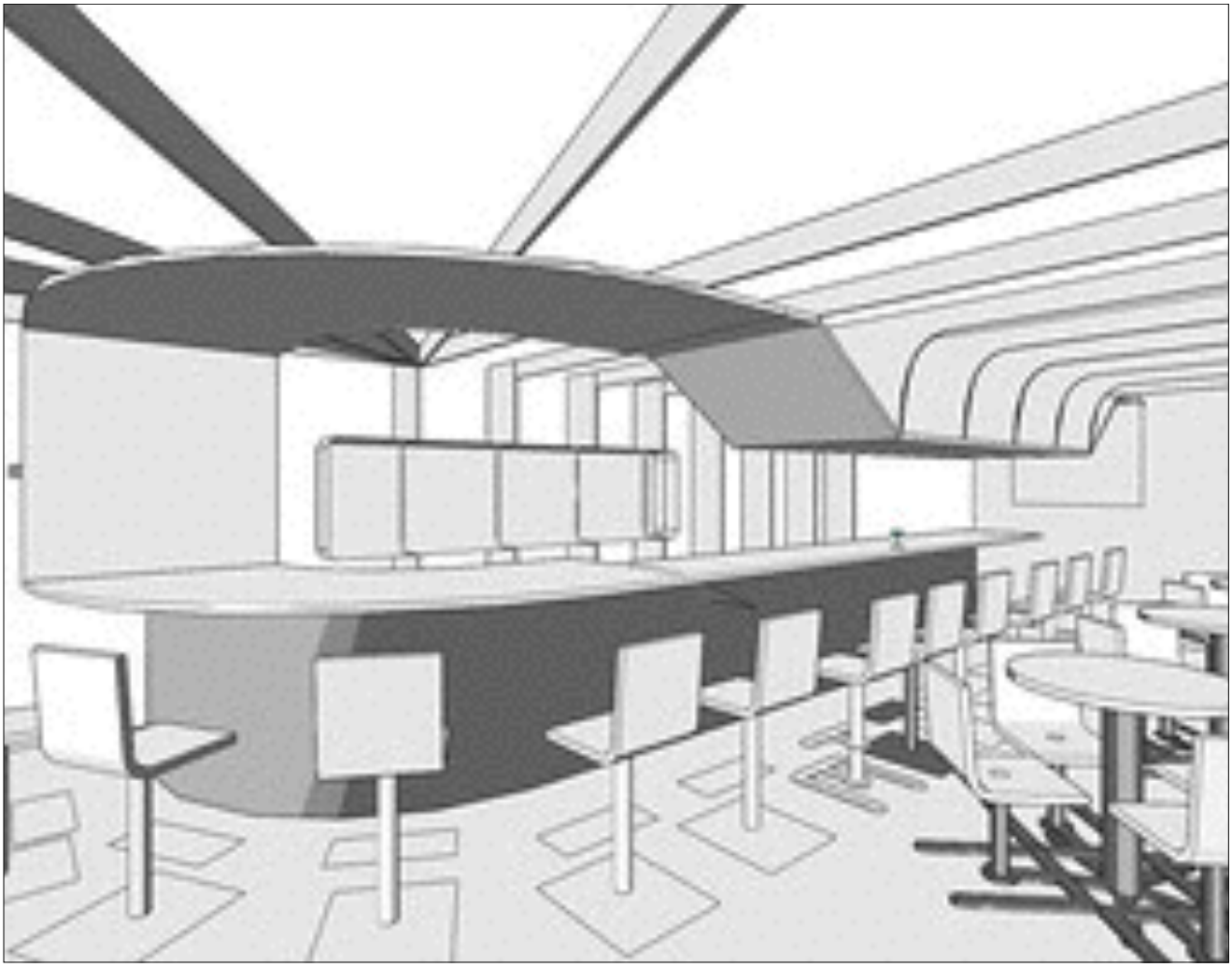
IBM DATA SCIENCE PROFESSIONAL CERTIFICATE: CAPSTONE



BATTLE OF THE NEIGHBORHOOD: Hennepin County, MN

Prepared for: IBM Data Science Capstone
Prepared by: Danielle Dumonceaux
April 12, 2020

1. Business Understanding

Innovation and strong business climate are in the DNA of the state of Minnesota. From being a pioneer in medical devices and bioscience technology, to a leader in water technology, food and agricultural innovation, Internet of Things, and new technologies.

In 2019, Fortune 500 named the Twin Cities (Hennepin County) in Minnesota number one with having the most Fortune 500 companies per Capita among the top 30 metro areas. Politico 2018 named Minnesota as the second Strongest State in the Nation, while the 2019 U.S. News & World Report ranked Minnesota as third Best State Overall.

The state of Minnesota welcomes cutting-edge ideas, from big, small, and debuting companies. While Minnesota is home to 17 Fortune 500 companies and six of the nation's largest private companies, the prosperous economy of the state would not be such without the burgeoning small and debuting businesses. Startup businesses add to the robust job market found in Minnesota, which helped Minnesota acquire the 2018 title of "Happiest State in the U.S." from MagnifyMoney, and the 2019 U.S. News & World Report ranking of number three for opportunity. Startup businesses have been successful in the state of Minnesota, helping the state obtain the ranking of third in Startup Early Survival Rate from Kauffman Indicators of Entrepreneurship for 2017, and an equal rating from the Bureau of labor Statistics from 2013-2018 as the best 5-Year Business Survival Rate.

Minnesota is an ever growing culturally diverse state whose residents enjoy indulging on familiar and innovative foods. Each year, the residents of Minnesota join together at the Great Minnesota Get Together, other known as the Minnesota State Fair, to experience new twists to their favorite dishes, and explore novel treats from around the gastronomic globe. However, outside of the August fair, the residents of Minnesota are avid restaurant goers, where each dish has a touch of modernization and innovation.

Interest

A multi-gastronomical world-renowned chef desires to bring his innovative culinary perspective to the Twin Cities — Hennepin County — to open up a type of restaurant that is already present, yet under represented in the area. They desire the restaurant to be in an urban area geared towards the middle-class budget— at least 64% of the households making $3,334 a month or more, therefore located in an area that is populated by said class -- between 21,000 and 90,000 residents -- to support the business, but also one that has a rich culinary scene.

Note

In this analysis, basic data analysis will be used to determine the most optimal city ('Neighborhood') in Hennepin County, MN that meets the guidelines of the interest, focusing on population, salary of households within the various neighborhoods, and the quantity and type of venues in the most optimal cities. There are various additional factors, such as space availability, unemployment, and difficulty of starting a new business that are important to explore before choosing a specific location, but that can be completed after the optimal city and type of cuisine are identified. Therefore, they will not be performed within the scope of this project.

2. Data Understanding

There are three types of data that will be used to resolve the problem at hand.

1. **Metropolitan Council** (https://metrocouncil.org/Data-and-Maps.aspx) : Used to acquire a list of the various neighborhoods in Hennepin County, their collective population, and their average salary. This information is to identify the various neighborhoods in play, to see if there is a boosting population that could support the new small business, and to identify which community would be identified as middle-class. Data is from the year 2017.
2. **FourSqure**: Used to acquire a comprehensive list of restaurant establishments currently in Hennepin County. A filtered list is used to identify which restaurant type is already present but least represented.
3. **Geospatial Data**: Used to integrate the restaurants to the neighborhoods by their longitude and latitude coordinates.

Data Preparation

Upon properly filtering the data to meet the needs of the interest, the main data frame was built as seen in Figure 2.1. Please note, the cities ('Neighborhoods') are labeled as CTU_NAME until later in the exploratory data analysis.

	CTU_CODE	CTU_NAME	POPULATION	HOUSEHOLDS	COMMUNITY_DESIGNATION	Workers earning equal or more than $3,334 per month	Workers earning between $1,251 and $3,333 per month	Workers earning equal or less than $1,250 per month
0	6616	Bloomington	88885	38022	Urban	23384	12127	7148
1	7948	Brooklyn Center	31145	11063	Urban	6312	5966	2779
2	14158	Crystal	22929	9488	Urban	6702	3646	2048
3	18188	Edina	52497	22557	Urban	14579	3939	3236
4	21965	Fort Snelling (Unorganized)	271	190	Urban	31	40	21
5	24308	Golden Valley	21646	9449	Urban	6687	2126	1467
6	30140	Hopkins	19079	8765	Urban	4979	2710	1531
7	43000	Minneapolis	423990	180340	Urban	100309	56845	33553
8	45628	New Hope	21545	8851	Urban	5458	3227	1838
9	49012	Osseo	2792	1285	Urban	693	376	219
10	54214	Richfield	36544	15179	Urban	9684	5673	3122
11	54808	Robbinsdale	14860	6281	Urban	4143	2192	1102
12	56680	St. Anthony (Hennepin Co. part)	5444	2266	Urban	1775	554	375

Figure 2.1 Main Data Frame

The initial data that was imported covered all the cities found within Hennepin County. This amounted to a total of 46 cities. For each city, household salary was divided into three categories as seen in Figure 2.2.

	CTU_CODE	CTU_NAME	WAGE_LEVEL	COUNT
3	6616	Bloomington	Number of Workers with earnings $1,250 per mon...	7148
4	6616	Bloomington	Number of Workers with earnings $1,251 to 3,3...	12127
5	6616	Bloomington	Number of Workers with earnings greater than $...	23384
9	7948	Brooklyn Center	Number of Workers with earnings $1,250 per mon...	2779
10	7948	Brooklyn Center	Number of Workers with earnings $1,251 to 3,3...	5966

Figure 2.2. Cities' Wage Level split into three monthly earning brackets

In addition, in the initial data set, community designations were identified as seen in Figure 2.3. This helped to clean the dataset, as the interest desired to open a business in an urban city. Therefore, those cities who did not fit this criteria were removed from the main data frame.

	CTU_CODE	CTU_NAME	POPULATION	HOUSEHOLDS	COMMUNITY_DESIGNATION
0	6616	Bloomington	88885	38022	Urban
1	7948	Brooklyn Center	31145	11063	Urban
2	7966	Brooklyn Park	80866	27660	Suburban
3	10846	Champlin	23690	8519	Suburban
4	10918	Chanhassen (Hennepin Co. part)	0	0	Emerging Suburban Edge
5	13168	Corcoran	5592	1957	Emerging Suburban Edge
6	14158	Crystal	22929	9488	Urban
7	15022	Dayton (Hennepin Co. part)	5427	1910	Emerging Suburban Edge
8	15148	Deephaven	3948	1438	Suburban
9	18116	Eden Prairie	63726	24893	Suburban

Figure 2.3 Community Designation's identified

3. Modeling

Once data preparation was complete, it was time to explore the data. In order to determine which city ('Neighborhood') was to be best suite for our interest, we needed to perform the following:

1. Apply basic exploratory analysis to our data and identify the percentage of workers' earnings in relation to the number of households per city. This would be done via three bar chart analysis.
2. Pull venue lists per neighborhood from the Foursquare API and convert the categorical values into a form that could be used for machine learning algorithms. This would be done using One Hot Encoding.
3. Identify the frequency of all venues within the identified neighborhoods by grouping rows by neighborhood and taking the mean of frequency of occurrence of each venue category.
4. Identify groups with similar characteristics, using K-Means.

3.1 Exploratory Data Analysis

In order to identify which neighborhood would be best suited for our interest, bar graphs were used to visualize a main filter: household salary. Figure 3.1.1 depicts the percentages of workers' earnings based off their neighborhood household totals. It shows that regardless of the population, amongst a majority of neighborhoods there was a neighborhood to income percentage of greater than 50% earning equal to or more than $3,334 per month. However, not all neighborhoods met the interest's 64% salary mark.

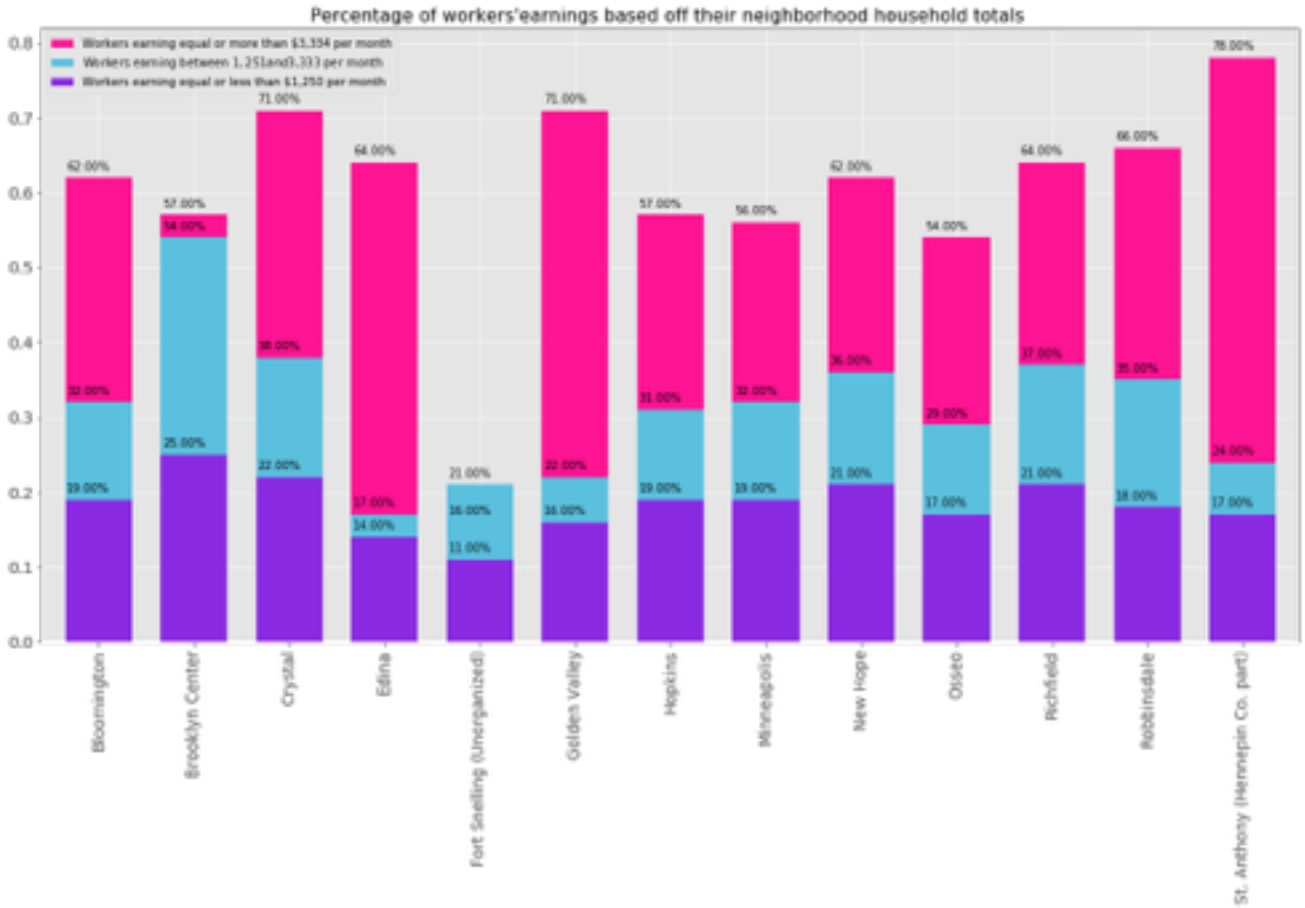


Figure 3.1.1 Percentage of workers' earnings based off their neighborhood household totals

While salary was important to note, it was also key to visualize how many homes populated each neighborhood. Figure 3.1.2 depicts the total households per neighborhood. From this illustration, we can see that the city ('Neighborhood') of Minneapolis has too great of households for the interest.

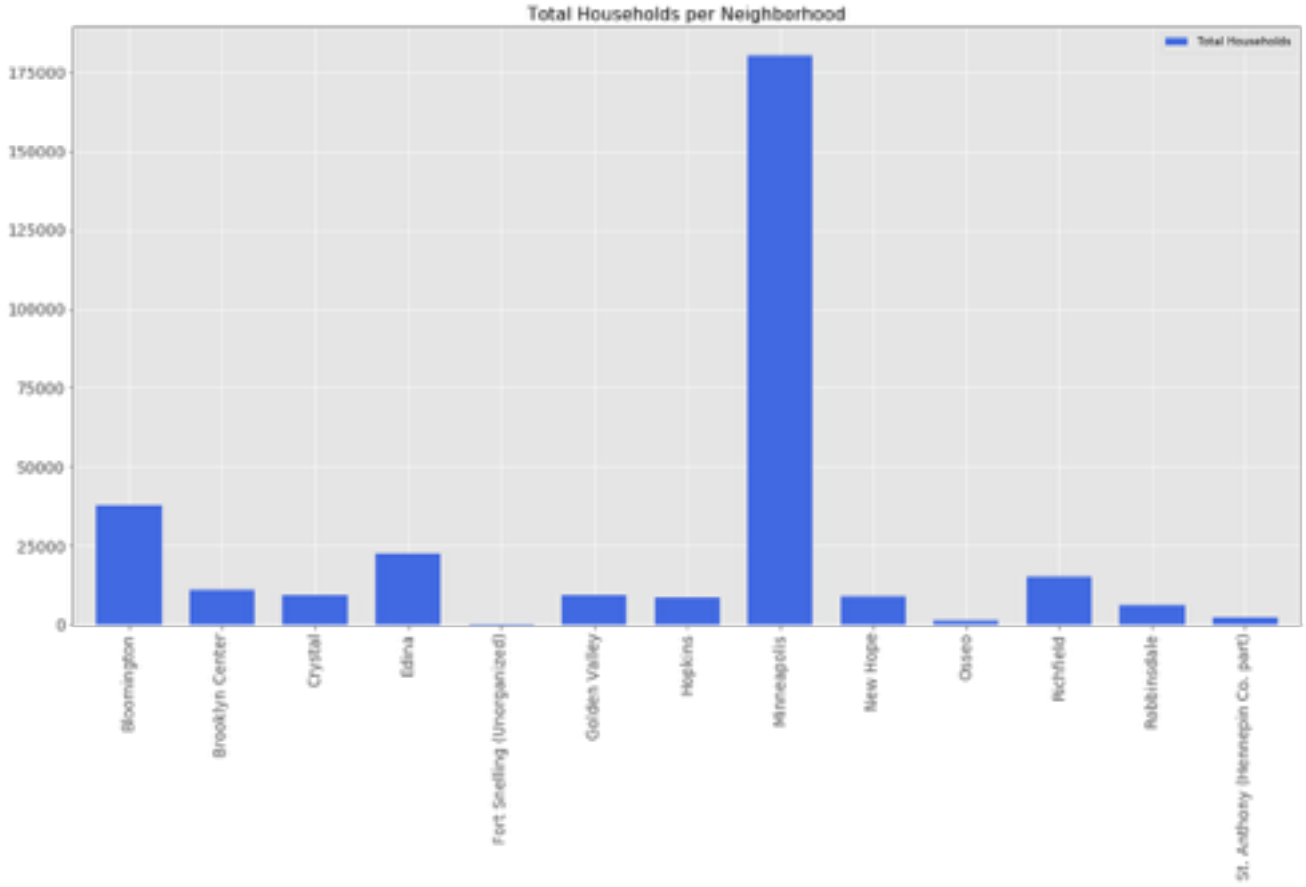


Figure 3.1.2 Total Households per Neighborhood

The final exploratory data analysis figure combined the first two figures to visualize the two key citizen driven location deciding factors: households numbers, and the amount of workers' in each neighborhood that would be considered middle class by earning equal or more than $3,334 per month. This can be seen in figure 3.1.3.

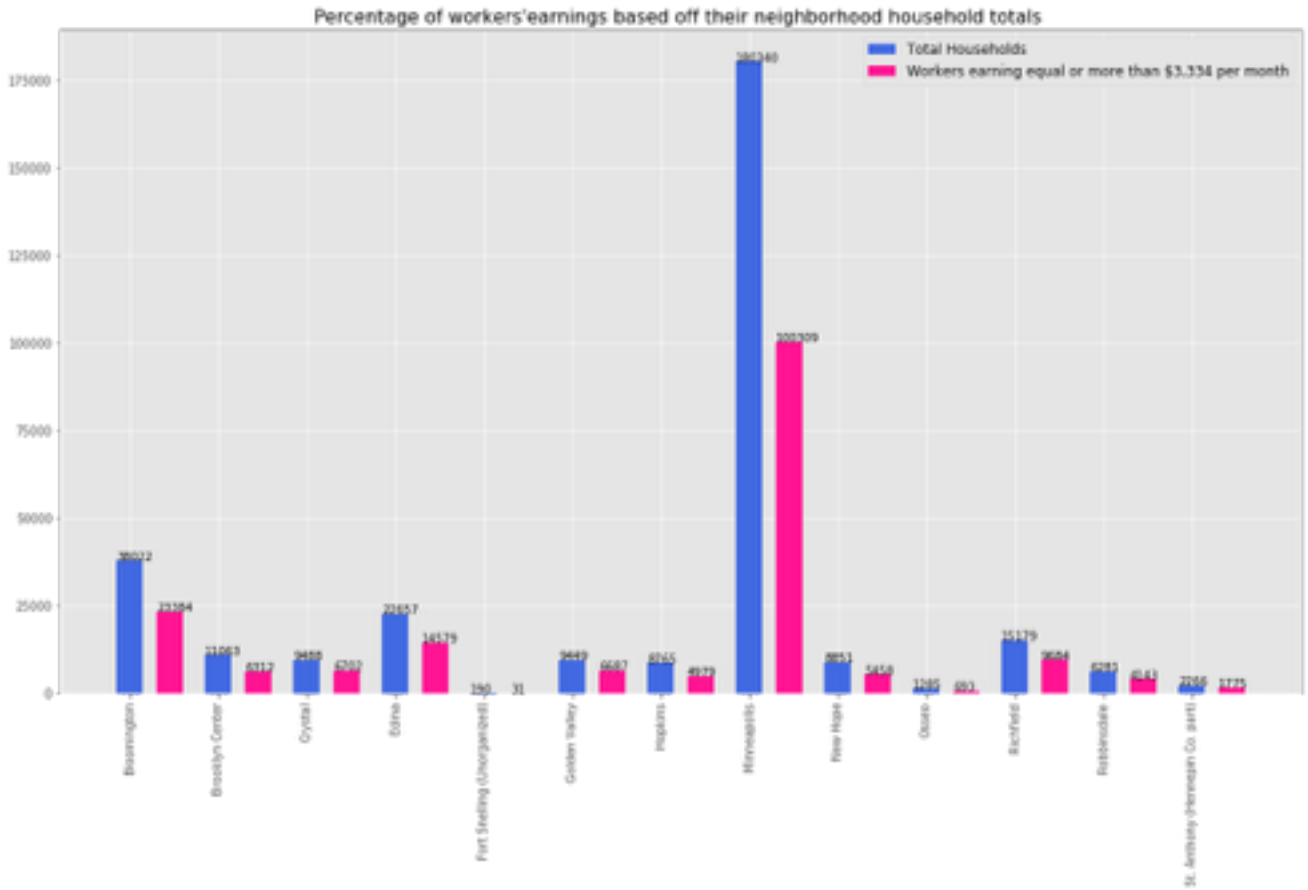


Figure 3.1.3 Percentage of workers' earnings based off their neighborhood household totals

Neighborhoods of interest are then narrowed to fit the interest. The population is to be between 21,000 and 90,000 residents and the middle class workers's population of households should be equal or greater than 64% of households per neighborhood. Figure 3.1.4 depicts the percentages of middle class worker households for all neighborhoods in question, followed by figure 3.1.5 that applies the interest's criteria to narrow the list to four neighborhoods.

	CTU_NAME	POPULATION	Percentage of Middle Class Workers
0	Bloomington	88885	62.0
1	Brooklyn Center	31145	57.0
2	Crystal	22929	71.0
3	Edina	52497	64.0
4	Fort Snelling (Unorganized)	271	16.0
5	Golden Valley	21646	71.0
6	Hopkins	19079	57.0
7	Minneapolis	423990	56.0
8	New Hope	21545	62.0
9	Osseo	2792	54.0
10	Richfield	36544	64.0
11	Robbinsdale	14860	66.0
12	St. Anthony (Hennepin Co. part)	5444	78.0

Figure 3.1.4 Percentage of Middle Class Worker Households per Neighborhood

	CTU_NAME	POPULATION	Percentage of Middle Class Workers
2	Crystal	22929	71.0
3	Edina	52497	64.0
5	Golden Valley	21646	71.0
10	Richfield	36544	64.0

Figure 3.1.5 Narrowed neighborhood list that meets interest's criteria

3.2 One Hot Encoding

Upon identifying four neighborhoods that fit within the perimeters of the interest, it is now time to identify what venues, their types, and frequency are present in each neighborhood. To do this, Foursquare is accessed. In order to access Foursquare, the geopy library was accessed to acquire the needed latitude and longitude coordinates of each neighborhood so to communicate with the Foursquare API. Figure 3.2.1 depict the results of accessing the geopy library for the four neighborhoods.

	index	Neighborhood	POPULATION	Percentage of Middle Class Workers	LONGITUDE	LATITUDE
0	2	Crystal	22929	71.0	-93.357536	45.032421
1	3	Edina	52497	64.0	-93.350122	44.889703
2	5	Golden Valley	21646	71.0	-93.378462	44.986118
3	10	Richfield	36544	64.0	-93.287788	44.876643

Figure 3.2.1 Latitude and Longitude coordinates acquired through the geopy library.

In accessing the folium library, a map of Hennepin County was able to be rendered. This permitted to visualize Hennepin County and superimpose the neighborhoods on top (Figure 3.2.2)

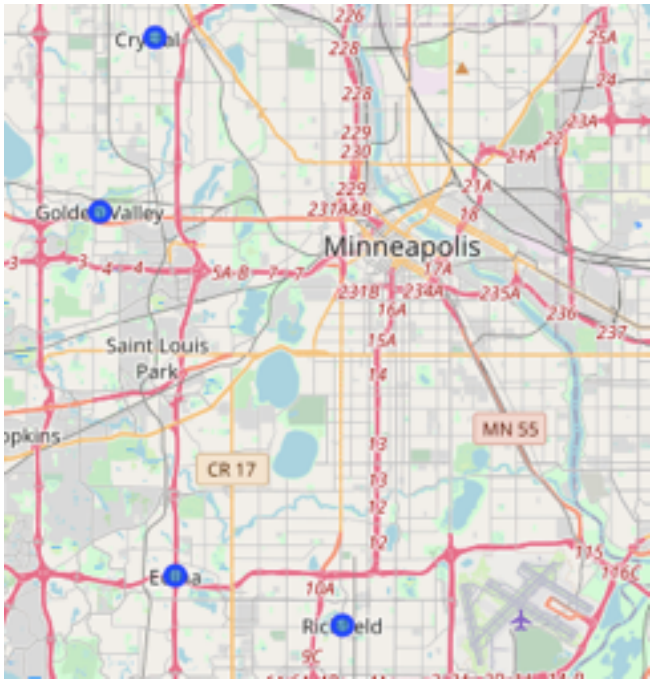


Figure 3.2.2 Hennepin County with neighborhoods superimposed.

Foursquare API was accessed to explore the neighborhoods and segment them. A call brought back forty-nine venues. These venues were then grouped by their neighborhood, resulting in thirty-six unique categories. To analyze each neighborhood, One Hot Encoding was used. This permitted to convert the categorical values, which were received through the Foursquare API call, into a form that could later be used to for machine learning algorithms. Figure 3.2.3 depict the final outcome of applying one hot encoding to the information received from accessing the Foursquare API.

Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Crystal	7	7	7	7	7	7
Edina	4	4	4	4	4	4
Golden Valley	33	33	33	33	33	33
Richfield	5	5	5	5	5	5

Figure 3.2.3 Results of calling the Foursquare API and applying One Hot Encoding

3.3 Mean of Frequency

Upon converting the categorical values with One Hot Encoding, Mean of Frequency can now be calculated. This will provide a smooth estimate of the center of derivative power. In other words, all venues within the identified neighborhoods are grouped into rows by neighborhood and an estimate of the central tendency of the derivative power distribution is taken to identify the frequency of occurrence of each venue category in each neighborhood. Figure 3.31 depicts the results after having found the mean of frequency.

	Neighborhood	ATM	American Restaurant	Arts & Crafts Store	Burger Joint	Chinese Restaurant	Coffee Shop	Dive Bar	Electronics Store	Fast Food Restaurant	...	Noodle House	Pizza Place	Pool	Record Shop	Salon / Barbershop	Sandwich Place	Shipping Store	Soccer Field
0	Crystal	0.000000	0.000000	0.142857	0.000000	0.142857	0.000000	0.000000	0.000000	0.0	...	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.0
1	Edina	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.0	...	0.000000	0.000000	0.25	0.000000	0.000000	0.000000	0.000000	0.0
2	Golden Valley	0.030303	0.060606	0.000000	0.030303	0.060606	0.060606	0.030303	0.030303	0.0	...	0.030303	0.060606	0.00	0.030303	0.030303	0.060606	0.030303	0.0
3	Richfield	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.200000	0.000000	0.2	...	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.4

Figure 3.3.1 Mean of Frequency of occurrences of each category

```
----Crystal----
                 venue  freq
0         Liquor Store  0.14
1  Arts & Crafts Store  0.14
2   Chinese Restaurant  0.14
3              Library  0.14
4   Italian Restaurant  0.14


----Edina----
                   venue  freq
0          Movie Theater  0.25
1           Home Service  0.25
2  Gym / Fitness Center  0.25
3                  Pool  0.25
4                   ATM  0.00


----Golden Valley----
                 venue  freq
0   Mexican Restaurant  0.06
1   Chinese Restaurant  0.06
2          Coffee Shop  0.06
3      Sandwich Place  0.06
4  American Restaurant  0.06


----Richfield----
                   venue  freq
0          Soccer Field   0.4
1              Dive Bar   0.2
2                Museum   0.2
3  Fast Food Restaurant   0.2
4                   ATM   0.0
```

For means of better visualization, the results of the mean of frequency are grouped to produce individual tables per neighborhood that represent the top five venues and their frequency in that neighborhood. This can be seen in figure 3.3.2.

Figure 3.3.2 Neighborhoods with their top 5 venues and their mean of frequency counts of each venue.

3.4 Clustering neighborhoods with K-Mean Clusters

In order to identify groups with similar characteristics, the unsupervised leaning method of K-Means was applied to the data. As the data had been condensed down to four cities, four clusters were identified (Figure 3.4.1). However, the desire to use K-Means was to identify in which city or region might there be the highest frequency of restaurants. Therefore, all neighborhoods within Hennepin County were brought back to see if it was possible to identify which region, closest to the remaining four neighborhoods, has the highest frequency of venues. Similar methods of one hot encoding (Figure 3.4.2), and finding the mean of frequency (Figure 3.4.3) were applied prior to clustering with K-Means (Figure 3.4.4). However, this proved to be unsuccessful, as all four cities were identified within the same cluster (Figure 3.4.5).

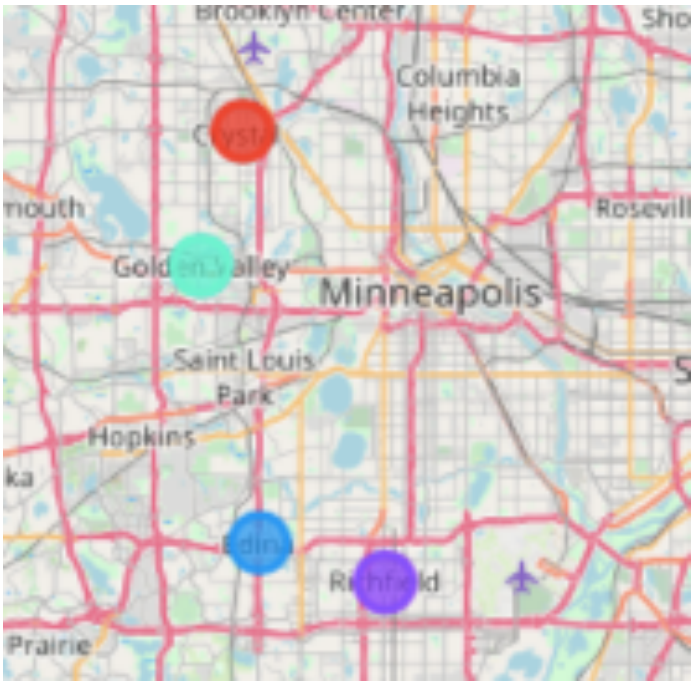


Figure 3.4.1 K-Means Cluster of Four Cities ('Neighborhoods')

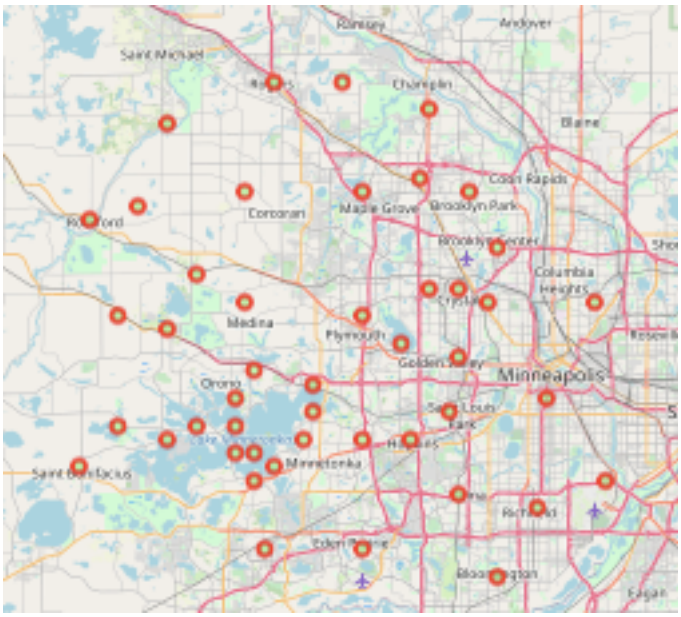
Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Bloomington	3	3	3	3	3	3
Brooklyn Center	4	4	4	4	4	4
Brooklyn Park	5	5	5	5	5	5
Champlin	8	8	8	8	8	8
Chanhassen (Hennepin Co. part)	3	3	3	3	3	3
Crystal	3	3	3	3	3	3
Deephaven	2	2	2	2	2	2
Eden Prairie	3	3	3	3	3	3
Edina	4	4	4	4	4	4
Excelsior	17	17	17	17	17	17
Fort Snelling (Unorganized)	4	4	4	4	4	4
Golden Valley	5	5	5	5	5	5
Hanover (Hennepin Co. part)	1	1	1	1	1	1
Hopkins	7	7	7	7	7	7
Maple Plain	5	5	5	5	5	5
Medicine Lake	1	1	1	1	1	1
Minneapolis	14	14	14	14	14	14
Minnetonka	1	1	1	1	1	1
Minnetonka Beach	4	4	4	4	4	4
Minnetrista	1	1	1	1	1	1
Mound	1	1	1	1	1	1
New Hope	1	1	1	1	1	1
Osseo	12	12	12	12	12	12
Plymouth	5	5	5	5	5	5
Richfield	14	14	14	14	14	14
Robbinsdale	15	15	15	15	15	15
Rockford (Hennepin Co. part)	5	5	5	5	5	5
Rogers	18	18	18	18	18	18
Spring Park	1	1	1	1	1	1
St. Anthony (Hennepin Co. part)	7	7	7	7	7	7
St. Bonifacius	1	1	1	1	1	1
St. Louis Park	8	8	8	8	8	8
Tonka Bay	3	3	3	3	3	3
Wayzata	32	32	32	32	32	32
Woodland	1	1	1	1	1	1

Figure 3.4.2 One Hot Encoding applied to all Hennepin County Neighborhoods

	Neighborhood	ATM	American Restaurant	Antique Shop	Art Gallery	Art Museum	Arts & Entertainment	Asian Restaurant	Athletics & Sports
0	Bloomington	0.000000	0.000000	0.000000	0.000000	0.000000	0.00000	0.000000	0.0
1	Brooklyn Center	0.000000	0.250000	0.000000	0.000000	0.000000	0.00000	0.000000	0.0
2	Brooklyn Park	0.000000	0.000000	0.000000	0.000000	0.000000	0.00000	0.000000	0.0
3	Champlin	0.000000	0.125000	0.000000	0.000000	0.000000	0.00000	0.000000	0.0
4	Chanhassen (Hennepin Co. part)	0.000000	0.000000	0.000000	0.000000	0.000000	0.00000	0.000000	0.0
5	Crystal	0.000000	0.000000	0.000000	0.000000	0.000000	0.00000	0.000000	0.0
6	Deephaven	0.000000	0.000000	0.000000	0.000000	0.000000	0.00000	0.000000	0.0
7	Eden Prairie	0.000000	0.000000	0.000000	0.000000	0.000000	0.00000	0.000000	0.0
8	Edina	0.000000	0.000000	0.000000	0.000000	0.000000	0.00000	0.000000	0.0
9	Excelsior	0.000000	0.058824	0.000000	0.000000	0.000000	0.00000	0.000000	0.0
10	Fort Snelling (Unorganized)	0.000000	0.000000	0.000000	0.000000	0.000000	0.00000	0.000000	0.0
11	Golden Valley	0.000000	0.000000	0.000000	0.000000	0.000000	0.00000	0.000000	0.2
12	Hanover (Hennepin Co. part)	0.000000	0.000000	0.000000	0.000000	0.000000	0.00000	0.000000	0.0
13	Hopkins	0.000000	0.000000	0.000000	0.000000	0.000000	0.00000	0.000000	0.0
14	Maple Plain	0.000000	0.000000	0.000000	0.000000	0.000000	0.00000	0.000000	0.0
15	Medicine Lake	0.000000	0.000000	0.000000	0.000000	0.000000	0.00000	0.000000	0.0
16	Minneapolis	0.000000	0.000000	0.000000	0.071429	0.071429	0.00000	0.000000	0.0
17	Minnetonka	0.000000	0.000000	0.000000	0.000000	0.000000	0.00000	0.000000	0.0
18	Minnetonka Beach	0.000000	0.000000	0.000000	0.000000	0.000000	0.00000	0.000000	0.0
19	Minnetrista	0.000000	0.000000	0.000000	0.000000	0.000000	0.00000	0.000000	0.0
20	Mound	0.000000	0.000000	0.000000	0.000000	0.000000	0.00000	0.000000	0.0
21	New Hope	0.000000	0.000000	0.000000	0.000000	0.000000	0.00000	0.000000	0.0
22	Osseo	0.000000	0.000000	0.000000	0.000000	0.000000	0.00000	0.000000	0.0
23	Plymouth	0.000000	0.000000	0.000000	0.000000	0.000000	0.00000	0.000000	0.0
24	Richfield	0.000000	0.000000	0.000000	0.000000	0.000000	0.00000	0.000000	0.0
25	Robbinsdale	0.066667	0.000000	0.000000	0.000000	0.000000	0.00000	0.000000	0.0
26	Rockford (Hennepin Co. part)	0.000000	0.000000	0.000000	0.000000	0.000000	0.00000	0.000000	0.0
27	Rogers	0.000000	0.000000	0.055556	0.000000	0.000000	0.00000	0.055556	0.0
28	Spring Park	0.000000	0.000000	0.000000	0.000000	0.000000	0.00000	0.000000	0.0
29	St. Anthony (Hennepin Co. part)	0.000000	0.000000	0.000000	0.000000	0.000000	0.00000	0.000000	0.0
30	St. Bonifacius	0.000000	0.000000	0.000000	0.000000	0.000000	0.00000	0.000000	0.0
31	St. Louis Park	0.000000	0.000000	0.000000	0.000000	0.000000	0.00000	0.000000	0.0
32	Tonka Bay	0.000000	0.000000	0.000000	0.000000	0.000000	0.00000	0.000000	0.0
33	Wayzata	0.000000	0.125000	0.000000	0.000000	0.000000	0.03125	0.000000	0.0
34	Woodland	0.000000	0.000000	0.000000	0.000000	0.000000	0.00000	0.000000	0.0

Figure 3.4.3 Means of Frequency applied to all Hennepin County Neighborhoods

Figure 3.4.4 K-Means clusters applied to all Hennepin County Neighborhoods.



	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
0	Bloomington	Skating Rink	Home Service	Trail	Yoga Studio	College Rec Center	Comedy Club	Concert Hall	Construction & Landscaping
1	Brooklyn Center	American Restaurant	Park	Discount Store	Shopping Mall	Yoga Studio	Fried Chicken Joint	Comedy Club	Concert Hall
2	Brooklyn Park	College Rec Center	Hockey Arena	Gym / Fitness Center	Park	Bank	Discount Store	Food Truck	Fast Food Restaurant
3	Champlin	Coffee Shop	Video Store	Mobile Phone Shop	Fast Food Restaurant	Thrift / Vintage Store	Big Box Store	Gym	American Restaurant
6	Crystal	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
8	Deephaven	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
9	Eden Prairie	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
10	Edina	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
11	Excelsior	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
12	Fort Snelling (Unorganized)	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
13	Golden Valley	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
25	Minneapolis	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
27	Minnetonka Beach	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
32	Osseo	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
33	Plymouth	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
34	Richfield	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
35	Robbinsdale	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
36	Rockford (Hennepin Co. part)	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
37	Rogers	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
40	St. Anthony (Hennepin Co. part)	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
42	St. Louis Park	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
43	Tonka Bay	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
44	Wayzata	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
46	Woodland	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

Figure 3.4.5 Cluster zero of all Hennepin County Neighborhoods with K-Means. Here, all four neighborhoods of interest are in one cluster.

4. Results and Discussion

As a result of preforming basic exploratory and cluster analysis on the neighborhoods of Hennepin County, the most optimal city ('Neighborhood') for the interest was identified. Based on meeting the criteria in population — between 21,000 and 90,000 residents, household wages — at or above $3,334 per month, and restaurant venue frequency, it can be concluded that the city that would be of best fit for the interest would be Golden Valley, Minnesota (Figure 4.1). While the city of Crystal, MN was a close second in regard to population and percentage of middle class households, the difference in frequency of venues was substantial enough to eliminate this option . The other two cities, Edina and Richfield, were substantially lower in both the percentage of households that were middle class workers and the frequency of restaurant venues in their neighborhoods. Therefore, these two cities were eliminated, presenting Golden Valley as the optimal city for the interest.

	index	Neighborhood	POPULATION	Percentage of Middle Class Workers	Total Restaurant Venues
0	2	Crystal	22929	71.0	0.285714
1	3	Edina	52497	64.0	0.000000
2	5	Golden Valley	21646	71.0	0.515152
3	10	Richfield	36544	64.0	0.200000

Figure 4.1 Comparison of all components of the interest's criteria to determine the most optimal city.

Upon identifying the most optimal city, we then could use the Foursquare API data to determine which type of restaurant in Golden Valley was least represented. In Figure 4.2, all types of cuisines and their mean of frequency are displayed, making it visually easy to identify the least represented, but already present, type of cuisine. It can be seen that American Restaurant, Chinese Restaurant, Coffee Shop, Mexican Restaurant, Pizza Place, and Sandwich Shop all tied for the most frequent type of venue. However, the least represented ended in a five-way tie consisting of the Burger Joint, Gastropub, the Ice Cream Shop, the Italian Restaurant, and the Noodle House.

	Neighborhood	American Restaurant	Burger Joint	Chinese Restaurant	\
2	Golden Valley	0.060606	0.030303	0.060606	

	Coffee Shop	Fast Food Restaurant	Gastropub	Ice Cream Shop	\
2	0.060606	0.0	0.030303	0.030303	

	Italian Restaurant	Mexican Restaurant	Noodle House	Pizza Place	\
2	0.030303	0.060606	0.030303	0.060606	

	Sandwich Place	Total Restaurant Venues
2	0.060606	0.515152

Figure 4.2 Types of restaurant venues/frequency in Golden Valley, MN

While enough information was retrieved from Foursquare to make this initial restaurant type recommendation, the next steps would be to narrow the options for types of cuisine. This would require gathering additional information and venues that are not found on the Foursquare API, but rather available in other databases such as Google maps or Yelp. Having a comprehensive list of venues, stemming from a multitude of sources, would help to identify the actual types and frequency of restaurants in and around the city ('Neighborhood') of Golden Valley, Minnesota, so that a more definite restaurant type recommendation could be made.

5. Conclusion

After examining Hennepin County's population, wages, and venue frequency data through basic exploratory analysis, one hot encoding, and mean of frequency, we can conclude that the city that would be of best fit for the interest would be Golden Valley, Minnesota. Given the Foursquare data, their desire to open a cuisine that is least represented in that city would be their choice amongst the five least represented--Burger Joint, Gastropub, Ice Cream Shop, Italian Restaurant, or Noodle House. The utilization of K-Means to cluster venues in Hennepin County was unsuccessful for the four cities that met the city criteria for the interest fell into the same cluster. The next steps to narrow the cuisine choice would be to gather additional venues that are not on Foursquare, but found in other databases to get a more comprehensive understanding of the actual restaurants and types in and around the city of Golden Valley, MN.