

Capstone Project Final Project – Battle of Neighborhoods

Final Report

1. BACKGROUND

Relocation process is often stressful and fraught with uncertainty. Regardless of it being an inter-state relocation due to new job or finding a new apartment in the same city, the relocation process demands a good amount of time to research and shortlist neighborhoods based on individual requirements. Shortlisting an apartment includes considering key factors like neighborhood safety, rent affordability and accessibility around a neighborhood.

1.1 MOTIVATION

This project is inspired by the challenges me and my wife faced in shortlisting the apartments while relocating to Michigan due to my new job. It took a good bit of time for us to shortlist a few apartments and we found that the overall process of filtering out the neighborhoods was a bit tedious and time consuming. So, I decided to apply the data science principles in streamlining the relocation process which can potentially save consumer time and efforts.

2. BUSINESS PROPOSAL

The principles of data science can be applied to understand customer preferences about the frequently visited places over the past years and analyzing their historical data to suggest the best neighborhoods to live when relocating and further provide customers with a list of apartments to consider for a potential virtual or an in-person tour.

2.1 Problem Description

The problem that will be addressed through this capstone project focuses on streamlining the relocation process based on customer preferences on places they visit the most through a small survey. Every one of us go out and eat at a restaurant, bar, café or even at a food truck. Grocery picking, supermarket visits or shopping at a clothing store – everyone of us has some sort of preferences, taste and spending style which is a core and essential part of our daily life. Some might shop a lot and would like to stay closer to a neighborhood that has good variety of shopping stores while others might love spending time at a bar or a café and would prefer neighborhood with bar/cafes as their primary choice.

Data science tools and techniques can be leveraged to shortlist best locations based on customer's top preferences. Furthermore, the customers can provide their work location, commute time requirements and rent budget to get top list of apartments which can be considered for a potential tour.

3. UNDERSTANDING DATA

4.1 Data Collection

a) **Customer Survey** - I did a survey with my wife to understand the places we have visited the most in the past couple years which includes everything ranging from grocery picking to the type of cuisine we like the most. The data is stored in a csv file and imported into pandas dataframe.

b) **Python Package** - The geopy python client is used to retrieve latitude and longitude positions of different neighborhoods and places. This would be needed to further make queries to the foursquare API.

c) **Application Programming Interface (API)** - The Foursquare API is used to retrieve information about the places and venues - category of places, in particular. The Realtor API will be used to retrieve information about apartments like rent, type of apartment, address etc.

4. DATA METHODOLOGIES

4.1 Applying Data for Problem Solving

a) **Analyzing Survey Data** - Based on the survey, the most visited places by the customer is stored in a pandas Dataframe. We can see that other than groceries and supermarket visits, the customer frequently visits Arts & Crafts store and has a variety of preferred food choices like Mexican, Mediterranean and Pizza places. Each customer will have their own list of frequently visited places.

Index	User Category
0	Supermarket
1	Grocery Store
2	Fuel
3	Noodle House
4	Mexican Restaurant
5	Pharmacy
6	Arts & Crafts Store
7	Big Box Store
8	Mediterranean Restaurant
9	Shipping Store
10	Pizza Place
11	Taco Place
12	Convenience Store
13	Hardware Store

b) List of Neighborhoods around User's Work Location - I have decided to use my work address which is located in Auburn Hills, MI. The commute time requirements can be factored in to get the list of neighborhoods. For the sake of simplicity, all the neighborhoods within 11 mile of radius of this location is retrieved along with their latitudes and longitude coordinates. These are manually fetched due to timeout limitations with Geocoder python package. It can be seen that around 70 neighborhoods exist near Auburn Hills.

Index	Zip Code	Neighborhood	State	Latitude	Longitude
0	48070	Hutington Woods	MI	42.5	-83.2
1	48067	Royal Oak L1	MI	42.5	-83.1
2	48034	Southfield L1	MI	42.5	-83.3
3	48076	Southfield L2	MI	42.5	-83.2
4	48072	Berkley	MI	42.5	-83.2
5	48068	Royal Oak L2	MI	42.5	-83.1
6	48334	Farmington	MI	42.5	-83.4
7	48071	Madison Heights	MI	42.5	-83.1
8	48073	Royal Oak L3	MI	42.5	-83.2
9	48025	Franklin	MI	42.5	-83.3
10	48017	Clawson	MI	42.5	-83.1
11	48322	West Bloomfield L1	MI	42.5	-83.4
12	48012	Birmingham L1	MI	42.5	-83.2
13	48301	Bloomfield Hills L1	MI	42.5	-83.3
14	48009	Birmingham L2	MI	42.5	-83.2
15	48083	Troy L1	MI	42.5	-83.1
16	48312	Sterling Heights L1	MI	42.6	-83
17	48324	West Bloomfield L2	MI	42.6	-83.4
18	48084	Troy L2	MI	42.6	-83.2
19	48301	Bloomfield township	MI	42.5	-83.3
20	48303	Bloomfield Hills L2	MI	42.6	-83.2

•

•

•

•

58	48094	Washington L1	MI	42.7	-83
59	48350	Davisburg	MI	42.8	-83.5
60	48360	Lake Orion L2	MI	42.8	-83.3
61	48348	Clarkston L3	MI	42.8	-83.4
62	48363	Oakland	MI	42.8	-83.2
63	48095	Washington L2	MI	42.8	-83
64	48361	Lake Orion L3	MI	42.8	83.2
65	48362	Lake Orion L4	MI	42.8	-83.2
66	48366	Lakeville	MI	42.8	-83.1
67	48367	Leonard	MI	42.8	-83.1
68	48371	Oxford L1	MI	42.8	-83.3
69	48370	Oxford L2	MI	42.8	-83.2

c) Unique Outlets for Top User Categories Across Neighborhoods - The foursquare is queried for all the outlets within 3.5 miles radius of a neighborhood. The outlets whose categories matches the shortlisted top user categories is stored in a data frame. It can be seen that Huntington Woods has just 1 grocery store as compared to 3 grocery stores in Lake Orion L2.



Index	Neighborhood	Outlets	Categories
0	Hutington Woods	Kroger Grocery Pickup and Delivery	Grocery Store
1	Hutington Woods	Noodles & Company	Noodle House
2	Hutington Woods	Noodle Topia	Noodle House
3	Hutington Woods	Margaritas Mexican Restaurant	Mexican Restaurant
4	Hutington Woods	Camelia's Mexican Grill	Mexican Restaurant
5	Hutington Woods	Margarita's Authentic Mexican Cuisine	Mexican Restaurant
6	Hutington Woods	Qdoba Mexican Grill	Mexican Restaurant
7	Hutington Woods	CVS pharmacy	Pharmacy
8	Hutington Woods	Pharmacy - Ascension Michigan Pharmacy - Southfield	Pharmacy
9	Hutington Woods	Meijer Pharmacy	Pharmacy
10	Hutington Woods	Ascension Michigan Pharmacy Long Term Care	Pharmacy
11	Hutington Woods	Ascension Michigan Pharmacy at Nine Mile	Pharmacy
12	Hutington Woods	ZMC Pharmacy	Pharmacy
13	Hutington Woods	Hope Pharmacy	Pharmacy

15821	Lake Orion L2	Meijer	Supermarket
15822	Lake Orion L2	Kroger Grocery Pickup and Delivery	Grocery Store
15823	Lake Orion L2	Walmart Grocery Pickup and Delivery	Grocery Store
15824	Lake Orion L2	Vg's Grocery	Grocery Store
15825	Lake Orion L2	Noodles & Company	Noodle House
15826	Lake Orion L2	Señoritas Mexican Restaurant	Mexican Restaurant
15827	Lake Orion L2	Mexican Village	Mexican Restaurant
15828	Lake Orion L2	Azteca's Mexican Cuisine	Mexican Restaurant
15829	Lake Orion L2	Chipotle Mexican Grill	Mexican Restaurant
15830	Lake Orion L2	Restaurante Don Chano's	Mexican Restaurant
15831	Lake Orion L2	Qdoba Mexican Grill	Mexican Restaurant
15832	Lake Orion L2	CVS pharmacy	Pharmacy
15833	Lake Orion L2	Walmart Pharmacy	Pharmacy
15834	Lake Orion L2	Henry Ford Pharmacy - Lakeside	Pharmacy
15835	Lake Orion L2	Henry Ford Pharmacy - Henry Ford Macomb Hospital	Pharmacy

d) One Hot Encoding and Grouping Dataframe - A dataframe is created using pandas 'one hot encoding' for the customer's top visited places. The dataframe is grouped by neighborhoods and total count of each category across neighborhoods is calculated.

```
df_onehot = pd.get_dummies(df_mod[['Categories']], prefix="",
prefix_sep="")
df_onehot['Neighborhood'] = df_mod['Neighborhood']
fixed_columns = [df_onehot.columns[-1]]
list(df_onehot.columns[:-1])
df_onehot = df_onehot[fixed_columns]

df_grouped = df_onehot.groupby('Neighborhood').sum().reset_index()
```

Neighborhood	Arts & Crafts Store	Convenience Store	Grocery Store	Hardware Store	Mediterranean Restaurant	Mexican Restaurant	Noodle House	Pharmacy	Pizza Place	Shipping Store	Supermarket	Taco Place
Auburn Hills L1	3	23	7	19	9	20	1	34	109	3	1	3
Auburn Hills L2	3	24	7	20	9	20	1	39	123	3	1	3
Berkley	4	9	3	13	11	10	3	48	57	2	0	4
Birmingham L1	6	13	7	24	16	20	5	71	99	3	0	4
Birmingham L2	6	13	7	24	17	20	5	72	102	3	0	4
Bloomfield Hills L1	6	13	7	24	17	20	5	72	101	3	0	4
Bloomfield Hills L2	6	19	8	29	21	26	5	85	134	3	0	4
Bloomfield Hills L3	6	22	8	29	21	26	5	86	142	3	0	4
Bloomfield Hills L4	6	22	8	29	21	26	5	86	143	3	0	4
Bloomfield township	6	16	8	28	21	25	5	82	133	3	0	4
Clarkston L1	3	25	7	20	10	22	1	40	133	3	1	3
Clarkston L2	3	26	7	21	10	27	1	41	138	3	1	3
Clarkston L3	3	26	7	22	10	29	1	43	143	3	1	3
Clawson	5	12	7	19	14	17	5	67	90	3	0	4
Commerce Township	6	25	9	29	21	27	5	88	147	3	0	6
Davisburg	3	26	7	22	10	29	1	43	142	3	1	3
Drayton Plains	3	24	7	20	9	20	1	37	121	3	1	3
Farmington	5	11	4	16	13	12	3	57	63	2	0	4
Franklin	5	11	5	18	14	14	4	67	85	2	0	4
Huntington Woods	2	2	1	10	6	4	2	20	29	2	0	4
Keego Harbor	6	33	11	35	24	32	5	97	179	4	1	7
Lake Orion L1	3	26	7	21	10	27	1	41	138	3	1	3
Lake Orion L2	3	26	7	22	10	29	1	43	143	3	1	3
Lake Orion L3	3	27	7	22	10	30	1	43	145	3	1	3
Lake Orion L4	3	27	7	24	10	30	1	45	148	3	1	3
Lakeville	3	27	7	24	10	30	1	45	148	3	1	3

e) Applying Clustering Algorithm - k-means clustering algorithm is selected to cluster the neighborhoods based on the total count of each category across neighborhoods. The total number of clusters is selected as 5 and the corresponding cluster labels are added to the dataframe.

```
df_grouped = df_onehot.groupby('Neighborhood').sum().reset_index()
X = df_grouped.values[:,1:]
X = np.nan_to_num(X)
cluster_group = StandardScaler().fit_transform(X)
```

```

k_cluster = 5
cluster_group = df_grouped.drop('Neighborhood',1)

kmeans = KMeans(n_clusters=k_cluster,
random_state=0).fit(cluster_group)

```

Cluster Labels	Neighborhood	Arts & Crafts Store	Convenience Store	Grocery Store	Hardware Store	Mediterranean Restaurant	Mexican Restaurant	Noodle House	Pharmacy	Pizza Place	Shipping Store	Supermarket	Taco Place
0	Auburn Hills L1	3	23	7	19	9	20	1	34	109	3	1	3
0	Auburn Hills L2	3	24	7	20	9	20	1	39	123	3	1	3
3	Berkley	4	9	3	13	11	10	3	48	57	2	0	4
4	Birmingham L1	6	13	7	24	16	20	5	71	99	3	0	4
4	Birmingham L2	6	13	7	24	17	20	5	72	102	3	0	4
4	Bloomfield Hills L1	6	13	7	24	17	20	5	72	101	3	0	4
1	Bloomfield Hills L2	6	19	8	29	21	26	5	85	134	3	0	4
1	Bloomfield Hills L3	6	22	8	29	21	26	5	86	142	3	0	4
1	Bloomfield Hills L4	6	22	8	29	21	26	5	86	143	3	0	4
1	Bloomfield township	6	16	8	28	21	25	5	82	133	3	0	4
2	Clarkston L1	3	25	7	20	10	22	1	40	133	3	1	3
2	Clarkston L2	3	26	7	21	10	27	1	41	138	3	1	3
2	Clarkston L3	3	26	7	22	10	29	1	43	143	3	1	3
4	Clawson	5	12	7	19	14	17	5	67	90	3	0	4
1	Commerce Township	6	25	9	29	21	27	5	88	147	3	0	6
2	Davisburg	3	26	7	22	10	29	1	43	142	3	1	3
0	Drayton Plains	3	24	7	20	9	20	1	37	121	3	1	3
3	Farmington	5	11	4	16	13	12	3	57	63	2	0	4
4	Franklin	5	11	5	18	14	14	4	67	85	2	0	4
3	Hutington Woods	2	2	1	10	6	4	2	20	29	2	0	4
1	Keego Harbor	6	33	11	35	24	32	5	97	179	4	1	7
2	Lake Orion L1	3	26	7	21	10	27	1	41	138	3	1	3
2	Lake Orion L2	3	26	7	22	10	29	1	43	143	3	1	3
2	Lake Orion L3	3	27	7	22	10	30	1	43	145	3	1	3
2	Lake Orion L4	3	27	7	24	10	30	1	45	148	3	1	3
2	Lakeville	3	27	7	24	10	30	1	45	148	3	1	3

f) Visualizing Clusters - To better visualize the clusters, folium library is used to represent all the clusters on a map. Customer work location is marked in red 'info' and the circle represents all the neighborhoods retrieved within 11 mile radius of the work location. We can see that the neighborhoods are clustered based on total count of each categories. For instance, Berkeley and Hutington Woods (part of orange cluster) has the least number of stores/venues across each category (Arts and Crafts Store, Grocery Store etc).

```

map_clusters = folium.Map(location = [42.655822,-83.229792],
zoom_start=10)
folium.Marker([42.655822,-83.229792], popup='User Work Location -
FCA Auburn Hills', icon=folium.Icon(color='red')).add_to(map_clusters)
folium.Circle([42.655822,-
83.229792], radius=25749.5).add_to(map_clusters)

x = np.arange(k_cluster)

```

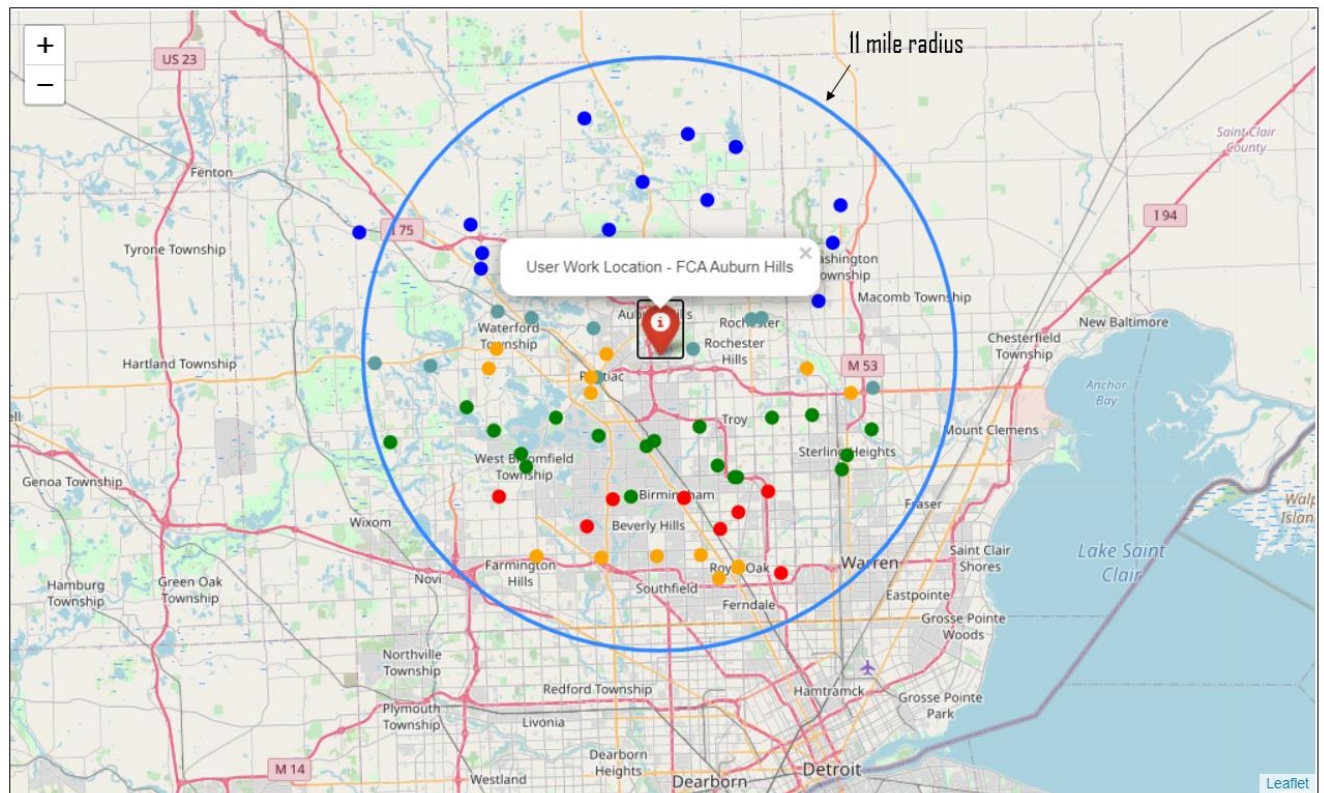


```

ys = [i + x + (i*x)**2 for i in range(k_cluster)]
colors_array = ['green', 'blue', 'orange', 'red', 'cadetblue']
for lat, lon, n, cluster in zip(df_final['Latitude'],
df_final['Longitude'],
df_final['Neighborhood'],
df_final['Cluster Labels']):
    label = folium.Popup(str(n) + ' Cluster' + str(cluster),
parse_html=True)
    folium.CircleMarker(
    [lat, lon],
    radius=4,
    popup=label,
    color=colors_array[cluster-1],
    fill=True,
    fill_color=colors_array[cluster-1],

    fill_opacity=1).add_to(map_clusters)

```



g) Analyzing Clusters - The five clusters are averaged across each category (Arts and Crafts Store, Grocery Store etc) as represented by the last row. This is useful to rank and select the best clusters. The color code of clusters is as follows -

Cluster 1 - Cadet Blue, **Cluster 2** - Green, **Cluster 3** - Blue, **Cluster 4** - Orange, **Cluster 5** - Red

```
df_cluster1 = df_cluster.loc[df_cluster['Cluster Labels']==1,
df_cluster.columns[[1]+list(range(2,df_cluster.shape[1]))]]
df_cluster1.reset_index(drop=True, inplace=True)
av_cluster1 = df_cluster1.mean(axis=0).round(2)
df_cluster_avg = df_cluster1.append(av_cluster1,ignore_index=True)
```

Average

Index	Neighborhood	Arts & Crafts Store	Convenience Store	Grocery Store	Hardware Store	Mediterranean Restaurant	Mexican Restaurant	Noodle House	Pharmacy	Pizza Place	Shipping Store	Supermarket	Taco Place
0	Auburn Hills L1	3	23	7	19	9	20	1	34	109	3	1	3
1	Auburn Hills L2	3	24	7	20	9	20	1	39	123	3	1	3
2	Drayton Plains	3	24	7	20	9	20	1	37	121	3	1	3
3	Pontiac L4	3	24	7	19	9	20	1	34	115	3	1	3
4	Rochester L1	3	23	7	17	9	19	1	32	97	3	1	3
5	Rochester L2	3	23	7	18	9	19	1	33	108	3	1	3
6	Rochester L3	3	24	7	20	9	20	1	37	121	3	1	3
7	Utica L3	3	24	7	19	9	20	1	34	114	3	1	3
8	Waterford L3	3	24	7	20	9	20	1	39	123	3	1	3
9	White Lake L1	2	19	6	11	9	19	1	30	86	2	1	2
10	White Lake L2	3	23	7	19	9	20	1	34	109	3	1	3
11	nan	2.91	23.2	6.91	18.4	9	19.7	1	34.8	111	2.91	1	2.91

Cluster 1

Average

Index	Neighborhood	Arts & Crafts Store	Convenience Store	Grocery Store	Hardware Store	Mediterranean Restaurant	Mexican Restaurant	Noodle House	Pharmacy	Pizza Place	Shipping Store	Supermarket	Taco Place
0	Bloomfield Hills L2	6	19	8	29	21	26	5	85	134	3	0	4
1	Bloomfield Hills L3	6	22	8	29	21	26	5	86	142	3	0	4
2	Bloomfield Hills L4	6	22	8	29	21	26	5	86	143	3	0	4
3	Bloomfield township	6	16	8	28	21	25	5	82	133	3	0	4
4	Commerce Township	6	25	9	29	21	27	5	88	147	3	0	6
5	Keego Harbor	6	33	11	35	24	32	5	97	179	4	1	7
6	Sterling Heights L1	6	15	8	27	20	24	5	81	128	3	0	4
7	Sterling Heights L2	6	19	8	29	21	26	5	85	141	3	0	4
8	Sterling Heights L3	6	30	9	34	24	29	5	93	166	4	1	7
9	Sterling Heights L4	6	33	11	35	24	33	5	97	184	4	1	7
10	Troy L2	6	16	8	28	21	25	5	82	133	3	0	4
11	Troy L3	6	19	8	29	21	26	5	85	135	3	0	4
12	Troy L4	6	19	8	29	21	26	5	85	135	3	0	4
13	Troy L5	6	28	9	31	21	28	5	91	155	3	0	7
14	Troy L6	6	31	9	35	24	29	5	97	170	4	1	7
15	Union Lake	6	32	11	35	24	30	5	97	176	4	1	7
16	West Bloomfield L2	6	16	8	28	20	25	5	81	132	3	0	4
17	West Bloomfield L3	6	19	8	29	21	26	5	85	134	3	0	4
18	West Bloomfield L4	6	27	9	30	21	27	5	91	150	3	0	6
19	nan	6	23.2	8.74	30.4	21.7	27.2	5	88.1	148	3.26	0.26	5.16

Cluster 2

Index	Neighborhood	Arts & Crafts Store	Convenience Store	Grocery Store	Hardware Store	Mediterranean Restaurant	Mexican Restaurant	Noodle House	Pharmacy	Pizza Place	Shipping Store	Supermarket	Taco Place
0	Clarkston L1	3	25	7	20	10	22	1	40	133	3	1	3
1	Clarkston L2	3	26	7	21	10	27	1	41	138	3	1	3
2	Clarkston L3	3	26	7	22	10	29	1	43	143	3	1	3
3	Davisburg	3	26	7	22	10	29	1	43	142	3	1	3
4	Lake Orion L1	3	26	7	21	10	27	1	41	138	3	1	3
5	Lake Orion L2	3	26	7	22	10	29	1	43	143	3	1	3
6	Lake Orion L3	3	27	7	22	10	30	1	43	145	3	1	3
7	Lake Orion L4	3	27	7	24	10	30	1	45	148	3	1	3
8	Lakeville	3	27	7	24	10	30	1	45	148	3	1	3
9	Leonard	3	27	7	24	10	30	1	45	148	3	1	3
10	Oakland	3	26	7	22	10	29	1	43	145	3	1	3
11	Oxford L1	3	27	7	24	10	30	1	45	148	3	1	3
12	Oxford L2	3	27	7	24	10	30	1	45	148	3	1	3
13	Rochester L4	3	26	7	21	10	27	1	41	138	3	1	3
14	Utica L4	3	25	7	20	10	22	1	40	132	3	1	3
15	Washington L1	3	26	7	22	10	28	1	43	142	3	1	3
16	Washington L2	3	27	7	22	10	30	1	43	145	3	1	3
17	nan	3	26.3	7	22.2	10	28.2	1	42.9	143	3	1	3

Average

Cluster 3

Index	Neighborhood	Arts & Crafts Store	Convenience Store	Grocery Store	Hardware Store	Mediterranean Restaurant	Mexican Restaurant	Noodle House	Pharmacy	Pizza Place	Shipping Store	Supermarket	Taco Place
0	Berkley	4	9	3	13	11	10	3	48	57	2	0	4
1	Farmington	5	11	4	16	13	12	3	57	63	2	0	4
2	Hutington Woods	2	2	1	10	6	4	2	20	29	2	0	4
3	Pontiac L1	1	10	3	8	6	10	1	28	48	2	1	2
4	Pontiac L2	2	18	5	10	8	17	1	28	69	2	1	2
5	Pontiac L3	2	18	5	10	8	17	1	28	69	2	1	2
6	Royal Oak L1	3	2	2	10	6	5	2	26	37	2	0	4
7	Royal Oak L2	4	9	3	13	11	10	3	48	57	2	0	4
8	Southfield L1	3	8	2	13	10	9	3	42	48	2	0	4
9	Southfield L2	4	9	2	13	10	10	3	46	55	2	0	4
10	Utica L1	1	4	3	7	5	6	1	18	35	2	1	2
11	Utica L2	2	18	5	11	9	19	1	30	80	2	1	2
12	Waterford L1	1	16	5	10	6	13	1	28	65	2	1	2
13	Waterford L2	2	18	5	11	9	19	1	30	80	2	1	2
14	nan	2.57	10.9	3.43	11.1	8.43	11.5	1.86	34.1	56.6	2	0.5	3

Average

Cluster 4

Index	Neighborhood	Arts & Crafts Store	Convenience Store	Grocery Store	Hardware Store	Mediterranean Restaurant	Mexican Restaurant	Noodle House	Pharmacy	Pizza Place	Shipping Store	Supermarket	Taco Place
0	Birmingham L1	6	13	7	24	16	20	5	71	99	3	0	4
1	Birmingham L2	6	13	7	24	17	20	5	72	102	3	0	4
2	Bloomfield Hills L1	6	13	7	24	17	20	5	72	101	3	0	4
3	Clawson	5	12	7	19	14	17	5	67	90	3	0	4
4	Franklin	5	11	5	18	14	14	4	67	85	2	0	4
5	Madison Heights	5	11	5	17	14	12	4	61	79	2	0	4
6	Royal Oak L3	5	11	5	18	14	14	4	66	85	2	0	4
7	Troy L1	6	13	7	24	18	22	5	74	108	3	0	4
8	West Bloomfield L1	6	12	7	20	16	20	5	70	94	3	0	4
9	nan	5.56	12.1	6.33	20.9	15.6	17.7	4.67	68.9	93.7	2.67	0	4

Average

Cluster 5

The clusters are compared with each other and are ranked based on the average count of stores/venues across each category. The table below shows the comparison between the cluster -

Cluster #	Color Code	Rank	Description
Cluster 2	Green	1	Ranks the best in all the categories.
Cluster 3	Blue	2	a. Maximum number of convenience stores and mexican restaurants. b. Good number of outlets for majority of the categories.
Cluster 1	Cadet Blue	3	Fair number of outlets for each category
Cluster 5	Red	4	a. Good number of mediterranean restaurants, noodle houses, taco places and pharmacy. b. Ranks lowest in convenience and grocery store
Cluster 4	Orange	5	Least outlets for majority of the categories.

A couple of observations -

1. Even though neighborhoods in cluster 5 have good number of restaurant chains and pharmacy, they have the **least number of convenience and grocery stores** and should not be a preferable choice for customer.
2. Few neighborhoods in Cluster 2, Cluster 4 and Cluster 5 are on the border of each other. Moving slightly North or South from neighborhoods in Cluster 2, we enter the neighborhoods of low ranked clusters.
3. Moving to neighborhoods in the extreme North is cluster 3 which is the second best cluster for the customer to consider living.

Cluster 2 and Cluster 3 are the selected as the two best neighborhoods where the customer can look for apartments which will be explored in the next section.

Note - The python code has the analysis for Cluster 3 and the apartments are shown on the map for that cluster only.

h) Retrieving apartments and community information using Realtor API - The Realtor API is queried for all the apartments and houses which are available for rent in Cluster 2 and Cluster 3.



List of Apartments in Cluster 2

Index	city	lat	line	lon	name	postal_code	price_max	price_min
0	Bloomfield Hills	42.6	1610-1681 Bloomfield Place Dr	-83.3	Bloomfield Place Apartments	48302	979	849
1	West Bloomfield	42.6	2510 Woodrow Wilson Blvd	-83.3	Bloomfield on the Green Apartments	48324	1023	778
2	Pontiac	42.6	850 Golf Dr	-83.3	Crystal Lake Apartments	48341	999	899
3	Bloomfield Hills	42.6	5940 Cohasset Ln	-83.3	Whethersfield	48301	1675	1110
4	Pontiac	42.6	900 Martin Luther King Jr Blvd S	-83.3	Fox Pointe Apartments	48341	855	725
5	Pontiac	42.6	311 S Telegraph Rd	-83.3	Bloomfield On The River	48341	720	595
6	Auburn Hills	42.6	3161 Bloomfield Ln	-83.2	Bloomfield Square Apartments	48326	1075	909
7	Auburn Hills	42.6	124 Optimist Rd	-83.2	Parkways of Auburn Hills	48326	2075	1625
8	Birmingham	42.5	1997 Villa Rd	-83.2	Eton Square	48009	2000	975
9	Auburn Hills	42.6	580 Bloomfield Village Blvd	-83.2	Bloomfield Villas Apartments	48326	930	900
10	Birmingham	42.5	211 E Merrill St	-83.2	Merrillwood Collection	48009	4400	1500
11	Auburn Hills	42.6	3280 S Adams Rd	-83.2	Adams Creek	48326	1099	971
12	Birmingham	42.5	555 S Woodward Ave	-83.2	555 Residences	48009	4900	1679
13	Birmingham	42.5	505 E Lincoln St	-83.2	Lincoln House	48009	1790	1790
14	Birmingham	42.5	245 Townsend St	-83.2	Shain Park Apartments	48009	1995	1995
15	Bloomfield Hills	42.5	7480 Bingham Rd	-83.3	Glens of Bloomfield Apartments & Townhomes	48301	1825	1205
16	Wolverine Lake	42.6	2799 Heron Hills Dr	-83.5	Redwood Wolverine Lake	48390	2009	1672

61	Troy	42.6	6919-6973 Venus Dr	-83.1	Amberwood Townhomes	48085	1975	1975
62	Waterford	42.7	1049 N Oakland Blvd	-83.4	Ashton Pines	48327	886	886
63	Waterford	42.7	7380 Arbor Trl	-83.4	Waterford Pines	48327	979	979
64	Waterford	42.6	7035 Round Hill Dr	-83.4	Round Hill	48327	999	809
65	Waterford	42.6	1340 Briarwood Dr	-83.4	Briarwood Apartments	48327	1340	790
66	Waterford	42.6	5302 Sara Ln	-83.4	Elizabeth Lake Estates	48327	1020	760
67	Waterford	42.6	504 Maplebrook Ln	-83.4	Rivers Edge Apartments	48327	952	720
68	Waterford	42.7	5331-5379 Highland Rd	-83.4	Embassy West Apartments	48327	815	715
69	Waterford	42.7	1744 Crescent Lake Rd	-83.4	Crescent Manor	48327	865	865
70	West Bloomfield	42.6	6350 Aldingbrooke Circle Rd N	-83.4	Aldingbrooke Apartments	48322	4305	1175
71	Bloomfield Hills	42.6	1733 Huntingwood Ln Apt C	-83.2	1733 Huntingwood Ln Apt C	48304	1650	1650

List of Apartments in Cluster 3

Index	city	lat	line	lon	name	postal_code	price_max	price_min
0	Auburn Hills	42.7	100 Lake Village Blvd	-83.3	Auburn Gate Apartments	48326	1722	1167
1	Clarkston	42.7	4855 Fox Crk	-83.3	Legends Fox Creek	48346	1064	884
2	Clarkston	42.7	5800 Deepwood Ct	-83.4	Hidden Lakes Apts	48346	1204	1008
3	Clarkston	42.7	5605 Parview Dr	-83.4	Greens Lake Apartments	48346	995	880
4	Clarkston	42.7	6935 Tuson Blvd	-83.4	Chalet Villa	48346	925	825
5	Clarkston	42.7	5147 Lancaster Hills Dr	-83.4	Lancaster Lakes	48346	1375	1040
6	Clarkston	42.7	5801 Bridgewater Dr	-83.4	Bridgewater Park	48346	999	784
7	Clarkston	42.7	5891-5901 Dixie Hwy	-83.4	Independence Square Apartments	48346	999	815
8	Clarkston	42.7	4000 Brookside Rd	-83.4	Encore at Deerhill Villas	48346	1900	1900
9	Clarkston	42.7	8863 Dixie Hwy	-83.5	Bavarian Village on the Lake	48348	1295	799
10	Davisburg	42.8	10771 Dixie Hwy	-83.5	Springrove	48350	999	799
11	White Lake	42.7	3951 Bentwood Cir W	-83.6	Meadow Lake Estates	48383	999	999
12	Lake Orion	42.7	4100 Heron Springs Blvd	-83.2	Heron Springs Townhomes and Apartments	48359	2050	1575
13	Lake Orion	42.7	3355 Thornwood Dr	-83.3	Sycamore Creek Apartments	48359	885	653
14	Lake Orion	42.7	3331 Towne Park Dr	-83.3	Redwood Lake Orion	48359	1724	1673
15	Lake Orion	42.7	2000 Elmhurst Cir	-83.3	The Orion - Orion, MI	48359	1420	990
16	Lake Orion	42.7	3001 Lake Village Blvd	-83.2	Indian Lake Village Apartments	48360	1195	905

36	Washington	42.7	57163 Cypress St	-83	Redwood Washington Township MI	48094	1824	1639
37	Washington	42.7	8442 Stanford N	-83	The Enclave	48094	1549	825
38	Washington	42.7	8155 S Stony Dr	-83	Stony Creek Apartments	48094	860	660
39	Washington	42.7	8204 Washington Blvd	-83	Eagles Landing of Washington	48094	1025	850
40	Romeo	42.8	231 Apple Blossom Way	-83	Apple Valley Apartments	48065	1075	750
41	Romeo	42.8	159 S Bailey St	-83	159 Bailey Street	48065	685	685
42	Oakland Township	42.8	3720 E Clarkston Rd	-83.2	3720 E Clarkston Rd	48363	3900	3900
43	Orion Township	42.8	911 Old Hickory Ln	-83.2	911 Old Hickory Ln	48362	2200	2200
44	Oxford	42.8	100 Lakeview Dr	-83.2	100 Lakeview Dr	48370	1800	1800

i) Customer Constraints on Apartments - The list of apartments can further be narrowed down based on certain factors like rent, type of apartment - single living or family, number of bedrooms etc. The customer reviews can be used to further filter the apartments. For the sake of simplicity, I have defined a range of rent the customer is willing to pay. The following has been chosen -

Min Rent - \$1100, Max Rent - \$2000. A \$50 buffer is kept to add some flexibility in narrowing down the apartments.



List of Apartments in Cluster 2

Index	Neighborhood	Latitude	Address	Longitude	Apartment Name	Zip Code	Max Price(\$)	Min Price(\$)
0	Troy	42.6	1070 Villa Park Dr	-83.1	Rochester Villas Townhomes	48085	1105	1105
1	Sterling Heights	42.6	34792 Oceanview Dr	-83	Sterling Landings Apartments	48312	1300	1075
2	Shelby Township	42.6	14834 Lakeside Blvd N	-83	Harvard Oaks	48315	1390	1290
3	Clinton Township	42.6	35700 Moravian Dr	-83	Woodland Meadows Apts LLC	48035	1409	1059
4	Commerce Township	42.6	2649 Grove Cir	-83.4	Redwood Commerce Township	48382	1600	1600
5	Shelby Township	42.6	14747 Lakeside Blvd N	-83	Lakeside Boulevard Townhomes	48315	1650	1450
6	Bloomfield Hills	42.6	1733 Huntingwood Ln Apt C	-83.2	1733 Huntingwood Ln Apt C	48304	1650	1650
7	Bloomfield Hills	42.6	5940 Cohasset Ln	-83.3	Whethersfield	48301	1675	1110
8	Birmingham	42.5	505 E Lincoln St	-83.2	Lincoln House	48009	1790	1790
9	Bloomfield Hills	42.5	7480 Bingham Rd	-83.3	Glens of Bloomfield Apartments & Townhomes	48301	1825	1205
10	Royal Oak	42.5	4301-4595 Coolidge Hwy	-83.2	Amber Oak Townhomes	48073	1845	1825
11	Troy	42.6	3100 Gloucester Ave	-83.2	Buckingham Square Apts and TH	48084	1850	1050
12	Rochester Hills	42.6	1901 Village Green Blvd	-83.1	Village Green of Rochester Hills	48307	1915	1425



List of Apartments in Cluster 3

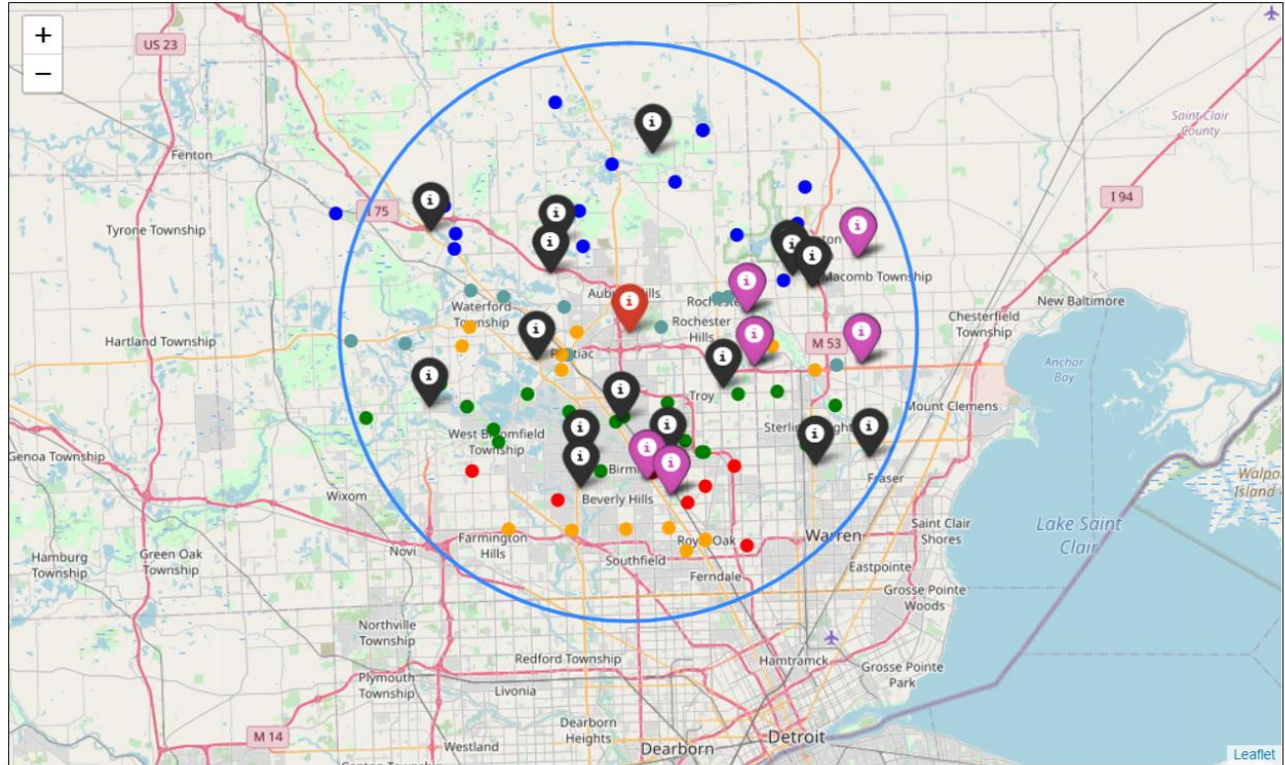
Index	Neighborhood	Latitude	Address	Longitude	Apartment Name	Zip Code	Max Price(\$)	Min Price(\$)
0	Shelby Township	42.7	8245 W Annsbury Cir	-83	Annsbury North Phase I	48316	1350	1350
1	Shelby Township	42.7	54189 Bay Pointe Dr	-83	Annsbury East Apartments	48316	1600	1450
2	Shelby Township	42.7	56114 Stoney Place Dr	-83.1	Stoney Park Place Apartments	48316	1647	1165
3	Shelby Township	42.7	6097 Windemere Ln	-83.1	Windemere East	48316	1700	1700
4	Auburn Hills	42.7	100 Lake Village Blvd	-83.3	Auburn Gate Apartments	48326	1722	1167
5	Lake Orion	42.7	3331 Towne Park Dr	-83.3	Redwood Lake Orion	48359	1724	1673
6	Oxford	42.8	100 Lakeview Dr	-83.2	100 Lakeview Dr	48370	1800	1800
7	Washington	42.7	57163 Cypress St	-83	Redwood Washington Township MI	48094	1824	1639
8	Rochester Hills	42.7	1515 Goldrush Ln	-83.1	Cidermill Village	48307	1876	1242
9	Clarkston	42.7	4000 Brookside Rd	-83.4	Encore at Deerhill Villas	48346	1900	1900

Key Observations -

The apartments highlighted with red in the list above are outliers. These apartments were returned by the Realtor API within 1-mile radius of a neighborhood but apparently exceed the 1-mile radius requirement when cross-checked with google maps. This can be a limitation of the Realtor API.

There are almost equal number of apartments in both the clusters which are in the range of \$1,100-\$2,000. The customers can look at these apartments online and pick a few that interests them the most. The list can further be narrowed down once all the constraints are accounted for.

j) Visualizing Apartment Location - The folium library is used to visualize the apartments location on top of cluster 1 and cluster 2. The apartments are marked with 'black' info. The apartments marked with 'purple' info are outliers and can be dropped. We can see that there are fair number of apartments in cluster 1 and cluster 2 with the rent in the range of \$1,100 to \$2,000.



5. RESULTS

1. The survey data provides an insight about the most frequent visited places by a customer which is used to retrieve neighborhoods having maximum outlets that matches the customer preferences on places they visit.
2. The neighborhoods are clustered using k-means algorithm and the pattern suggests that Cluster 2 (Green) and Cluster 3 (Blue) are the most suitable neighborhoods to consider for relocation as they have good number of outlets for each category (Mexican and Mediterranean Restaurants, Grocery & Convenience Store, Arts & Crafts Store and Pharmacy etc.)
3. A list of apartments is retrieved within a mile radius of these neighborhoods and are filtered based on customer requirements (apartment rent, number of beds etc.). These shortlisted apartments can be recommended to the customer.
4. The present scope of this analysis doesn't account for factors like neighborhood safety score, apartment reviews etc to filter the apartments further but can be extended in the future.

6. CONCLUSION

The capabilities of Data Science and Machine Learning can be extended to a wide variety of real life problems.

In this project, we applied the concept of Data Science to solve the process of shortlisting the neighborhoods and apartments during a relocation process. The new neighborhoods are selected

based on the survey data which tells about the most visited places by a customer in the past. Recommending a new neighborhood is tailored based on historical data of frequently visited places and can vary from customer to customer.