CAPSTONE PROJECT - THE BATTLE OF THE NEIGHBORHOODS

Florent GERALD

INTRODUCTION: BUSINESS PROBLEM

In this project we will try to suggest an optimal location for buying an appartment. Specifically, this report will be targeted to stakeholders interested in byuying an **real state** in **Dhaka Division**, Bangladesh.

Since there are lots of areas in Dhaka we will try to detect **locations that are included only in Dhaka district**. We are also particularly interested in **areas with restaurants and malls in vicinity**. We would also prefer locations **as close to public transportation as possible,** assuming that first two conditions are met.

We will use our data science powers to generate a few most promissing neighborhoods based on this criteria. Advantages of each area will then be clearly expressed so that best possible final location can be chosen by stakeholders.

DATA

Based on definition of our problem, factors that will influence our decision are:

1. numbers of existing venues in the neighborhood (any type of venues)

2. number of and distance to public transport in the neighborhood, if any

3. distance of neighborhood from restaurants and shopping malls

We decided to use the Sub-Office of postal service, to define our neighborhoods.

Following data sources will be needed to extract/generate the required information:

1. The neighborhood location will be determined by using **Geopy**

2. number of venues and their type and location in every neighborhood will be obtained using **Foursquare API**

NEIGHBORHOOD CANDIDATES

Let's create latitude & longitude coordinates for centroids of our candidate neighborhoods. First we need to scrape the Dhaka district's neighbourhood locations and data from this link and then find the latitude & longitude of the neighborhoods, using specific, well known address and geopy.
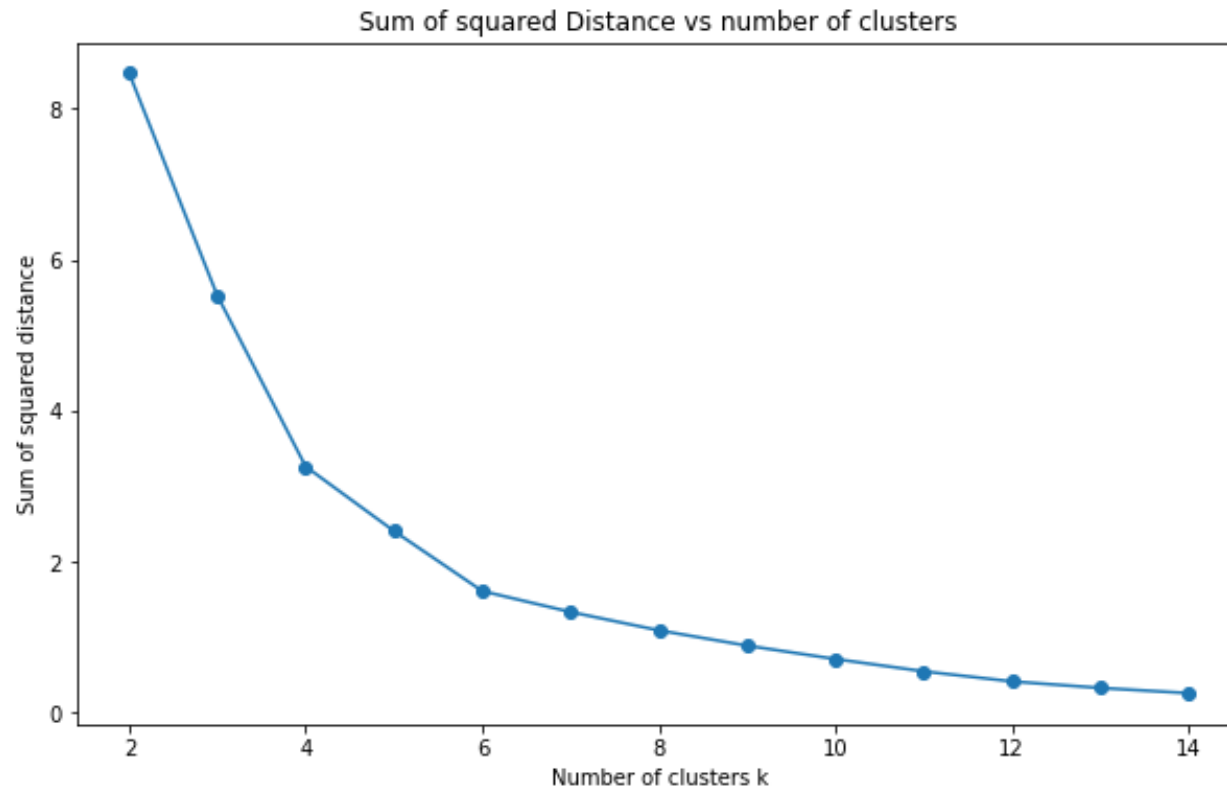
DATA PROCESSING

The dataset contains 1 District, 48 Thana, 48 Sub office (neighbor), 809 unique venues and 35 unique venue categories. We extracted the venue and venue categories by using the Foursquare API.

	Neighborhood	ATM	American Restaurant	Art Gallery	Arts & Crafts Store	Arts & Entertainment	Asian Restaurant	BBQ Joint	Bakery	Bar	Beer Garden	Big Box Store	Bike Shop	Bistro	Boat or Ferry	Burger Joint
0	Dhaka Cantonment--TSO	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	Dhaka Cantonment--TSO	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
2	Dhaka Cantonment--TSO	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	Dhaka Cantonment--TSO	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	Dhaka Cantonment--TSO	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

DATA PROCESSING CONT.

The Data frame consists of categorical values which we have to convert into numerical values for analyzing it. We used a one-hot encoding to convert the data set from the categorical variable.

	Neighborhood	ATM	American Restaurant	Art Gallery	Arts & Crafts Store	Arts & Entertainment	Asian Restaurant	BBQ Joint	Bakery	Bar	Beer Garden	Big Box Store	Bike Shop	Bistro	Boat or Ferry	Burger Joint
0	Dhaka Cantonment-- TSO	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	Dhaka Cantonment-- TSO	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
2	Dhaka Cantonment-- TSO	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	Dhaka Cantonment-- TSO	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	Dhaka Cantonment-- TSO	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

K MEAN CLUSTERING

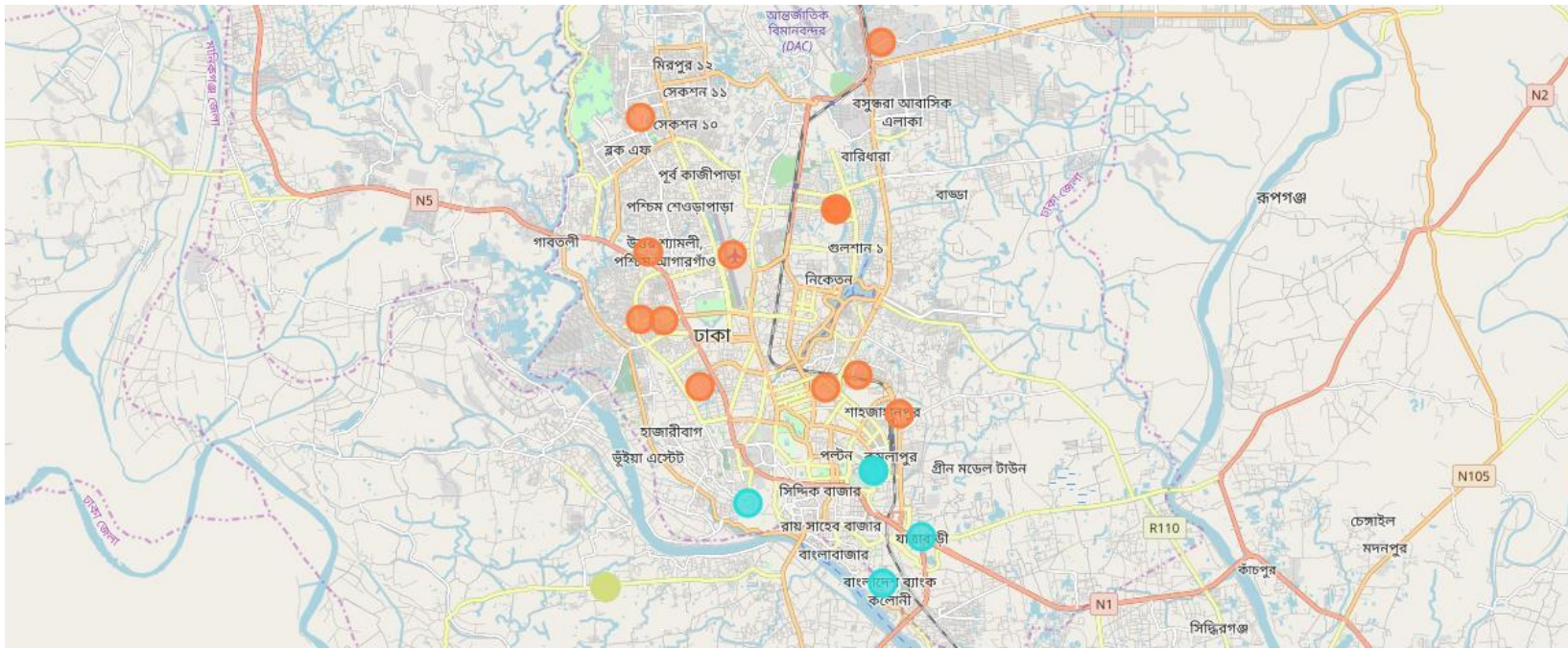Sum of squared Distance vs number of clusters

We used K mean algorithm for this purpose. And we evaluated the number of K by reducing the SSE

1. We used 'k-means++' as initializer.
2. n_init was set to 1000.
3. The max_iter: was set to 1000.

From the graph we can see that if we take 7 clusters then the SSE will be low enough.

MAPPING THE DATA

After clustering the dataset we plotted the location into folium map with different colors to visualize.

RESULT ANALYSIS: CLUSTER 0

Cluster 0 is in Thana Keraniganj with Ati, Keraniganj, Dhaka Jute mills, and kalatia locations which have common attributes within but they have distinct differences with other clusters. The most common restaurant categories for this cluster are American and Turkish restaurants. They also have a cricket ground Department store farmers market within 1.5 KM range.

	Thana	Neighborhood	Post Code	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	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue	Latitude	Longitude
19	Keraniganj	Kalatia	1313	0	American Restaurant	Turkish Restaurant	Fried Chicken Joint	Cricket Ground	Dentist's Office	Department Store	Dessert Shop	Donut Shop	Electronics Store	Farmers Market	23.6982	90.3505
1	Keraniganj	Ati	1312	0	American Restaurant	Turkish Restaurant	Fried Chicken Joint	Cricket Ground	Dentist's Office	Department Store	Dessert Shop	Donut Shop	Electronics Store	Farmers Market	23.6982	90.3505
21	Keraniganj	Keraniganj	1310	0	American Restaurant	Turkish Restaurant	Fried Chicken Joint	Cricket Ground	Dentist's Office	Department Store	Dessert Shop	Donut Shop	Electronics Store	Farmers Market	23.6982	90.3505
8	Keraniganj	Dhaka Jute Mills	1311	0	American Restaurant	Turkish Restaurant	Fried Chicken Joint	Cricket Ground	Dentist's Office	Department Store	Dessert Shop	Donut Shop	Electronics Store	Farmers Market	23.6982	90.3505

CLUSTER 1

Cluster 1 consists of 11 Thanas. They are Gulshan, Mohammadpur, Khilkhet, Uttara, Khilgaon, Ramna, Tejgaon, Dhaka Sadar Dhanmondi, shabujbag, and Mirpur. we have analyzed 14 neighborhoods in this area. they are Khilkhet-TSO, Khilgaon-TSO, Uttara Model Town-TSO, Jigatala TSO, Gulshan Model Town, Sangsad Bhaban TSO, Mirpur TSO, Mohammadpur Housing, Shantinagr TSO, Dhaka Cantonment TSO, Basabo TSO, Tejgaon TSO, Banani TSO, and Badda. The most common item in these areas is Café, Fast Food Restaurant, Shopping Mall, and Hotel.

	Thana	Neighborhood	Post Code	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	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue	Latitude
23	Khilkhet	KhilkhetTSO	1229	1	Shopping Mall	Lounge	Coffee Shop	Hookah Bar	Indian Restaurant	Tea Room	Clothing Store	Gymnastics Gym	Light Rail Station	Bar	23.8307
22	Khilgaon	KhilgaonTSO	1219	1	Fast Food Restaurant	Shopping Mall	Café	Indian Restaurant	Plaza	Market	Coffee Shop	Light Rail Station	Lake	Comfort Food Restaurant	23.7497
37	Uttara	Uttara Model TownTSO	1230	1	Mexican Restaurant	Café	Department Store	Pizza Place	Restaurant	Ice Cream Shop	Asian Restaurant	BBQ Joint	Bakery	Market	23.8755
16	Dhanmondi	Jigatala TSO	1209	1	Café	Fast Food Restaurant	Restaurant	Shopping Mall	Bakery	Ice Cream Shop	Coffee Shop	Art Gallery	Asian Restaurant	Bus Station	23.7471
14	Gulshan	Gulshan Model Town	1212	1	Café	Hotel	Coffee Shop	Korean Restaurant	Asian Restaurant	Fast Food Restaurant	Italian Restaurant	Indian Restaurant	Nightclub	Ice Cream Shop	23.79

CLUSTER 2

In this study, Cluster 2 consists of one thana, Jaypara, with three neighborhood Joypara, Palamganj, and Narisha. The most common items in this location are Big Box store, Turkish Restaurant, and Fried Chicken Joint.

	Thana	Neighborhood	Post Code	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	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue	Latitude	Longitude
17	Joypara	Joypara	1331	2	Big Box Store	Turkish Restaurant	Fried Chicken Joint	Cricket Ground	Dentist's Office	Department Store	Dessert Shop	Donut Shop	Electronics Store	Farmers Market	23.6076	90.125
27	Joypara	Palamganj	1331	2	Big Box Store	Turkish Restaurant	Fried Chicken Joint	Cricket Ground	Dentist's Office	Department Store	Dessert Shop	Donut Shop	Electronics Store	Farmers Market	23.6076	90.125
26	Joypara	Narisha	1332	2	Big Box Store	Turkish Restaurant	Fried Chicken Joint	Cricket Ground	Dentist's Office	Department Store	Dessert Shop	Donut Shop	Electronics Store	Farmers Market	23.6076	90.125

CLUSTER 3

In this study, Cluster 3 consists of one thana, Savar, with ten neighborhoods Amin Bazar, Dairy Farm, Saver P.A.T.C, Rajphulbaria, EPZ, Kashem Cotton Mills, Jahangirnagar University, Savar Cantonment, Savar, and Shimulia. The most common items in this location are Shopping mall, Bus station, and market

	Thana	Neighborhood	Post Code	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	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue	Latitude	Longitude
33	Savar	Saver P.A.T.C	1343	3	Shopping Mall	Bus Station	Market	Turkish Restaurant	Food	Cricket Ground	Dentist's Office	Department Store	Dessert Shop	Donut Shop	23.8477	90.2587
29	Savar	Rajphulbaria	1347	3	Shopping Mall	Bus Station	Market	Turkish Restaurant	Food	Cricket Ground	Dentist's Office	Department Store	Dessert Shop	Donut Shop	23.8477	90.2587
35	Savar	Shimulia	1345	3	Shopping Mall	Bus Station	Market	Turkish Restaurant	Food	Cricket Ground	Dentist's Office	Department Store	Dessert Shop	Donut Shop	23.8477	90.2587
32	Savar	Savar Canttonment	1344	3	Shopping Mall	Bus Station	Market	Turkish Restaurant	Food	Cricket Ground	Dentist's Office	Department Store	Dessert Shop	Donut Shop	23.8477	90.2587
31	Savar	Savar	1340	3	Shopping Mall	Bus Station	Market	Turkish Restaurant	Food	Cricket Ground	Dentist's Office	Department Store	Dessert Shop	Donut Shop	23.8477	90.2587

CLUSTER 4

The Dhamrai thana's Kalampur is a cluster of its own, we can say this is an anomaly. the most common item of kalampur is market, Turkish Restaurant, and fried chicken joint.

	Thana	Neighborhood	Post Code	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	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue	Latitude	Longitude
18	Dhamrai	Kalampur	1351	4	Market	Turkish Restaurant	Fried Chicken Joint	Cricket Ground	Dentist's Office	Department Store	Dessert Shop	Donut Shop	Electronics Store	Farmers Market	23.9202	90.2109

CLUSTER 5

Cluster 5 is a little different it has some of the historical sites. We call this cluster old Dhaka. It consists of Lalbag, Sutrapur, Motijheel and Jatrabari thana. The neighborhoods analyzed are Posta, Gandaria, Dilkusha, Dhania, and Bangabhaban. The most common items in this cluster are train stations, restaurants, and historic sites.

	Thana	Neighborhood	Post Code	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	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
28	Lalbag	Posta TSO	1211	5	Historic Site	Pizza Place	Restaurant	Asian Restaurant	Turkish Restaurant	Fast Food Restaurant	Cricket Ground	Dentist's Office	Department Store	Dessert Shop
13	Sutrapur	Gandaria TSO	1204	5	Train Station	Fast Food Restaurant	Restaurant	Outlet Store	Boat or Ferry	Cricket Ground	Dentist's Office	Department Store	Dessert Shop	Donut Shop
11	Motijheel	DilkushaTSO	1223	5	Restaurant	Hotel	Plaza	Soccer Field	Train Station	Fried Chicken Joint	Cosmetics Shop	Clothing Store	Bus Station	Indian Restaurant
10	Jatrabari	Dhania TSO	1236	5	Train Station	Intersection	Bus Station	Playground	Food	Cricket Ground	Dentist's Office	Department Store	Dessert Shop	Donut Shop
4	Motijheel	BangabhabanTSO	1222	5	Restaurant	Hotel	Plaza	Soccer Field	Train Station	Fried Chicken Joint	Cosmetics Shop	Clothing Store	Bus Station	Indian Restaurant

CLUSTER 6

Cluster 6 has two neighborhoods and they are both from sutrapur thana, they are Dhaka Sadar and wari. the most common item in this cluster in a food restaurant.

	Thana	Neighborhood	Post Code	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	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
9	Sutrapur	Dhaka Sadar HO	1100	6	Comfort Food Restaurant	Asian Restaurant	Turkish Restaurant	Fried Chicken Joint	Dentist's Office	Department Store	Dessert Shop	Donut Shop	Electronics Store	Farmers Market
38	Sutrapur	Wari TSO	1203	6	Comfort Food Restaurant	Asian Restaurant	Turkish Restaurant	Fried Chicken Joint	Dentist's Office	Department Store	Dessert Shop	Donut Shop	Electronics Store	Farmers Market

CONCLUSION

The purpose of this project was to do everything a data scientist does every day in his life. The project gives enough opportunity to learn and exercise exciting new things. Though this project is not impactful enough, it is a start. I want to thank IBM and Coursera to bring such a good course for learning data science.
Florent GERALD