BURAK CELAL AKYUZ

REPORT OF THE FINAL PROJECT OF THE "APPLIED DATA SCIENCE CAPSTONE" COURSE'S PEER GRADED ASSIGNMENT

SUBJECT:

"WHAT" to OPEN at "WHERE" in BURSA

FEBRUARY 9, 2020

Introduction

Bursa is one of the crowded cities in the Turkey. There are almost three million people in Bursa. Bursa's rank in the "World's Liveable Cities, 2016" is 28 in the world and first in the Turkey. According to the industrial statistics Bursa is the heart of automotive production in Turkey. There are big automotive companies such as Bosch, Renault etc. Industrial potential of Bursa is growing day by day. This is also the topic of the project. Imagine that you have seen a potential at "Industrial Tourism Sector". Business people is traveling around the world everyday to visit factories, to make new aggriments etc. and at this vacations they need some places such as hotels to rest/stay. And also the local people is looking for places to spent their time.

So imagine that you have a money to invest but you still have questions about opening "What" in "Where". Bursa has 17 boroughs and all of them has their own people, own lifestyles, own incoming levels. So you can not open everything in everywhere. You have to know the "most prefered places" in each borough. In the next section, we will discover the boroughs of the Bursa and their "most wanted" places to solve our problem.

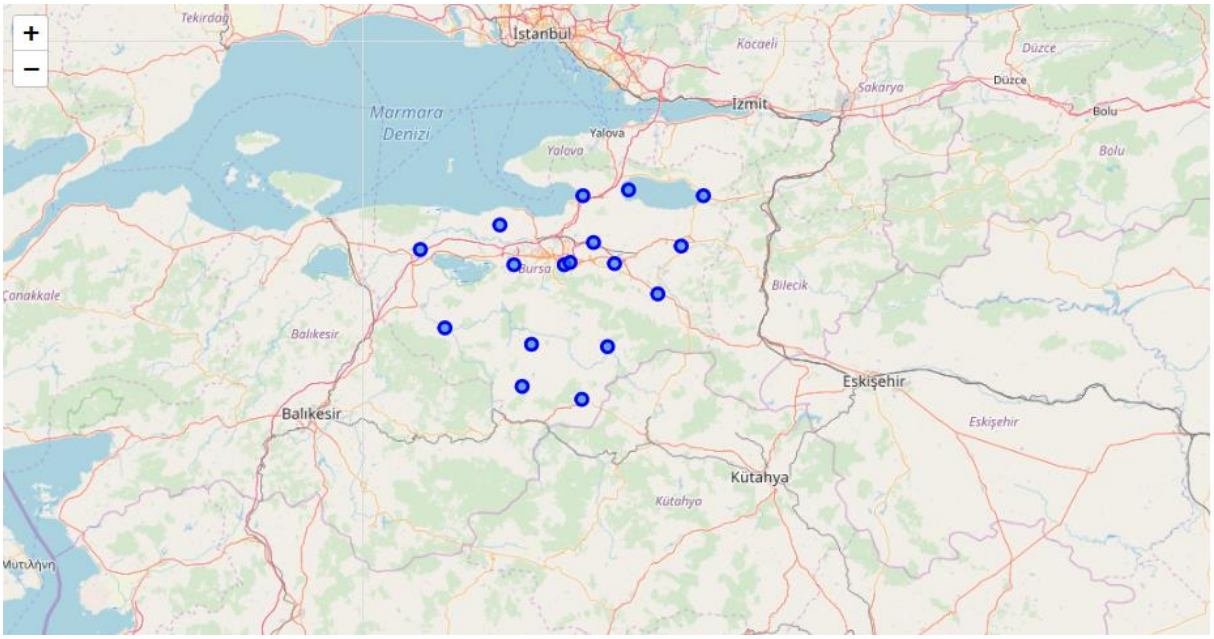
Data

At this section, we will take a look to the data which we are going to use to solve the problem. First I have used the location data of the boroughs of Bursa. Bursa has 17 boroughs. Before we find the venues we need to find the coordinates of the boroughs. I have used the geolocator library to find the latitude and longitude values for each borough. You can find the values below.

	City	Borough	Neighborhood	Latitude	Longitude
0	Bursa	Büyükorhan	Büyükorhan	39.744199	28.870146
1	Bursa	Gemlik	Gemlik	40.430165	29.157070
2	Bursa	Gürsu	Gürsu	40.264367	29.206952
3	Bursa	Harmancık	Harmancık	39.699869	29.151777
4	Bursa	İnegöl	İnegöl	40.080014	29.509645
5	Bursa	İznik	İznik	40.430345	29.722373
6	Bursa	Karacabey	Karacabey	40.237546	28.396002
7	Bursa	Keles	Keles	39.889890	29.271753
8	Bursa	Kestel	Kestel	40.186927	29.304732
9	Bursa	Mudanya	Mudanya	40.325485	28.769453
10	Bursa	Mustafakemalpaşa	Mustafakemalpaşa	39.955087	28.512252
11	Bursa	Nilüfer	Nilüfer	40.182864	28.832399
12	Bursa	Orhaneli	Orhaneli	39.897180	28.913627
13	Bursa	Orhangazi	Orhangazi	40.452753	29.374641
14	Bursa	Osmangazi	Osmangazi	40.182737	29.067835
15	Bursa	Yenişehir	Yenişehir	40.248492	29.621070
16	Bursa	Yıldırım	Yıldırım	40.190099	29.097653

Table 1 Latitude and Longitude values of boroughs

The neighborhood values was empty for the each borough. So I used borough values for the neighborhood values too. Now we can create the map of Bursa. We need the latitude and longitude values of the Bursa. Latitude value is equal to 40.182737 and longitude value is the 29.0678352. Now we can draw a map to see the distrubition of the Bursa's boroughs.



Map 1 Distrubition of the Boroughs

After we draw the map we will use Foursquare API to find the venues for each boroughs. I have used a function to find the venues at 4.1 kilometers radius. Bursa's littlest borough is 'Yıldırım' which is 110 kilometers square. If we want to draw a circle from the coordinate base of the Yıldırım radius of this circle will be 4180 meters. So I used this distance for all boroughs. I have found 904 venues for 17 boroughs. You can see the first and last five rows of the data frame.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Büyükorhan	39.744199	28.870146	Ekomini	39.771730	28.886716	Convenience Store
1	Büyükorhan	39.744199	28.870146	Büyükorhan Yıldırım Belediyesi Parkı	39.775058	28.885938	Theme Park
2	Büyükorhan	39.744199	28.870146	Kapalı Pazar	39.769340	28.886484	Market
3	Büyükorhan	39.744199	28.870146	Büyükorhan Teras Cafe	39.770685	28.886360	Mediterranean Restaurant
4	Büyükorhan	39.744199	28.870146	Büyükorhan Dag Yolu	39.774393	28.868063	Forest

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
899	Yıldırım	40.190099	29.097653	Setbaşı Roma Dondurma	40.180583	29.070860	Ice Cream Shop
900	Yıldırım	40.190099	29.097653	Tuz Pazarı Çarşısı	40.185066	29.065252	Farmers Market
901	Yıldırım	40.190099	29.097653	Ulu Cami	40.183965	29.062331	Mosque
902	Yıldırım	40.190099	29.097653	Bağdat Hurma Tatlıcısı	40.194273	29.064308	Dessert Shop
903	Yıldırım	40.190099	29.097653	Meşhur Kayhan Pideli Köfte	40.184169	29.069145	Turkish Restaurant

Table 2 First and Last Five Rows of the Data Frame

Now, I will group the venues according to counts for all boroughs to see the number of the venues in each borough.

Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Büyükorhan	10	10	10	10	10	10
Gemlik	100	100	100	100	100	100
Gürsu	33	33	33	33	33	33
Harmancık	17	17	17	17	17	17
Karacabey	95	95	95	95	95	95
Keles	17	17	17	17	17	17
Kestel	28	28	28	28	28	28
Mudanya	18	18	18	18	18	18
Mustafakemalpaşa	12	12	12	12	12	12
Nilüfer	90	90	90	90	90	90
Orhaneli	10	10	10	10	10	10
Orhangazi	15	15	15	15	15	15
Osmangazi	100	100	100	100	100	100
Yenişehir	88	88	88	88	88	88
Yıldırım	100	100	100	100	100	100
İnegöl	100	100	100	100	100	100
İznik	71	71	71	71	71	71

Table 3 Number of Venues at Boroughs

We need the categories of the venues to find most common venues. I will use the *unique* method of the dataframe. There are 180 unique categories in our data frame. In the next section, we will use this data to create a new data frame to use in clustering algorithm.

Methodology

In this section, first we will create a new data frame by using the data in Data section and we will use this data frame in clustering algorithm to create clusters which includes the Bursa's boroughs according to the number of boroughs. Let's dive in.

First we are using encoding operation to "Venue Category" column and creating a new dataframe by using this data which named as bursa_onehot_encode. We are assigning borough's neighborhood values to the bursa_onehot_encode data frame. And let's check the head and tail of the bursa_onehot_encode data frame.

	Neighborhood	Accessories Store	Airport Terminal	Arcade	Art Gallery	Arts & Crafts Store	Arts & Entertainment	Athletics & Sports	Auto Garage	BBQ Joint	...	Travel & Transport	Tree	Tunnel	Turkish Coffeehouse	Turkish Home Cooking Restaurant
0	Büyükorhan	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0
1	Büyükorhan	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0
2	Büyükorhan	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0
3	Büyükorhan	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0
4	Büyükorhan	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0

5 rows × 181 columns

	Neighborhood	Accessories Store	Airport Terminal	Arcade	Art Gallery	Arts & Crafts Store	Arts & Entertainment	Athletics & Sports	Auto Garage	BBQ Joint	...	Travel & Transport	Tree	Tunnel	Turkish Coffeehouse	Turkish Home Cooking Restaurant
899	Yıldırım	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0
900	Yıldırım	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0
901	Yıldırım	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0
902	Yıldırım	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0
903	Yıldırım	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0

5 rows × 181 columns

Table 4 Head and Tail of the bursa_onehot_encode Data Frame

We will group this data according to the neighborhoods mean for the each venue type. The gropued data frame has 17 rows which represents number the boroughs and the 181 columns which represents the borough name and 180 unique venues. You can see the data frame at the next page.

	Neighborhood	Accessories Store	Airport Terminal	Arcade	Art Gallery	Arts & Crafts Store	Arts & Entertainment	Athletics & Sports	Auto Garage	BBQ Joint	...	Travel & Transport	Tree	Tunnel	Turkish Coffeehouse
0	Büyükorhan	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.000000	0.000000	0.000000	0.000000
1	Gemlik	0.000000	0.000000	0.020000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.000000	0.000000	0.000000	0.000000
2	Gürsu	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.030303	0.000000	0.030303	...	0.000000	0.060606	0.030303	0.030303
3	Harmancık	0.000000	0.000000	0.058824	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.000000	0.000000	0.058824	0.000000
4	Karacabey	0.000000	0.000000	0.031579	0.000000	0.000000	0.010526	0.000000	0.010526	0.000000	...	0.000000	0.000000	0.000000	0.010526
5	Keles	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.000000	0.000000	0.000000	0.000000
6	Kestel	0.035714	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.000000	0.035714	0.000000	0.035714
7	Mudanya	0.000000	0.000000	0.000000	0.000000	0.000000	0.055556	0.000000	0.000000	0.055556	...	0.000000	0.000000	0.000000	0.000000
8	Mustafakemalpaşa	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.000000	0.000000	0.000000	0.166667
9	Nilüfer	0.000000	0.000000	0.011111	0.000000	0.000000	0.000000	0.022222	0.000000	0.000000	...	0.000000	0.000000	0.000000	0.000000
10	Orhaneli	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.000000	0.000000	0.000000	0.000000
11	Orhangazi	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.133333	0.000000	0.000000	...	0.000000	0.000000	0.000000	0.000000
12	Osmangazi	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.000000	0.000000	0.000000	0.000000
13	Yenişehir	0.000000	0.011364	0.034091	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...	0.000000	0.000000	0.000000	0.000000
14	Yıldırım	0.000000	0.000000	0.020000	0.000000	0.010000	0.000000	0.010000	0.000000	0.000000	...	0.000000	0.000000	0.000000	0.000000
15	İnegöl	0.000000	0.000000	0.030000	0.010000	0.000000	0.000000	0.000000	0.000000	0.010000	...	0.000000	0.000000	0.000000	0.000000
16	İznik	0.000000	0.000000	0.000000	0.042254	0.014085	0.000000	0.000000	0.000000	0.000000	...	0.014085	0.000000	0.000000	0.000000

Table 5 Mean Venue Type of Each Borough

As we can see entire data frame is not so meaningful. We can not tell anything with these 180 venues. We need to filter this data frame. We will use the most 10 venues for each boroughs to make analysis. If we call the head of the filtered data frame the results will be like that.

	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	Büyükorhan	Scenic Lookout	Food Truck	Forest	Convenience Store	Mediterranean Restaurant	Market	Theme Park	Café	Mountain	Fish & Chips Shop
1	Gemlik	Café	Seafood Restaurant	Convenience Store	Gym / Fitness Center	Coffee Shop	Turkish Restaurant	Dessert Shop	Restaurant	Burger Joint	Breakfast Spot
2	Gürsu	Mountain	Forest	Tree	Tea Room	Plaza	Kafenio	Convenience Store	Moving Target	Fast Food Restaurant	Soccer Stadium
3	Harmancık	Café	Campground	Mountain	Forest	Business Service	Soccer Stadium	Bus Station	Cafeteria	Tunnel	Park
4	Karacabey	Café	Convenience Store	Kofte Place	Arcade	Fast Food Restaurant	Soup Place	Kebab Restaurant	Nightclub	Cigkofte Place	Gym / Fitness Center

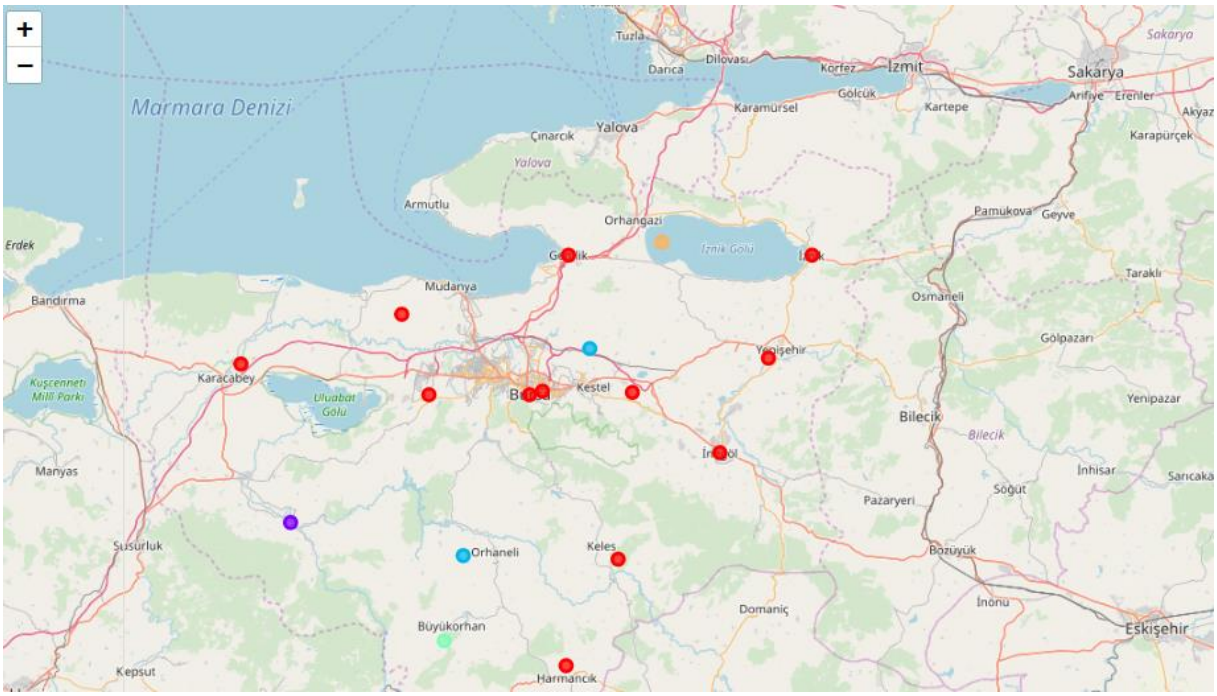
Table 6 Most Visited Venues at Boroughs

And now our data is ready to use in clustering algorithm. I have set 5 clusters for this project. KMeans algorithm from sklearn library is used to find out clusters. Let's check the data frame after clustering algorithm.

	City	Borough	Neighborhood	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue
0	Bursa	Büyükorhan	Büyükorhan	39.744199	28.870146	3	Scenic Lookout	Food Truck	Forest
1	Bursa	Gemlik	Gemlik	40.430165	29.157070	0	Café	Seafood Restaurant	Convenience Store
2	Bursa	Gürsu	Gürsu	40.264367	29.206952	2	Mountain	Forest	Tree
3	Bursa	Harmancık	Harmancık	39.699869	29.151777	0	Café	Campground	Mountain
4	Bursa	İnegöl	İnegöl	40.080014	29.509645	0	Café	Restaurant	Kofte Place

Table 7 Head of The Clustered Data Frame

Now, lets draw a map to see the distribution of our clusters. Each color represent another cluster. Red ones represent the Cluster #1, Purple represents Cluster #2, Blue ones are Cluster #3, Green one is the Cluster #4 and the last one I mean the Yellow is the Cluster #5.



Distribution of the Clusters

Now let’s analyze this results in the next section.

Results and Discussion

In this section, we will closely look to the clusters from our “Methodology” section. I will make my recommendations to you in this section too. So let’s dive in.

We have five clusters which includes the most common venues of boroughs. Let’s start with the first one.

1	Gemlik	0	Café	Seafood Restaurant	Convenience Store	Gym / Fitness Center	Coffee Shop	Turkish Restaurant	Dessert Shop	Restaurant	Burger Joint	Breakfast Spot
3	Harmancık	0	Café	Campground	Mountain	Forest	Business Service	Soccer Stadium	Bus Station	Cafeteria	Tunnel	Park
4	İnegöl	0	Café	Restaurant	Kofte Place	Gym / Fitness Center	Steakhouse	Hotel	Arcade	Motorcycle Shop	Soup Place	Bagel Shop
5	İznik	0	Café	Hotel	Art Gallery	Beach	Castle	Restaurant	Resort	Scenic Lookout	Kebab Restaurant	Turkish Restaurant
6	Karacabey	0	Café	Convenience Store	Kofte Place	Arcade	Fast Food Restaurant	Soup Place	Kebab Restaurant	Nightclub	Cigkofte Place	Gym / Fitness Center
7	Keles	0	Convenience Store	Steakhouse	Tea Room	Pub	Mountain	Bistro	Bed & Breakfast	Café	Restaurant	Clothing Store
8	Kestel	0	Beach	Breakfast Spot	Department Store	Lake	Kofte Place	Intersection	Garden	Rest Area	Moving Target	Knitting Store
9	Mudanya	0	Farm	Café	Lake	Forest	Beer Garden	Beach	Campground	BBQ Joint	Park	Soccer Stadium
11	Nilüfer	0	Café	Steakhouse	Buffet	Farm	Coffee Shop	Pide Place	Turkish Restaurant	Bakery	Restaurant	Soccer Field
14	Osmangazi	0	Historic Site	Café	Bakery	Seafood Restaurant	Turkish Restaurant	Cosmetics Shop	Mosque	Park	History Museum	Gym
15	Yenişehir	0	Café	Restaurant	Plaza	Electronics Store	Mobile Phone Shop	Park	Convenience Store	Arcade	Garden	Tea Room
16	Yıldırım	0	Turkish Restaurant	Café	Bakery	Dessert Shop	Seafood Restaurant	Kebab Restaurant	Mosque	Cigkofte Place	Kofte Place	Sandwich Place

Table 8 Boroughs and Venues in the First Cluster

First cluster is the most crowded cluster. It has 12 boroughs. If we evaluate this cluster the results will be like this.

- 12 of the 17 boroughs of the Bursa have the most common venues.
- %58 of the venues in the first cluster is 'Café' so if you are planning to open a Café in Gemlik, Harmancık, Inegol, Iznik, Karacabey, Nilufer, Yenisehir.
- If we closely look to the data frame, there are three more Café too as the 2nd most common venue at Mudanya, Osmangazi and Yildirim. The people in the first cluster really like to spend their time at Cafés.
- 7th borough has convenience store as the most common venue.
- And the last borough which is Kestel has "Beach" as the most common venue.

Now let's discuss about second cluster.

	Borough	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
10	Mustafakemalpaşa	1	Construction & Landscaping	Turkish Coffeehouse	River	Plaza	Café	Indian Chinese Restaurant	Pier	Forest	Nature Preserve	Flea Market

Table 9 Borough and Venues in the Second Cluster

Second cluster has only one borough which is "Mustafa Kemal Pasa". Mustafa Kemal Pasa's most common venue is constructing but once we look to the second one we can see "Turkish Coffehouse". This is a kind of Café too. So we can easily see and say that us Turkish people really likes the coffee. Let's analyze the third cluster now.

	Borough	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
2	Gürsu	2	Mountain	Forest	Tree	Tea Room	Plaza	Kafenio	Convenience Store	Moving Target	Fast Food Restaurant	Soccer Stadium
12	Orhaneli	2	Mountain	Hotel	Spa	River	Nature Preserve	Campground	Farm	Wings Joint	Farmers Market	Flower Shop

Table 10 Boroughs and Venues in the Third Cluster

At the third cluster we can see a combination. Both of the first common venue for the Gursu and Orhaneli boroughs is Mountain. The second ones are forest and hotel respectively. So we can say that the visitors of the these two boroughs are really like nature and probably they are resting at mountain hotels. Now let's go to the fourth cluster.

	Borough	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
0	Büyükorhan	3	Scenic Lookout	Food Truck	Forest	Convenience Store	Mediterranean Restaurant	Market	Theme Park	Café	Mountain	Fish & Chips Shop

Table 11 Borough and Venues in the Fourth Cluster

Again we have only one borough in the cluster 4th and it is “Buyukorhan”. Most common three venues are “Scenic Lookout”, “Food Truck” and “Forest”. Probabily people who visits here goes to the “Forest” to take some pictures against the “Scenic Lookout” and while doing this they eat somethings from the “Food Truck”. What a scenario. Now let’s look to our final cluster.

	Borough	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
13	Orhangazi	4	Breakfast Spot	Recreation Center	Athletics & Sports	Restaurant	Lake	Farm	Seafood Restaurant	Trail	Nature Preserve	Park

Table 12 Borough and Venues in the Fifth Cluster

In our last cluster we have “Breakfast Spot” , “Recreation Center” and “Athletics & Sports” as top three. The people who visits the Orhangazi are social people. Probabily, they started their visits with a breakfast maybe in a recreation center and they goes to the sport from there for a healthy life.

Conclusions

At this project we have analyzed the boroughs of Bursa according to the number and type of the venues in each borough. We started this project with finding the latitude and longitude values of each borough. Then we have used Foursquare API to find the venues at each borough. After that we have grouped this venues according to the categories and choosed the top 10 category for analysis. We have divided our data to the 5 clusters. Our major aim was answering the “What” to open at “Where” question. Now we can say the conclusions below.

- If you are planning to open a “Café” you can prefer Gemlik, Harmancık, Inegol, Iznik, Karacabey, Nilufer and Yenisehir boroughs as your primary borough and as your secondary you can choose Mudanya, Osmangazi and Yildirim.
- If you wish to open a “Convenience Store” you can prefer Keles borough which is the only borough has convenience store as first common. Even if Karacabey has “Café” as first common you still can try your chance to open a “Convenience Store” at there too.
- Kestel has “Beach” as the first most common venue. If you like sun, sea and sand trinity you can try your chance at Kestel.
- Mustafa Kemal Pasa is the only borough which has “Turkish Coffe House”. Name of this borough comes from the founder of the Turkish Republic Mustafa Kemal Ataturk. According to the historical data Mustafa Kemal loves coffe too much and now we can see that the people is at the Mustafa Kemal Pasa loves coffee too.
- Gursu and Orhaneli are for the nature tourism. If you love nature I suggest you to open a hotel at the mountains of these two boroughs.
- Buyukorhan has both “Scenic lookout” and “Forest”. So the visitors are coming to see these two beauties. You can choose two 2nd most common venue which is “Food Truck” two sell somethings to visitors of the nature.

- Orhangazi is the most social borough I think. The people in there go for “Breakfast Spot” first then they visit “Recreation Center” and go to the “Athletics & Sports” to burn their calories to live a healthy life. I think you can try your chance with a “Breakfast Spot” too.

Now we know the answer of the question. Now you can prefer your venue and your investment tool according to your budget.

That is all for this project. Thank you for your read and for your feedbacks.