Battle of Neighborhoods-The Smart Investor

Finding the best neighborhood of Istanbul to invest in Fastfood

A.1.Business Problem

Imagine that ,You are an investor who want invest into profitable businesses on all around the world. You have the capital and want to invest in Turkey.There are many options on the desk.

Invest in a fast food restaurant in Istanbul ?

If we ask this question without using our data science powers , it sounds quite bare and meaningless. Turkey's fast food market size 2.4 billion dollars and it continues to grow.The word of "2.4 billion dollars market size" Things are getting heating up. The financial health and stability the sector is strong and getting stronger. Increasingly , Turkey's young population promises a high potential on fast food sector. Aim of this project is to find optimal locations to open franchise fast food restaurants.Especially , This report will be targeted to guide investors who wants to invest in fast food restaurant in Istanbul. Istanbul is the 14th most populated metropolis of the world with its 15 million people, 38 borough and 968 neighborhood. Therefore, a little question comes into my mind, Which part of Istanbul we should invest ? My aim is to detect neighborhoods that have less rivals and have most potential customers. Now , Lets's dive into data ocean to find optimal answers.

A.2. Data Description

1. I built my project on 2 main factors,

a. I focused on percentage of fast food restaurants on all food venues.I categorized venues into 4 groups include fast food and venues which can be alternative to fast food restaurants. (venue_categories.csv)

Fast Food

Restaurants

Cafe (includes cafe's and bars)

Dessert shops

b. I focused on age -fast food consumption correlation, People between 15–35 age group are my target customers;

According to a statistical hypotesis, age and fast food consumption have strong correlation.(provided by "Ekev Academic Magazine")

In addition , According to a survey about fast food habits in Turkey , 45% of fast food consumers were aged 16–34 (provided by "Mediabrands/Insight Research Company")

2. Second-level Administrative Divisions of the Turkey json file to create coropleth map via ratio of potential customers.(Provided by Spatial Data Repository of NYU)(json file) [1]

3. Foursquare api to detect food venues in neighborhoods[2]

4. Turkish Statistical Institute , Medas platform –Demographic data of Istanbul boroughs to detect age and population distribution (istanbul_demographic.csv)[3]

5. Geocoder Library to detect coordinates of neighborhoods

6. İstanbul Metropolitan Municipality open data portal to detect neighborhood populations (neighborhood_pop.csv) [4]

7. Google Map, to get the coordinates of the each borough. (istanbul_geo.csv) [5].

8. Foursquare api lets 800 request per day , So I deleted some neighborhoods from 963 to 798 and saved as (neighborhood_coordinates.csv) Query takes nearly 30 minutesData Description

1. Second-level Administrative Divisions of the Turkey json file to create coropleth map via ratio of potential customers.(Provided by Spatial Data Repository of NYU)
2. Foursquare api to detect food venues in neighborhoods
3. Turkish Statistical Institute , Medas platform –Demographic data of Istanbul boroughs to detect age and population distribution
4. Geocoder Library to detect coordinates of neighborhoods
5. İstanbul Metropolitan Municipality open data portal to detect neighborhood populations

B. Methodology

B.1.Data Cleaning and Pre-processing

I got demographic data of Istanbul from Turkish Statistical Institute , sum up to target customer group.Target customer group is people aged between 15-34

	borough	15-19_divorced	15-19_partner_died	15-19_married	15-19_never_married	20-24_divorced	20-24_partner_died	20-24_married		sum 15-34	sum 35-90	sum_all	borough	ratio 15-34
0	Adalar	NaN	NaN	4	844	2	NaN	55	0	3912.0	10090.0	14002.0	Adalar	0.279389
1	Arnavutkoy	1.0	NaN	476	23960	134	4.0	6660	1	102163.0	106826.0	208989.0	Arnavutkoy	0.488844
2	Atasehir	NaN	1.0	191	27107	107	1.0	2811	2	124353.0	210113.0	334466.0	Atasehir	0.371795
3	Avcilar	1.0	NaN	268	27890	131	4.0	4250	3	125781.0	196295.0	322076.0	Avcilar	0.390532
4	Bagcilar	3.0	NaN	437	59664	202	11.0	8725	4	241586.0	303677.0	545263.0	Bagcilar	0.443063

B.2 Borough coordinates , neighborhood names and populations imported

	borough	neighborhood	neighborhood_pop
0	Adalar	Burgazada,Adalar	1418
1	Adalar	Heybeliada,Adalar	3837
2	Adalar	Kinaliada,Adalar	1731
3	Adalar	Maden,Adalar	3916
4	Adalar	Nizam,Adalar	3048

	Borough	Latitude	Longitude
0	Adalar	40.8619	29.1208
1	Arnavutkoy	41.1956	28.7352
2	Atasehir	40.9831	29.1279
3	Avcilar	40.9880	28.7170
4	Bagcilar	41.0450	28.8338

I used borough coordinates and neighborhood names on Geocoder library .Thus I found coordinates of neighborhoods.Therefore, I zipped neighborhood names and boroughs to avoid confusion of neighborhoods with same name. Example: There are 2 neighborhood have same name in two different boroughs. I zipped as "15 temmuz , Esenler" ," 15 Temmuz,Bagcilar"

B.3. I got neighborhood coordinates by using Geocoder Library

	borough	neighborhood	neighborhood_pop	lat	lon
0	Adalar	Burgazada,Adalar	1418	40.88113	29.07024
1	Adalar	Heybeliada,Adalar	3837	40.87682	29.10167
2	Adalar	Kinaliada,Adalar	1731	40.90941	29.05241
3	Adalar	Maden,Adalar	3916	40.87500	29.12984
4	Adalar	Nizam,Adalar	3048	40.87501	29.12913

B.4. I got venues names and coordinates using foursquare api

Foursquare api lets 800 request per day , So I deleted some neighborhoods from 963 to 798 and saved as "neighborhood_coordinates.csv" Query takes more than 30 minutes.

	neighborhood	nb_lat	nb_lon	venue	venue_lat	venue_lon	venue_cat
0	Burgazada,Adalar	40.88113	29.07024	Burgazada Sahil	40.881171	29.069600	Beach
1	Burgazada,Adalar	40.88113	29.07024	Sait Faik Abasıyanık Müzesi	40.881015	29.067458	History Museum
2	Burgazada,Adalar	40.88113	29.07024	Sinem Dondurma	40.880984	29.069779	Ice Cream Shop
3	Burgazada,Adalar	40.88113	29.07024	Burgazada Deniz Kulübü	40.879190	29.070356	Other Great Outdoors
4	Burgazada,Adalar	40.88113	29.07024	Adalar Su Sporları Kulübü	40.879114	29.072084	Beach

I got 55.760 venues from 798 neighborhoods

B.5. Food venues categorized

I categorized venues into 4 groups so that I can easily analyse neighborhoods.4 main venue groups are ;

1. Fast Food 2.Restaurants 3.Cafe (includes cafe's and bars) 4.Dessert shops

|	neighborhood	nb_lat	nb_lon	venue	venue_lat	venue_lon	venue_cat
0	Burgazada,Adalar	40.88113	29.07024	Sinem Dondurma	40.880984	29.069779	dessert shop
1	Burgazada,Adalar	40.88113	29.07024	Adalar Cemevi Çay Bahçesi	40.879195	29.068156	cafe
2	Burgazada,Adalar	40.88113	29.07024	Four Letter Word	40.880573	29.068838	cafe
3	Burgazada,Adalar	40.88113	29.07024	İndos Bar	40.879317	29.069743	cafe
4	Burgazada,Adalar	40.88113	29.07024	İskele Çardak 1947	40.880486	29.069798	restaurant

C.Analysis

C.1.Exploratory Data Analysis

C.1.1. Age Distribution by Borough

Let's take a first look at demographics of Istanbul. At first glance , Esenyurt borough attracted my attention. Crowdest borough of Istanbul in additon Esenyurt has the most potential customer ratio.
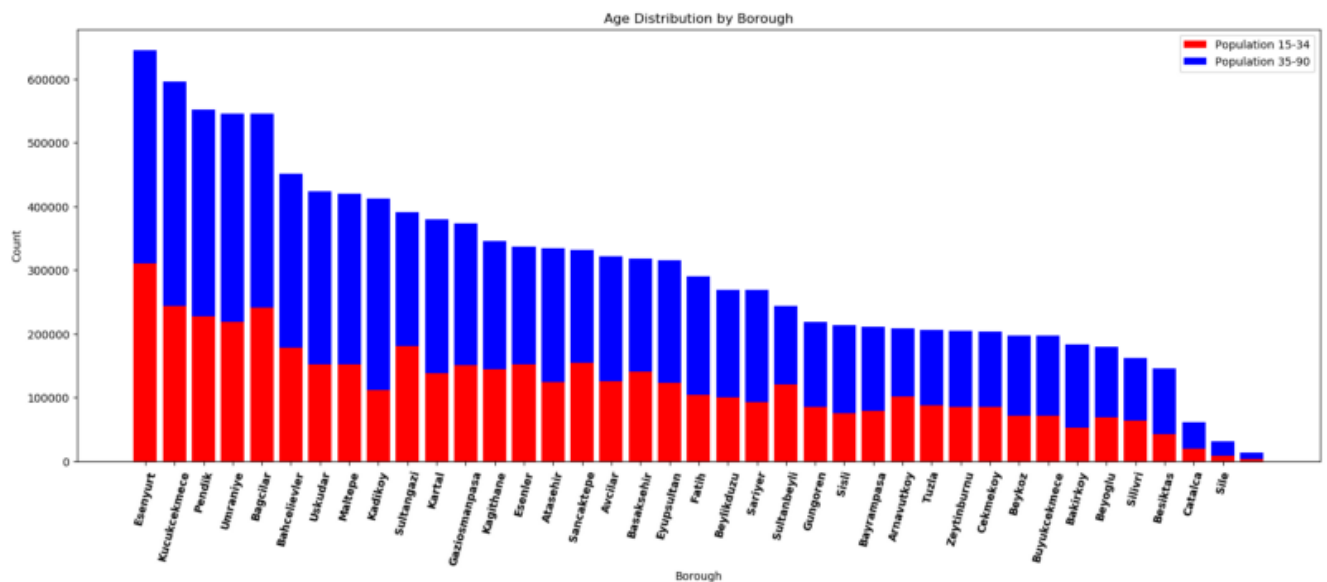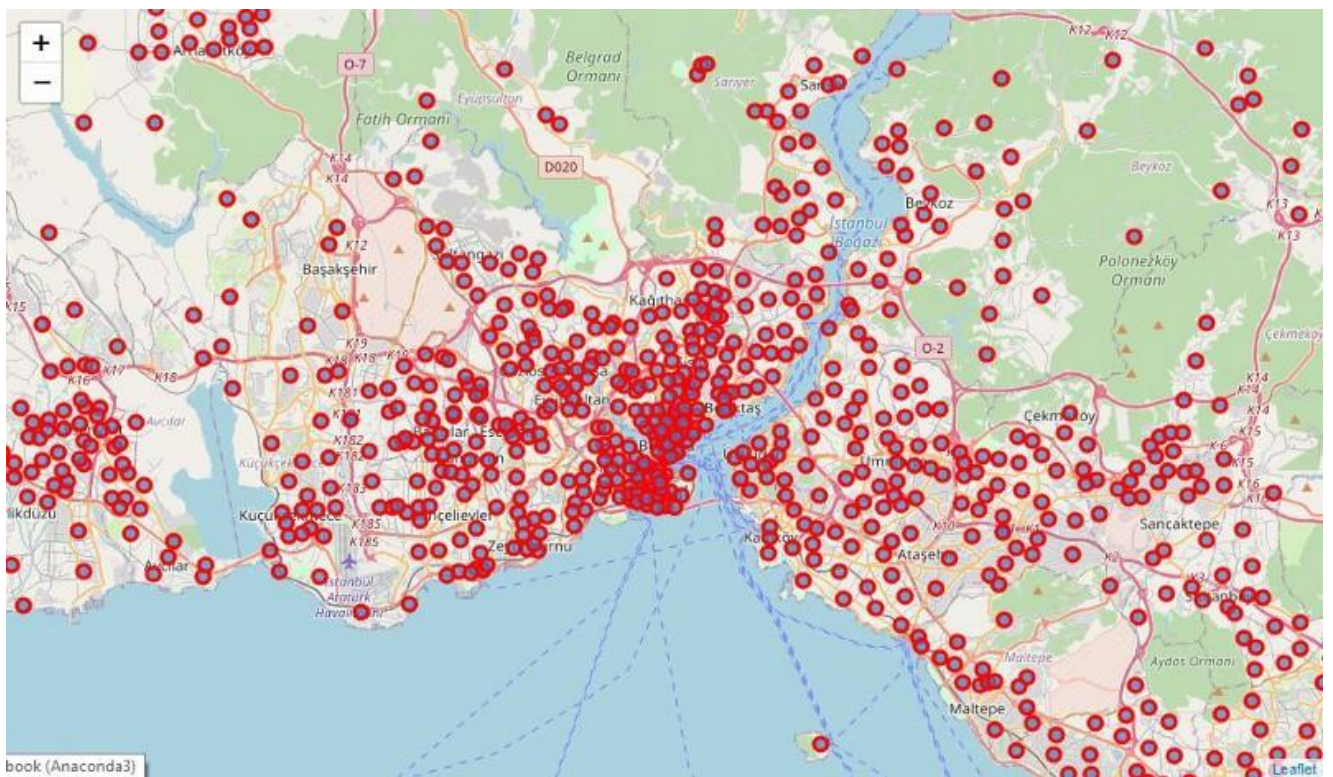

Age distribution by borough

C.1.2. Let's look at borough centers and how its distributed
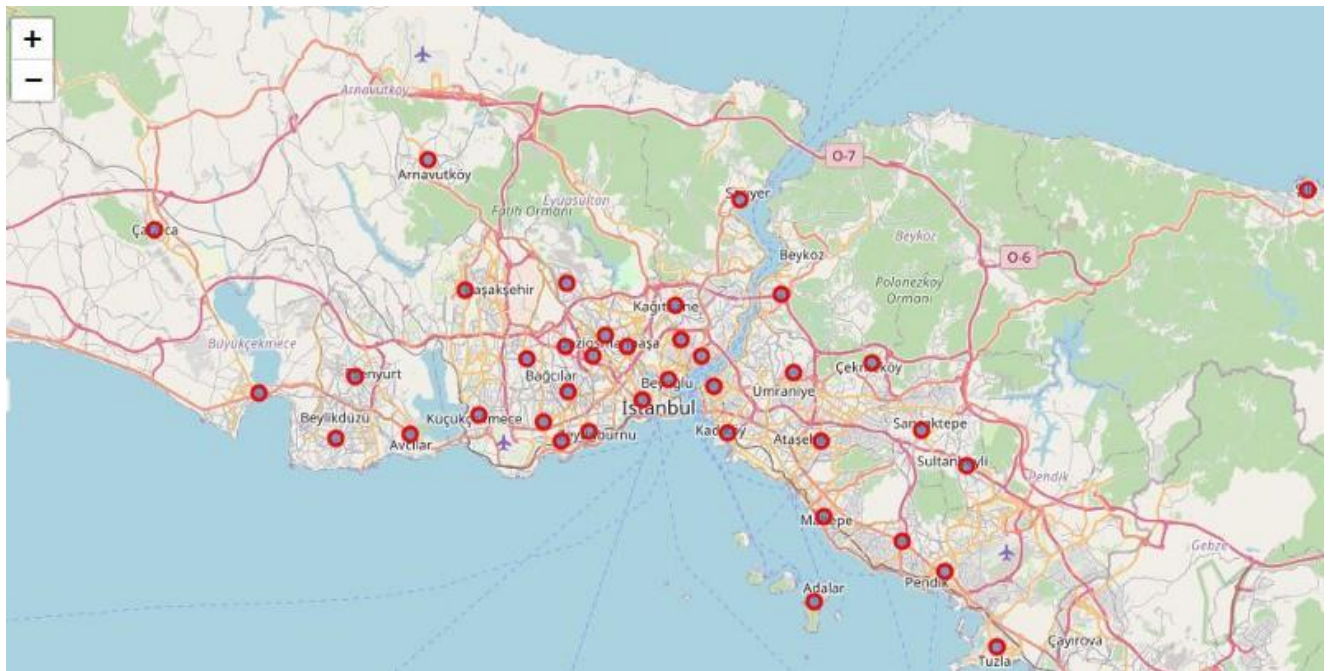
C.1.3. Neighborhoods on the map

By now , All we see is chaos , but We will discover hidden order in chaos

Neighborhoods of Istanbul

C.2.Clustering

My aim is to analyse neighborhoods by percentage of fast food restaurants on all food venues,so I grouped venues by neighborhood and I took mean of frequency.

1. First ,categorical data is transformed into numerical data with one-hot encoding.

	cafe	dessert shop	fast_food	restaurant	neighborhood
0	0	1	0	0	Burgazada,Adalar
1	1	0	0	0	Burgazada,Adalar
2	1	0	0	0	Burgazada,Adalar
3	1	0	0	0	Burgazada,Adalar
4	0	0	0	1	Burgazada,Adalar

2.I grouped venues by neighborhood and I took average of frequency of each venue category.Then , I clustered all neighborhoods with k-means algorithm.

	neighborhood	cafe	dessert shop	fast_food	restaurant	labels
0	100. Yil,Bagcilar	0.366667	0.033333	0.333333	0.266667	2
1	15 temmuz,Bagcilar	0.343750	0.062500	0.187500	0.406250	1
2	15 temmuz,Esenler	0.343750	0.062500	0.187500	0.406250	1
3	19 mayis,Buyukcekmece	0.409091	0.000000	0.181818	0.409091	1
4	19 mayis,Kadikoy	0.404255	0.063830	0.212766	0.319149	1

Here we have important question,How can I choose optimal number of clusters? Answer is Elbow technique;

The Elbow Method showing the optimal k
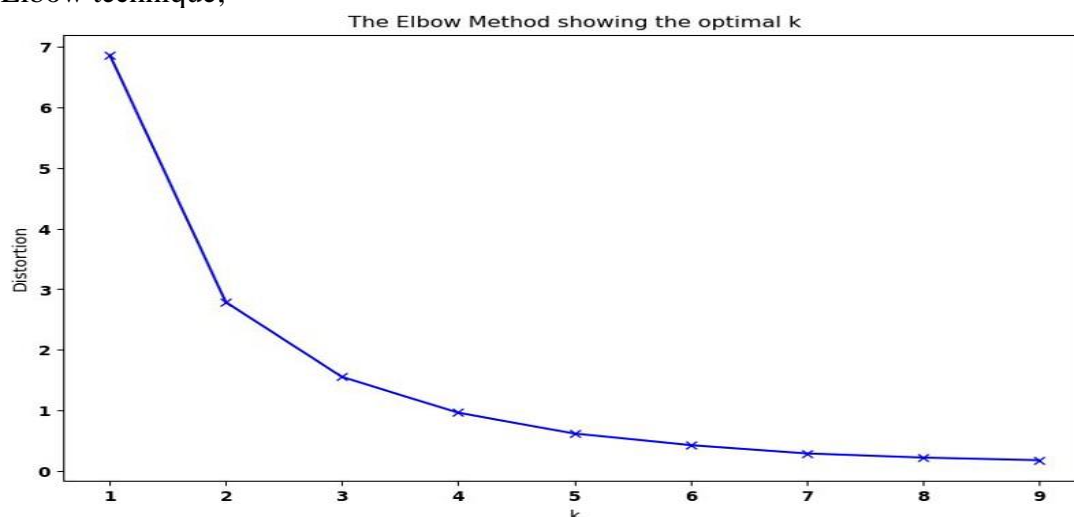
As we see , after k=4 , slop of graph is decreasing , That shows optimal k value is 4.

3.I labeled clusters

Let's check the labels, as we see , the most intense fast food venues labeled as 2.The more intense fast food ratio values means the more rivals we have. We should focus on less rivals and high potential customers.

	borough	neighborhood	neighborhood_pop	lat	lon	fast_food	cafe	restaurant	dessert shop	labels	label_name
519	Sancaktepe	Pasakoy,Sancaktepe	1236	41.00908	29.28492	0.666667	0.000000	0.333333	0.000000	2	Investment Priority 4
464	Sultanbeyli	Mimar Sinan,Sultanbeyli	14565	40.99424	29.26869	0.600000	0.000000	0.000000	0.400000	2	Investment Priority 4
333	Catalca	Kaleici,catalca	5760	41.15027	28.45312	0.500000	0.000000	0.500000	0.000000	2	Investment Priority 4
722	Pendik	camlik,Pendik	12349	40.92634	29.27361	0.500000	0.500000	0.000000	0.000000	2	Investment Priority 4
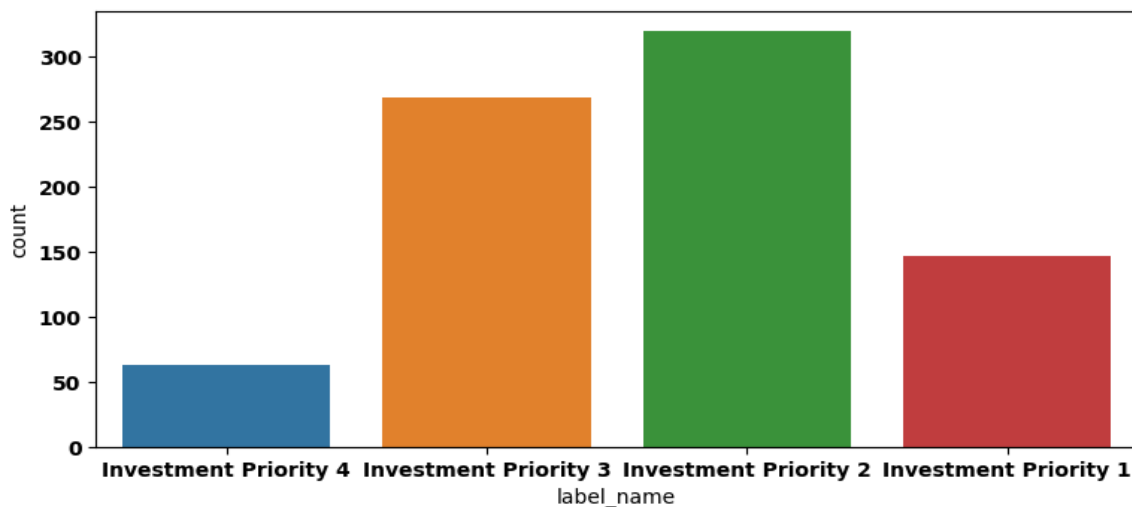
Total count of clusters

D.Results

D.1.Results

Which part of Istanbul we should invest ?

Here comes the answer for our vital question that we asked in "Introduction: Business Problem section "

D.2.Visualization

Cluster 2 : Investment Priority 4, color:turquoise

Cluster 1 : Investment Priority 3 ,color:purple

Cluster 3 : Investment Priority 2 ,color:yellow

Cluster 0 : Investment Priority 1 ,color:red

Data **visualization** gives us a clear idea of what the information means by giving it through map .This makes the data more natural for the human mind to comprehend and therefore
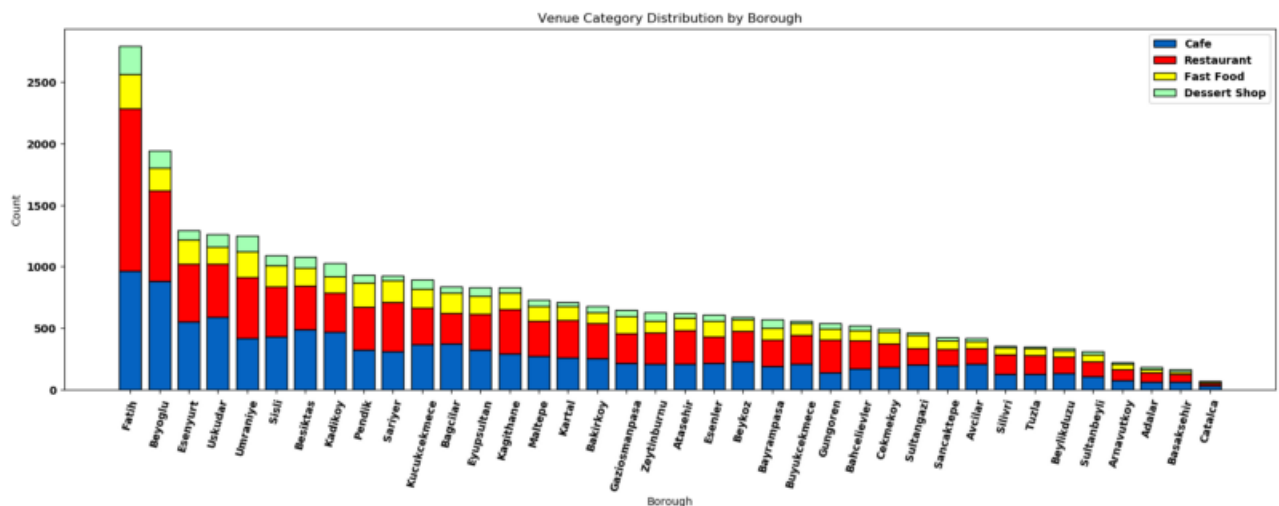
makes it easier to identify .I proposed 4 information in pop-ups; 1.Neighborhood name 2. Borough name 3.Investment Priority 4.Neighborhood population .

Coropleth map visualises 15–34 population ratio from yellow to maroon.
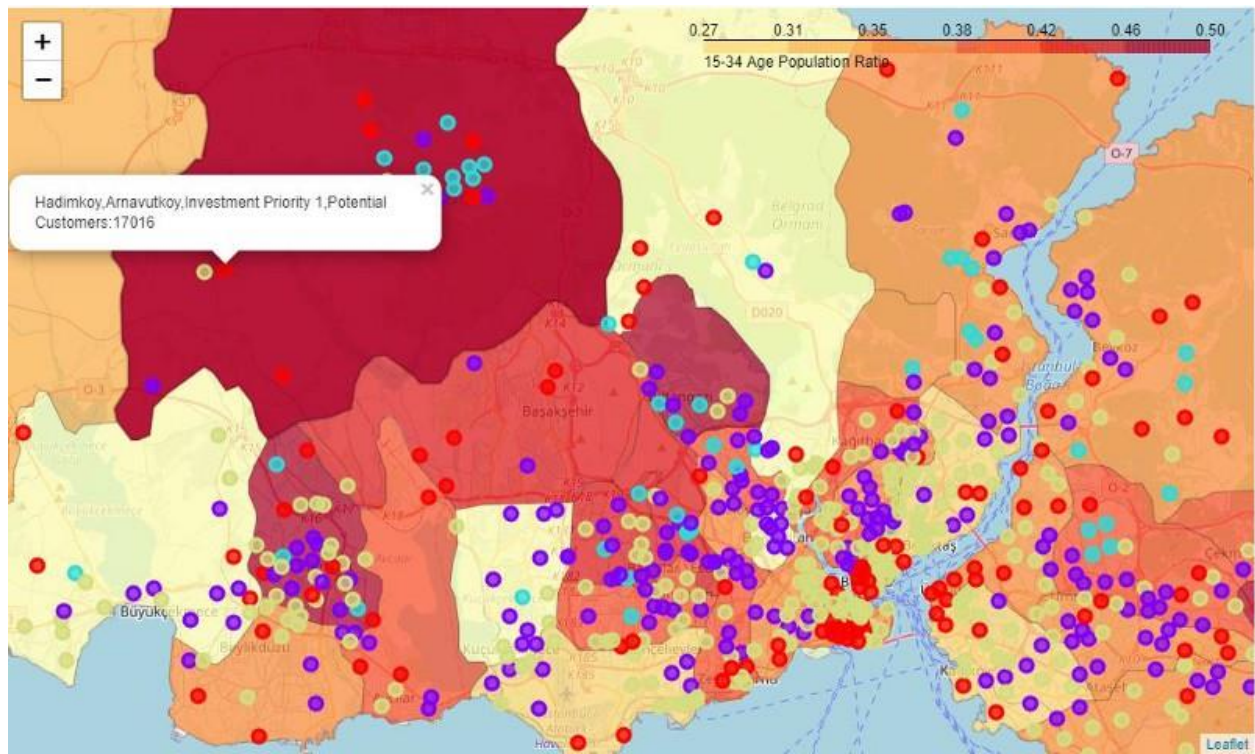
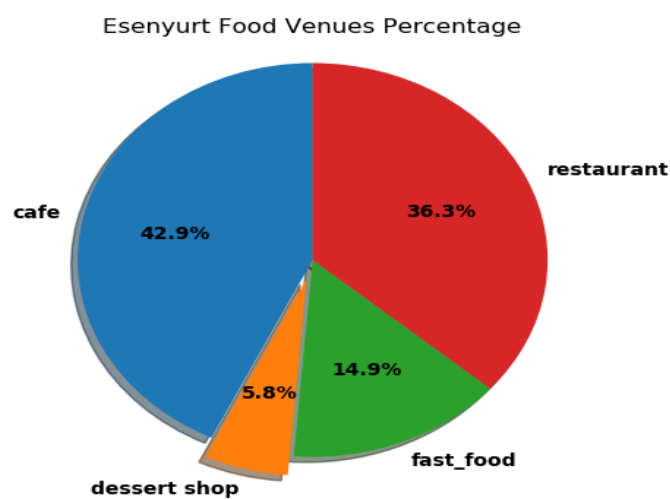Clustered Neighborhoods of Istanbul

E. Discussion

As wee see on graph,Fatih first , Beyoglu second most crowded places by venue count. These boroughs are very attractive places for tourists. Esenyurt is more residential area also the most populated borough.

Smart investor wants to pay less and to take more . At first glance , we'll focus on red points on maroon area. That shows "Investment Priority 1" also potential customer ratio high.We will clarify right place to invest at 2 steps.

E.1.First we will clarify boroughs ;

The boroughs have maroon color are : Arnavutkoy , Esenyurt, Sultangazi ,Sancaktepe and Sultanbeyli.

Esenyurt is the most populated borough also has the biggest potential customer ratio , So my first investment choice is Esenyurt , and the second one is Basaksehir Let's take a close look to Esenyurt , At Esenyurt , population between the age of 15 and 90 is 645782 and there are 1293 food venues. Esenyurt has only % 14.9 percentage fast food venues although it has 1293 food venues.


E.2. Second we will clarify neighborhoods;

Selahaddin Eyyubi , Hurriyet , Sultaniye , Talatpasa neighborhoods of Esenyurt are in "Investment Priority 1" cluster . Also their populations are 17515,17657,14553, 18615 . İstanbul's average population by neighborhood is 11963.Comperatively , These neighborhoods has overpopulated by far .
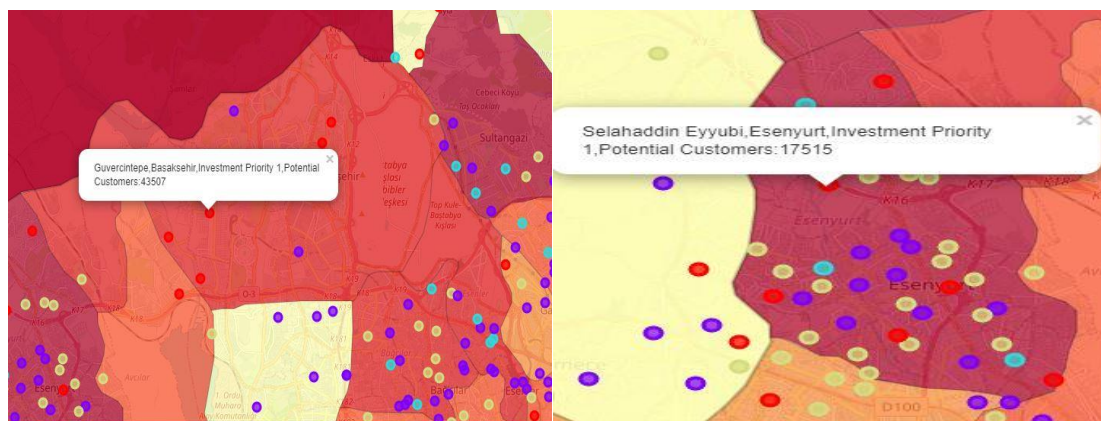
Guvercintepe neighborhood,Başakşehir   Selahaddin Eyyubi Neighborhood,Esenyurt

Another shining star is Basaksehir;
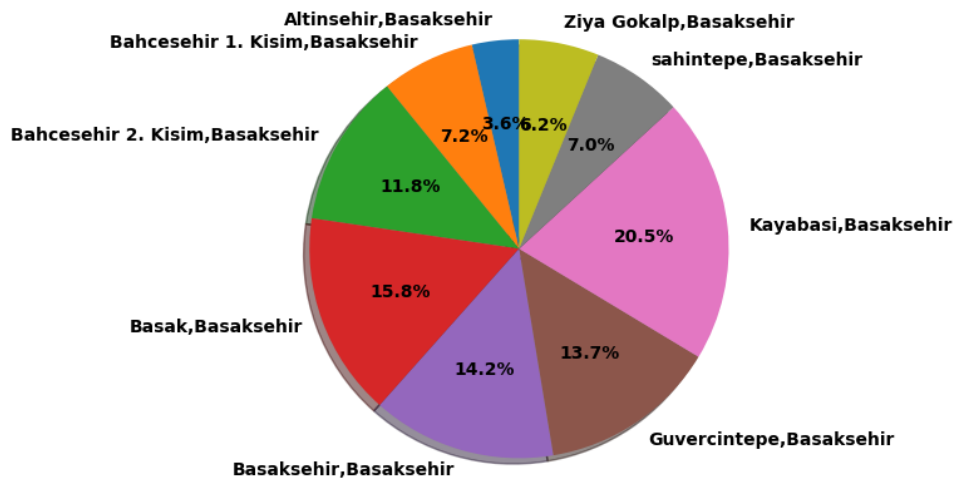
Eventhough Basaksehir borough's potential customer ratio lower than Esenyurt , Guvercintepe,Basak and Basaksehir neighborhoods are good options for investment.**Guvercintepe,Basak and Basaksehir neighborhood** populations are 43507, 49899 and 44928 .In other words, nearly 3 times bigger than Istanbul's average.These 3 neighborhoods has %43.7 percentage of Basaksehir. Half of Basaksehir's population packed in these 3 neighborhoods
In additon , as we see on the map , food venue concentration around these neighborhoods are low .**That's attractive.**

F.Conclusion

At this covid 19 era , We are witnessing capstone moments of mankind history . Our daily routines are changing at a tremendous speed. Pandemic boosted digitalization. As a natural result of this revolutionary changes, our customer habits are changing.Online food delivery has became more important for both business owners and customers as more people tend to order takeout during pandemic. **Superstar of online food delivery is fast food sector .**

Turkish people are exceptionally young people. The median age around 30. In additon, working hours are long. Turkey's average working hours is 9 hours and significantly higher than European countries.Working young Turkish people has less time to cook at home .This is fueling fast food sector and makes it profitable business model . That s why I decided to examine fast food business.

Data driven decisions are more precise than heuristic decisions . This project was my first end to end data science project . I experienced every step of a data science project from defining business problem to having managerial insights . I also learn to use different data sources and web scraping techniques . Especially the most important earning of this project to me is to having vision what can I do with data. It was exciting experience .

E.References

1. [Second-level Administrative Divisions of the Turkey](https://geo.nyu.edu/catalog/stanford-nj696zj1674)

2. [Foursquare API](https://developer.foursquare.com/)

3. [Medas Platform](https://biruni.tuik.gov.tr/medas/?kn=95&locale=tr)

4. [Neighborhood populations](https://en.wikipedia.org/wiki/Istanbul)

5. [Google Map](https://www.google.com/maps/)

6. https://worldfood-istanbul.com/Articles/turkish-fast-food-sector-keeps-on-growing

7. https://www.thebarbecuelab.com/fast-food/

8. http://www.ekevakademi.org/Makaleler/1527107404_03%20Ece%20Armagan-Yasin%20ESKICI.pdf