

The Battle of the Neighborhoods – Istanbul

By Omer Korkmaz

Wednesday, December 16, 2020

Report Subject: Searching for the best neighborhood to open up a theatre in Istanbul using Data Science

This report contains the project called "The Battle of the Neighborhoods - Istanbul". The project is the final assignment of IBM Data Science Professional Certificate Courses (9 of 9).

Index

a. Business Understanding

Includes shared business understanding with stakeholders.

b. Data Understanding

Includes source and description of the data that will be used to solve the problem.

c. Target Audience and Expectations

Defines target audience and their expectations

d. Methodology & Analysis

Includes both approach and details of data acquisition, exploratory data analysis, testing, machine learnings and other analyses.

e. Results & Discussion

Discusses the results and related issues.

f. Conclusion

Concludes the report.

a. Business Understanding

According to Istanbul's Wikipedia page, with a total population of around fifteen million residents, Istanbul is one of the world's largest cities by population, ranking as the world's fifteenth-largest city and the largest city in Europe. Since it is a city which "its' streets are paved with gold", and prominent in geopolitical and cultural affairs, it is the best city in the country for entrepreneurs to start a business, especially if the business is based on art or cultural products.

Istanbul is the best city in the country to start a business based on art, however, which neighborhood is optimal for a specific venue is a problem remains to be solved. This project aims to solve this problem. We are going to try to find the optimal neighborhood in Istanbul to open a theatre hall, using data science.

Optimal neighborhood would have a lot of other venues in certain categories, in terms of number and variety, so that people would want to come to the neighborhood frequently. We want to enable them to combine going to a play with other activities, eating outside, going to a pub, meeting friends at a coffee house, etc. On the other hand, it is supposed to have no theatre halls at all or might have a few. This way, competition won't be fierce.

b. Data Understanding

Based on the business understanding, we are looking at 2 dimensions here, in order to solve the problem at hand:

- There has to be a lot of other venues in the neighborhood in certain categories, in terms of number and variety.
- Neighborhood is supposed to have no theatre halls at all or might have a few.

Therefore, two sets of analysis will be used together. Both analyses will leverage;

- List of neighborhoods and postal codes from Turkish Postal Service's website: List is maintained in an Excel table on Turkish Postal Service's official website: <https://postakodu.ptt.gov.tr/> (Can be downloaded by clicking on "Posta Kodu Özet Tablosu"). This table contains postal codes, city and neighborhood names for entire Turkey. Data regarding Istanbul will be filtered and used.
- Geographical location data of the neighbourhoods from Geocoder package: This data will be used to create maps. (Afternote: Geocoder package didn't produce reliable results. Due to this fact Geopandas package is used)
- Venue data from Foursquare: Venue data is fundamental for this project, it will enable us to find the optimal neighborhood in Istanbul to open a theatre hall.

c. Target Audience and Expectations

This project is meant to help entrepreneurs to decide which neighborhood is the optimal place to open a theatre hall in Istanbul. Moreover, it is meant to help them to convince potential investors regarding the viability of their future business.

d. Methodology & Analysis

In this project we are going to try to find neighborhoods of Istanbul that have high restaurant, pub and coffee house variety and density, and have no or few theater halls.

In first step we will collect the required data:

- List of neighborhoods and postal codes (From Turkish Postal Service's official website)
- Geodata for neighborhoods (From Arcgis)
- Geodata for venues (From Foursquare)

After data tidying and combining related data together, we will concentrate on finding optimal neighborhoods with following attributes, in accordance with our shared understanding with stakeholders:

- High density of venues in desired categories (Restaurants, pubs and coffee houses)
- High variety of venues in desired categories (Restaurants, pubs and coffee houses)
- No or few theater hall existence

We will present map of all such locations. Moreover, we will create clusters (using k-means clustering) of those locations in order to identify general zones / neighborhoods which should be a starting point for final exploration and search for optimal venue location that has desired attributes.

d.1. Gathering & Tidying Data

d.1.1. List of neighborhoods and postal codes

List of neighborhoods and postal codes is acquired Turkish Postal Service's official website: <https://postakodu.ptt.gov.tr/> (Can be downloaded by clicking on "Posta Kodu Özet Tablosu").

A new database is created and uploaded to Github for the purposes of this project: https://github.com/Omer-Korkmaz/Coursera_Capstone/blob/master/TUR_POSTAL_CODES.csv Headers are translated from Turkish to English and 4th level of detail is eliminated from data (Data included

following levels of details. 1st: City, 2nd: Borough, 3rd: Neighborhood, 4th: sub-neighborhood. Postal codes are given based on 3rd level and 4th level is not useful or related with this project).

This database has data regarding entire Turkey, 2469 postcodes and neighborhoods. First few lines are as following:

	City	Postcode	Borough	Neighborhood
0	ADANA	1720	ALADAG	ALADAG
1	ADANA	1722	ALADAG	MADENLI
2	ADANA	1922	CEYHAN	BUYUKMANGIT
3	ADANA	1920	CEYHAN	CEYHAN
4	ADANA	1924	CEYHAN	MUSTAFABEYLI

Since project requires data regarding Istanbul, a new dataframe is created. An excerpt from is below:

	City	Postcode	Borough	Neighborhood
1099	ISTANBUL	34975	ADALAR	BURGAZADA
1100	ISTANBUL	34970	ADALAR	BUYUKADA
1101	ISTANBUL	34973	ADALAR	HEYBELIADA
1102	ISTANBUL	34977	ADALAR	KINALIADA
1103	ISTANBUL	34275	ARNAVUTKOY	ARNAVUTKOY
1104	ISTANBUL	34277	ARNAVUTKOY	BAKLALI
1105	ISTANBUL	34285	ARNAVUTKOY	BOGAZKOY
1106	ISTANBUL	34287	ARNAVUTKOY	BOLLUCA
1107	ISTANBUL	34555	ARNAVUTKOY	HADIMKOY
1108	ISTANBUL	34281	ARNAVUTKOY	HARACCI
1109	ISTANBUL	34283	ARNAVUTKOY	TASOLUK
1110	ISTANBUL	34758	ATASEHIR	ATATURK
1111	ISTANBUL	34888	ATASEHIR	FERHATPASA
1112	ISTANBUL	34752	ATASEHIR	ICERENKOY
1113	ISTANBUL	34755	ATASEHIR	KAYISDAGI
1114	ISTANBUL	34750	ATASEHIR	KUCUKBAKKALKOY
1115	ISTANBUL	34707	ATASEHIR	MUSTAFAKEMAL
1116	ISTANBUL	34704	ATASEHIR	ORNEK
1117	ISTANBUL	34779	ATASEHIR	YENICAMLICA
1118	ISTANBUL	34746	ATASEHIR	YENISAHRA

d.1.2. Geodata for neighborhoods

Using Geopy package, we establish that geographical coordinate of Istanbul are 41.0096334, 28.9651646.

Thereafter, using Geopandas package, we create a dataframe, providing coordinates for each postcode:

	geometry	address	Longitude	Latitude	Postcode
1099	POINT (29.06197 40.88023)	34975	29.061974	40.880225	34975
1100	POINT (29.12311 40.85572)	34970	29.123110	40.855719	34970
1101	POINT (29.09094 40.87615)	34973	29.090943	40.876150	34973
1102	POINT (29.04850 40.90841)	34977	29.048496	40.908410	34977
1103	POINT (28.74328 41.20992)	34275	28.743275	41.209924	34275
...
1360	POINT (29.08501 41.04300)	34690	29.085005	41.043003	34690
1361	POINT (28.89779 40.99265)	34025	28.897785	40.992650	34025
1362	POINT (28.91192 41.01077)	34015	28.911920	41.010765	34015
1363	POINT (28.90521 40.99211)	34020	28.905205	40.992106	34020
1364	POINT (28.91145 41.01956)	34010	28.911448	41.019560	34010

This alone can not be sufficient for our analyses, as a consequence, we merge it with neighborhood&postcode dataframe. Following is the result table:

	Postcode	Borough	Neighborhood	Adress	geometry	address	Longitude	Latitude
0	34975	ADALAR	BURGAZADA	34975, ISTANBUL	POINT (29.06197 40.88023)	34975	29.061974	40.880225
1	34970	ADALAR	BUYUKADA	34970, ISTANBUL	POINT (29.12311 40.85572)	34970	29.123110	40.855719
2	34973	ADALAR	HEYBELIADA	34973, ISTANBUL	POINT (29.09094 40.87615)	34973	29.090943	40.876150
3	34977	ADALAR	KINALIADA	34977, ISTANBUL	POINT (29.04850 40.90841)	34977	29.048496	40.908410
4	34275	ARNAVUTKOY	ARNAVUTKOY	34275, ISTANBUL	POINT (28.74328 41.20992)	34275	28.743275	41.209924
...
261	34690	USKUDAR	YAVUZTURK	34690, ISTANBUL	POINT (29.08501 41.04300)	34690	29.085005	41.043003
262	34025	ZEYTINBURNU	CIRPICI	34025, ISTANBUL	POINT (28.89779 40.99265)	34025	28.897785	40.992650
263	34015	ZEYTINBURNU	SEYITNIZAM	34015, ISTANBUL	POINT (28.91192 41.01077)	34015	28.911920	41.010765
264	34020	ZEYTINBURNU	TELSIZ	34020, ISTANBUL	POINT (28.90521 40.99211)	34020	28.905205	40.992106
265	34010	ZEYTINBURNU	TOPKAPI	34010, ISTANBUL	POINT (28.91145 41.01956)	34010	28.911448	41.019560

d.1.3. Geodata for venues

Geodata for venues is gathered from Foursquare.

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	BURGAZADA	40.880225	29.061974	Burgazada Kalpazan Kayaları	40.880424	29.065728	Campground
1	BURGAZADA	40.880225	29.061974	Sait Faik Abasıyanık Müzesi	40.881015	29.067458	History Museum
2	BURGAZADA	40.880225	29.061974	Peyote Cennet Bahçesi	40.884498	29.063283	Beer Garden
3	BURGAZADA	40.880225	29.061974	Cennet Bahçesi/ Burgaz Ada	40.884500	29.063331	Garden
4	BURGAZADA	40.880225	29.061974	Bayraktepe	40.877576	29.063138	Scenic Lookout

We used this data along with geodata for neighborhoods.

d.2. Exploratory Data Analysis

d.2.1. General Analysis

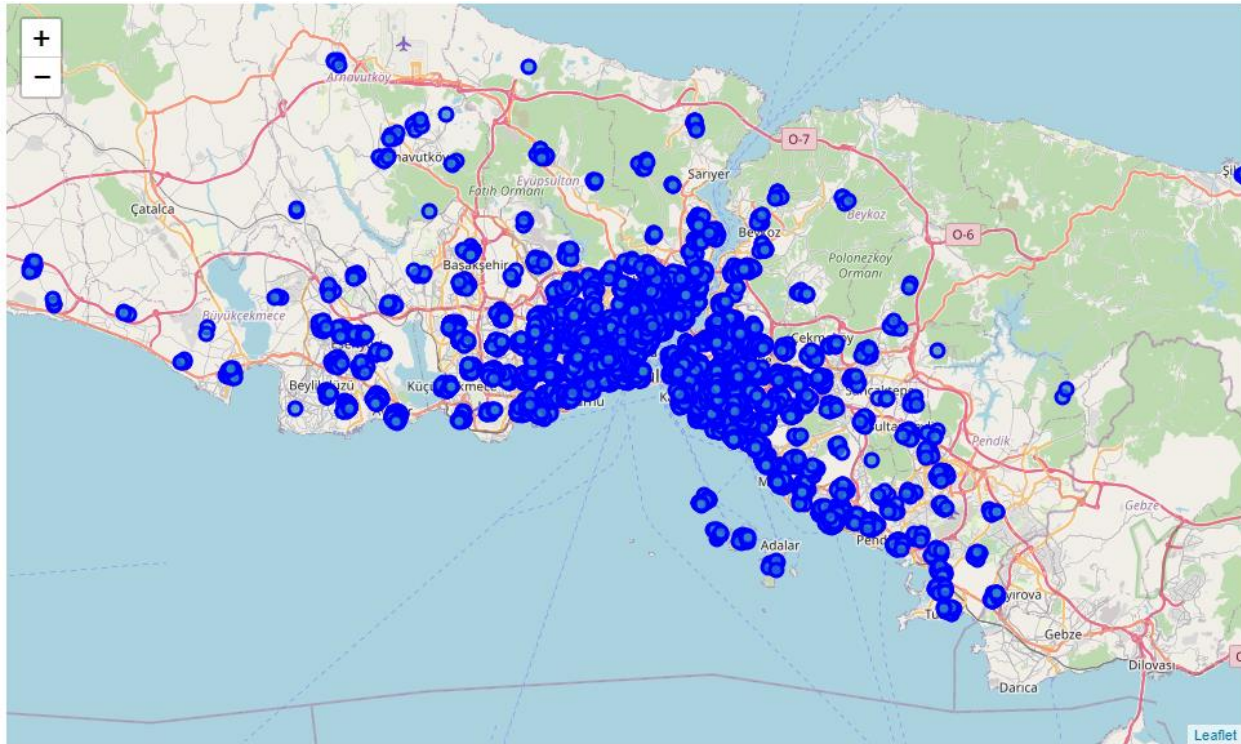
We created a table a grouped count to see density of venues in each neighborhood. This gave us an understanding about popular neighborhoods:

	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Neighborhood						
ABBASAGA	100	100	100	100	100	100
ACIBADEM	39	39	39	39	39	39
AKATLAR	57	57	57	57	57	57
AKINCILAR	83	83	83	83	83	83
AKSARAY	40	40	40	40	40	40
...
ZEYREK	59	59	59	59	59	59
ZEYTINLIK	100	100	100	100	100	100
ZUBEYDEHANIM	18	18	18	18	18	18
ZUHURATBABA	11	11	11	11	11	11
ZUMRUTEVLER	5	5	5	5	5	5

For further analysis, we turn to absolute values, checked out how many venues are there in each neighborhood. Top 30 as is following:

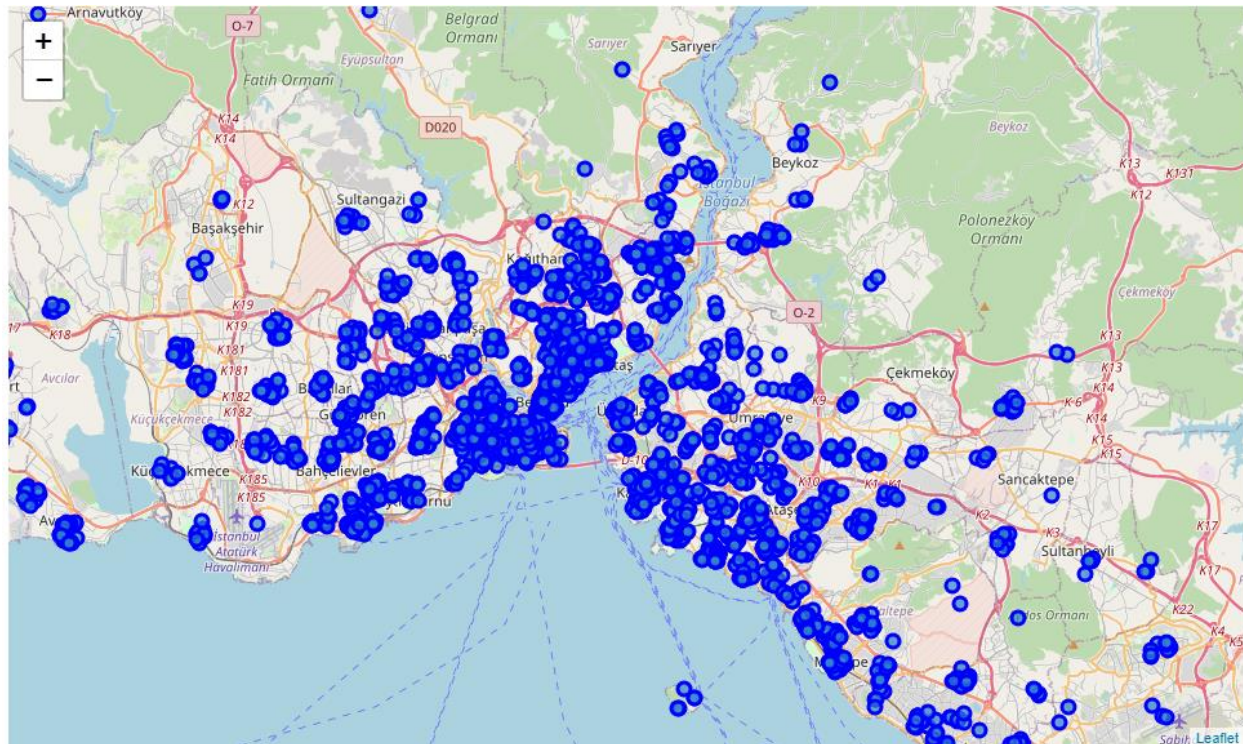
Neighborhood	Count	Neighborhood	Count	Neighborhood	Count
0 MIMARSINAN	157	10 TARLABASI	100	20 SIRKECI	100
1 BAHCELIEVLER	157	11 HALASKARGAZI	100	21 ZEYTINLIK	100
2 CUMHURİYET	132	12 TAKSIM	100	22 SAHRAYICEDIT	100
3 TOPKAPI	115	13 HARBIYE	100	23 YALI	100
4 ISTIKLAL	100	14 TAHTAKALE	100	24 CIHANGIR	100
5 TESVİKİYE	100	15 FENERBAHCE	100	25 KEMANKES	100
6 BAYAZIT	100	16 ABBASAGA	100	26 ARAPCAMI	100
7 OSMANAGA	100	17 SEHREMINI	100	27 CAFERAGA	100
8 EMINONU	100	18 SUADIYE	100	28 CADDEBOSTAN	98
9 TURKALI	100	19 MAHMUTPASA	100	29 KUCUKBAKKALKOY	97
10 TARLABASI	100				

Istanbul has a lot of neighborhoods with a great number of venues. Seeing them on the map might provide a picture:



Summary table and the concentration on the map above provides us with neighborhoods that have highest number of venues. At this point, it is important to remember the problem to solve. Even though neighborhoods that have highest number of venues provides an idea regarding popular neighborhoods, according to business understanding, it is assumed that only coffee houses, restaurants, bars and pubs are important for analysis. Following table and map help us understand the distribution of venues in these categories:

	Neighborhood	Count		Neighborhood	Count
0	ABBASAGA	62	16	TURKALI	44
1	CAFERAGA	59	17	SEHREMİNİ	43
2	TOPKAPI	58	18	BAHCELİEVLER	42
3	MİMARŞİNAN	58	19	İSKENDERPAŞA	42
4	OSMANAGA	56	20	ÇİHANGİR	42
5	YALI	54	21	SUADIYE	42
6	ARAPCAMI	54	22	TARLABASI	41
7	KEMANKES	53	23	BALAT	41
8	CUMHURİYET	51	24	HALASKARGAZI	41
9	İSTİKLAL	50	25	VEFA	41
10	TESVİKİYE	48	26	BOZKURT	39
11	ZEYTLİK	47	27	ESENTEPE	39
12	CADDEBOSTAN	46	28	BAYAZIT	39
13	AMBARLI	45	29	KARTAL	39
14	FENERBAHCE	44			
15	TAKSİM	44			



For further analysis, we take mean of the frequency of occurrence of each category. This enables us to see if desired venues in desired categories are popular in neighborhoods. For instance, «AKSARAY» and «ZEYTINLIK» have high occurrences for «Bar».

	Neighborhood	Afghan Restaurant	African Restaurant	American Restaurant	Arepa Restaurant	Argentinian Restaurant	Asian Restaurant	Austrian Restaurant	Bar	Beach Bar	...	Tantuni Restaurant	Thai Restaurant	f
0	ABBASAGA	0.0	0.0	0.00000	0.0	0.000000	0.000000	0.0	0.048387	0.0	...	0.0	0.0	
1	ACIBADEM	0.0	0.0	0.00000	0.0	0.000000	0.000000	0.0	0.000000	0.0	...	0.0	0.0	
2	AKATLAR	0.0	0.0	0.00000	0.0	0.000000	0.000000	0.0	0.000000	0.0	...	0.0	0.0	
3	AKINCILAR	0.0	0.0	0.03125	0.0	0.000000	0.000000	0.0	0.000000	0.0	...	0.0	0.0	
4	AKSARAY	0.0	0.0	0.00000	0.0	0.071429	0.071429	0.0	0.071429	0.0	...	0.0	0.0	
...	
212	ZEYREK	0.0	0.0	0.00000	0.0	0.000000	0.000000	0.0	0.000000	0.0	...	0.0	0.0	
213	ZEYTINLIK	0.0	0.0	0.00000	0.0	0.000000	0.000000	0.0	0.042553	0.0	...	0.0	0.0	
214	ZUBEYDEHANIM	0.0	0.0	0.00000	0.0	0.000000	0.000000	0.0	0.000000	0.0	...	0.0	0.0	
215	ZUHURATBABA	0.0	0.0	0.00000	0.0	0.000000	0.000000	0.0	0.000000	0.0	...	0.0	0.0	
216	ZUMRUTEVLER	0.0	0.0	0.00000	0.0	0.000000	0.000000	0.0	0.000000	0.0	...	0.0	0.0	

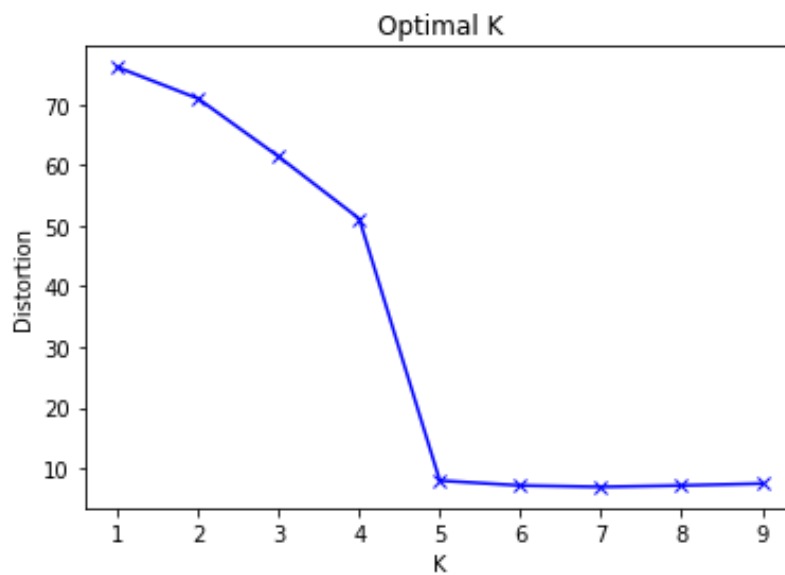
Seeing top 10 venues in each neighborhood might help us understand social landscape of each neighborhood. For example In «ZUHURATBABA» bars and ethnical restaurants are at the top.

	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	ABBASAGA	Coffee Shop	Café	Pub	Turkish Restaurant	Bar	Hookah Bar	Restaurant	Falafel Restaurant	Fondue Restaurant	Seafood Restaurant
1	ACIBADEM	Café	Coffee Shop	Restaurant	Italian Restaurant	Turkish Restaurant	Kebab Restaurant	Eastern European Restaurant	Ethiopian Restaurant	Falafel Restaurant	Fast Food Restaurant
2	AKATLAR	Café	Coffee Shop	Italian Restaurant	Kebab Restaurant	Modern European Restaurant	Sushi Restaurant	Turkish Restaurant	Hookah Bar	Filipino Restaurant	Eastern European Restaurant
3	AKINCILAR	Café	Turkish Restaurant	Hookah Bar	Restaurant	Comfort Food Restaurant	American Restaurant	Coffee Shop	Middle Eastern Restaurant	Doner Restaurant	Fast Food Restaurant
4	AKSARAY	Café	Turkish Restaurant	Restaurant	Doner Restaurant	Argentinian Restaurant	Asian Restaurant	Bar	French Restaurant	Eastern European Restaurant	Ethiopian Restaurant
...
212	ZEYREK	Café	Turkish Restaurant	Coffee Shop	Restaurant	Fast Food Restaurant	Wine Bar	Fondue Restaurant	Eastern European Restaurant	Ethiopian Restaurant	Falafel Restaurant
213	ZEYTINLIK	Café	Coffee Shop	Restaurant	Turkish Restaurant	Seafood Restaurant	Fast Food Restaurant	Bar	Salon / Barbershop	Kebab Restaurant	Sushi Restaurant
214	ZUBEYDEHANIM	Turkish Restaurant	Restaurant	Kebab Restaurant	Wine Bar	Fondue Restaurant	Dumpling Restaurant	Eastern European Restaurant	Ethiopian Restaurant	Falafel Restaurant	Fast Food Restaurant
215	ZUHURATBABA	Cocktail Bar	Café	Wine Bar	French Restaurant	Eastern European Restaurant	Ethiopian Restaurant	Falafel Restaurant	Fast Food Restaurant	Filipino Restaurant	Fondue Restaurant
216	ZUMRUTEVLER	Turkish Restaurant	Café	Wine Bar	French Restaurant	Dumpling Restaurant	Eastern European Restaurant	Ethiopian Restaurant	Falafel Restaurant	Fast Food Restaurant	Filipino Restaurant

d.3. Clustering

Created clusters (using k-means clustering) of neighborhoods in order to identify general zones / neighborhoods which should be a starting point for final exploration.

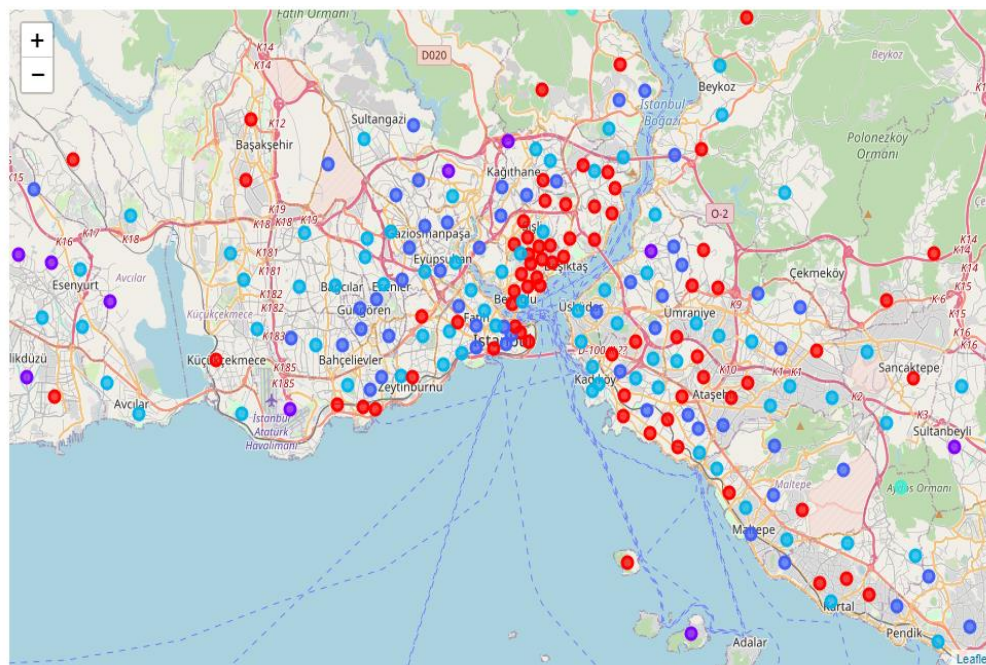
Using elbow method, we establish 5 is the optimal K. We are going to see which cluster is optimal in terms of venue variety on the following page.



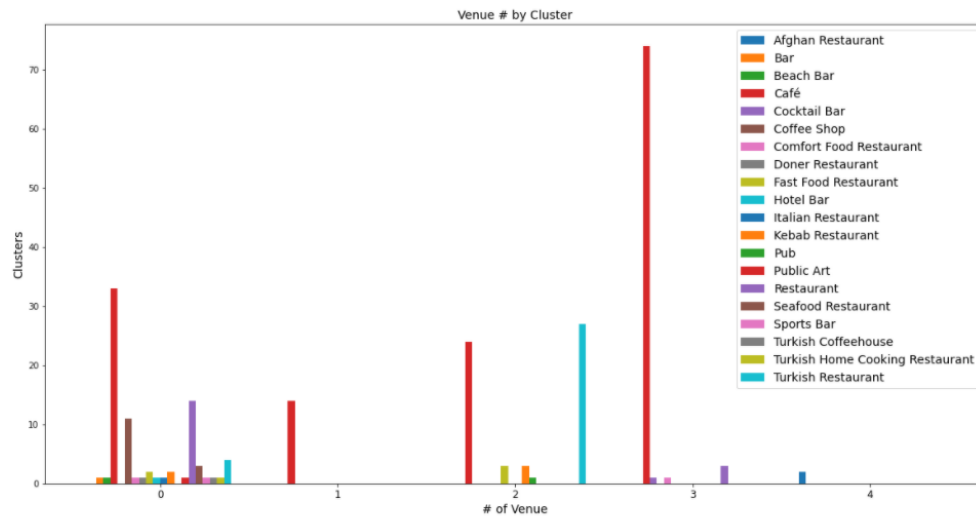
	Postcode	Borough	Neighborhood	Latitude	Longitude	Clusters	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue
1	34970	ADALAR	BUYUKADA	40.855719	29.123110	3.0	Café	Restaurant	Wine Bar	Fondue Restaurant	Dumpling Restaurant	Eastern European Restaurant	Ethiopian Restaurant
2	34973	ADALAR	HEYBELIADA	40.876150	29.090943	1.0	Café	Wine Bar	Doner Restaurant	Eastern European Restaurant	Ethiopian Restaurant	Falafel Restaurant	Fast Food Restaurant
3	34977	ADALAR	KINALIADA	40.908410	29.048496	0.0	Beach Bar	Café	Kebab Restaurant	Seafood Restaurant	Wine Bar	Eastern European Restaurant	Ethiopian Restaurant
4	34275	ARNAVUTKOY	ARNAVUTKOY	41.209924	28.743275	0.0	Seafood Restaurant	Restaurant	Café	Coffee Shop	Cocktail Bar	Turkish Restaurant	Hotel Bar
9	34281	ARNAVUTKOY	HARACCI	41.184955	28.707398	1.0	Café	Wine Bar	Doner Restaurant	Eastern European Restaurant	Ethiopian Restaurant	Falafel Restaurant	Fast Food Restaurant
...
261	34690	USKUDAR	YAVUZTURK	41.043003	29.085005	2.0	Turkish Restaurant	Coffee Shop	Wine Bar	French Restaurant	Dumpling Restaurant	Eastern European Restaurant	Ethiopian Restaurant
262	34025	ZEYTINBURNU	CIRPICI	40.992650	28.897785	3.0	Café	Restaurant	Turkish Restaurant	Dumpling Restaurant	Coffee Shop	Fast Food Restaurant	Fast Food Restaurant
263	34015	ZEYTINBURNU	SEYITNIZAM	41.010765	28.911920	3.0	Café	Restaurant	Turkish Restaurant	Kebab Restaurant	Indonesian Restaurant	Turkish Home Cooking Restaurant	Eastern European Restaurant
264	34020	ZEYTINBURNU	TELSIZ	40.992106	28.905205	0.0	Turkish Restaurant	Restaurant	Coffee Shop	Fast Food Restaurant	Dumpling Restaurant	Hotel Bar	Hotel Bar
265	34010	ZEYTINBURNU	TOPKAPI	41.019560	28.911448	0.0	Café	Comfort Food Restaurant	Coffee Shop	Turkish Restaurant	Restaurant	Pub	Fast Food Restaurant

We made an assumption and included certain venue categories in our analysis. Consequently, some neighborhoods did not fall under any cluster. Since we do not have any interest over these neighborhoods, they are not included on the table above.

Seeing clustered venues on the map gives a more clear picture. Cluster 0: Red, 1: Purple, 2: Blue, 3: Light Blue, 4: Green



We can see distribution of venues in clusters so that we can make a decision regarding which cluster fits our purpose best:



Above figure shows us that, in terms of variety of venue categories, cluster 0 has optimal neighborhoods.

In accordance with business needs, we are going to focus on cluster 0, i.e. red dots on the map.

d.4. Neighborhoods and Theatre Hall Analysis

Until this point, we established neighborhoods in cluster 0 are most diverse in terms of venue category. Moreover, we listed neighborhoods with highest number of venues. Neighborhoods in the intersection set of these two valuable deductions will have our eyes on them. Number of theaters in these neighborhoods supply the missing piece of this puzzle.

	Neighborhood	No. of Related Venues	No. of Theatre Halls
0	ABBASAGA	62	1.0
1	TOPKAPI	58	0.0
2	ARAPCAMI	54	2.0
3	CUMHURIYET	51	0.0
4	ISTIKLAL	50	1.0
5	TESVIKIYE	48	0.0
6	ZEYTINLIK	47	0.0
7	CADDEBOSTAN	46	0.0
8	TURKALI	44	0.0
10	FENERBAHCE	44	0.0
9	TAKSIM	44	1.0
12	SUADIYE	42	1.0
11	CIHANGIR	42	2.0
14	HALASKARGAZI	41	1.0
13	TARLABASI	41	3.0
15	BOZKURT	39	1.0
16	ESENTEPE	39	1.0
17	HARBIYE	38	1.0
18	IDEALTEPE	37	0.0
19	SIRKECI	36	0.0

e. Results & Discussion

Our analysis revealed that, even though there are great number of venues in Istanbul, there are certain clusters of neighborhoods that stands out, in terms of variety of venues and among them, there are certain neighborhoods that stands out in terms of number of venues.

We focused on venues in certain categories, restaurants, pubs and coffee houses. Based on these venues, we created clusters of neighborhoods and saw that, there are differences between zones, in terms of venue variety. After this point, in accordance with our shared business understanding with stakeholders, we concentrated on cluster 0 which has highest variety of venues in our desired venue categories. Thereafter, we took 20 most dense neighborhoods in terms of venue number in cluster 0 and looked for theater halls.

Result of all this is a table with 20 neighborhoods, that have:

- High density of venues in desired categories (Restaurants, pubs and coffee houses)
- High variety of venues in desired categories (Restaurants, pubs and coffee houses)
- No or few theater hall existence

f. Conclusion

This project is aimed at finding an optimal neighborhood in Istanbul to open a theater hall.

As a result of our analyses, we identified 20 neighborhoods, that have high density of venues in desired categories, high variety of venues in desired categories and no or few theater hall existence in them. This information shed a light on one aspect of a complicated problem.

Final decision regarding optimal neighborhood must be based on several more analyses alongside this one, i.e. on rent levels, levels of income, education levels and other demographical aspects of neighborhoods. Then stakeholders can make an informed decision.