Identifying Optimal Neighborhood to Open up a Theatre Hall in Istanbul

Identifying which neighborhood is optimal for a specific type of venue is crucial for the success of a future business.

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.

It is prominent in geopolitical and cultural affairs, therefore 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.

Identifying which neighborhood is optimal for a specific type of venue is crucial for the success of a future business.

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 (Restaurants, ğpubs and coffee houses), in terms of number and variety, so that people would want to come to the neighborhood frequently. 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.

Data Understanding

Data will be acquired from following sources:

- 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

Methodology

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.

Istanbul has a lot of neighborhoods with a great number of venues. Seeing them on the map and checking out the top 30's venue count (30 out of 266) might provide a picture:
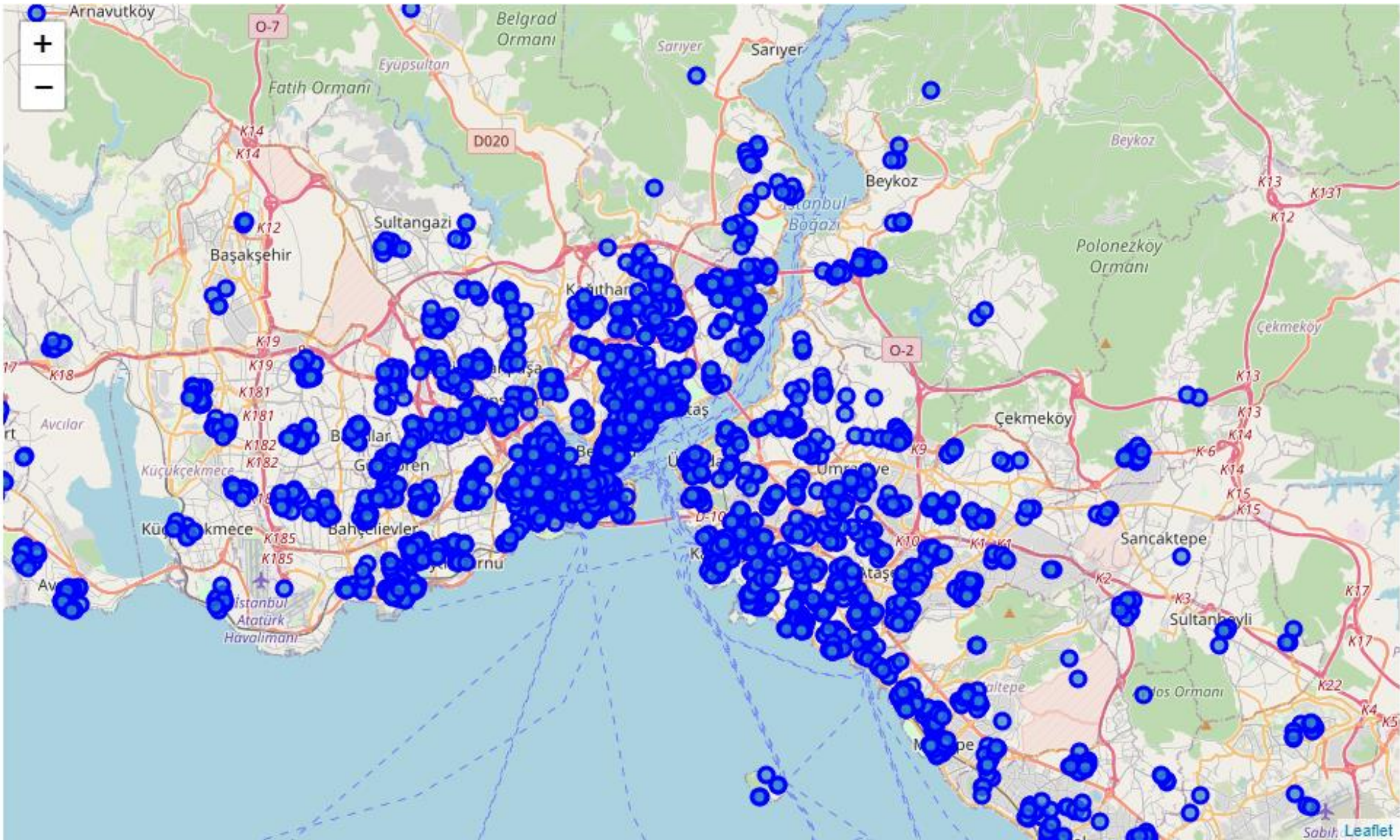
	Neighborhood	Count
0	MIMARSINAN	157
1	BAHCELIEVLER	157
2	CUMHURIYET	132
3	TOPKAPI	115
4	ISTIKLAL	100
5	TESVIKIYE	100
6	BAYAZIT	100
7	OSMANAGA	100
8	EMINONU	100
9	TURKALI	100
10	TARLABASI	100
11	HALASKARGAZI	100
12	TAKSIM	100
13	HARBIYE	100
14	TAHTAKALE	100
15	FENERBAHCE	100

	Neighborhood	Count
16	ABBASAGA	100
17	SEHREMINI	100
18	SUADIYE	100
19	MAHMUTPASA	100
20	SIRKECI	100
21	ZEYTINLIK	100
22	SAHRAYICEDIT	100
23	YALI	100
24	CIHANGIR	100
25	KEMANKES	100
26	ARAPCAMI	100
27	CAFERAGA	100
28	CADDEBOSTAN	98
29	KUCUKBAKKALKOY	97
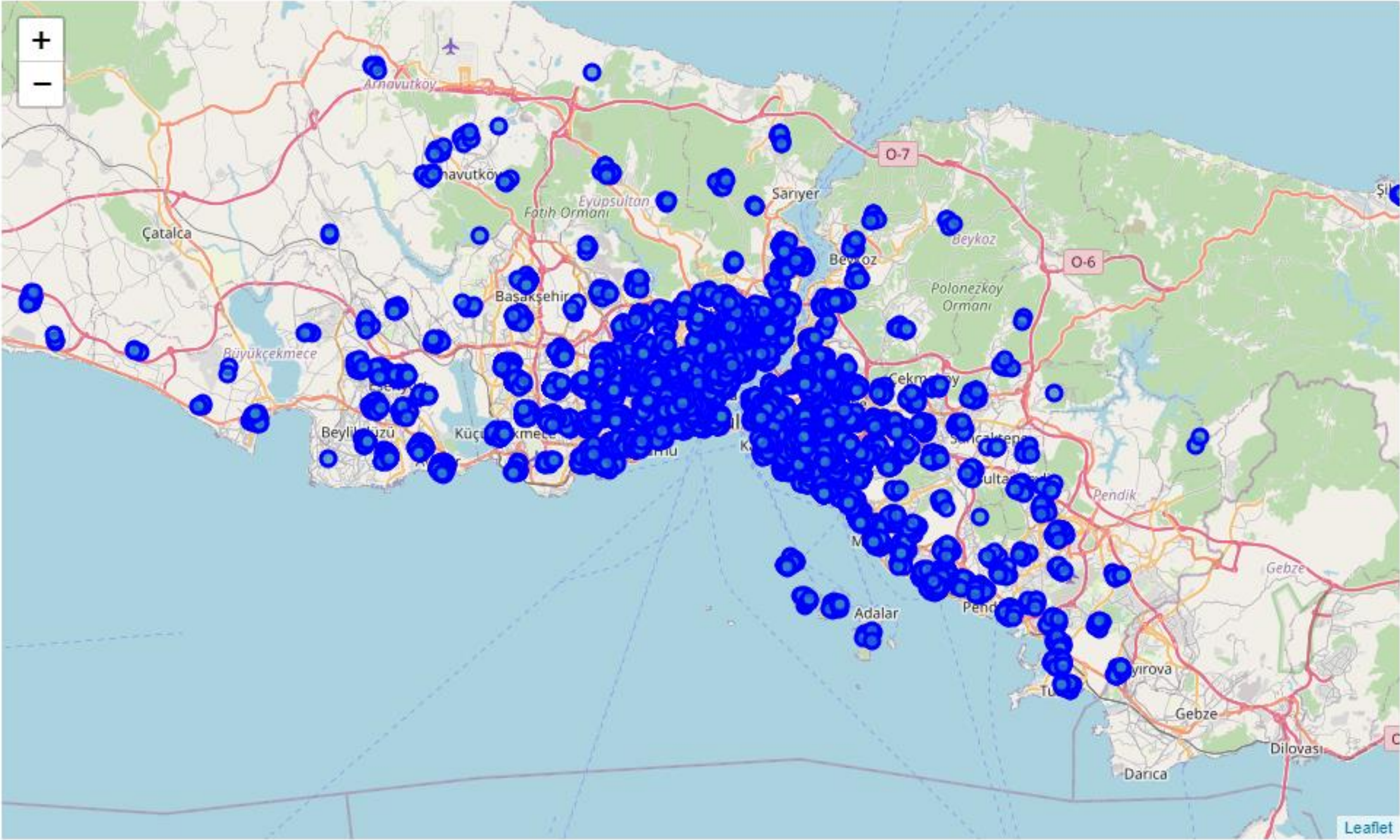
It is important to focus on the issue: concentrating on the related categories provides a narrow land to cover.

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	SEHREMINI	43
2	TOPKAPI	58		18	BAHCELIEVLER	42
3	MIMARSINAN	58		19	ISKENDERPASA	42
4	OSMANAGA	56		20	CIHANGIR	42
5	YALI	54		21	SUADIYE	42
6	ARAPCAMI	54		22	TARLABASI	41
7	KEMANKES	53		23	BALAT	41
8	CUMHURIYET	51		24	HALASKARGAZI	41
9	ISTIKLAL	50		25	VEFA	41
10	TESVIKIYE	48		26	BOZKURT	39
11	ZEYTINLIK	47		27	ESENTEPE	39
12	CADDEBOSTAN	46		28	BAYAZIT	39
13	AMBARLI	45		29	KARTAL	39
14	FENERBAHCE	44				
15	TAKSIM	44				

Looking for further cues regarding neighborhoods…

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 occurances for «Bar».

	Neighborhood	Afghan Restaurant	African Restaurant	American Restaurant	Arepa Restaurant	Argentinian Restaurant	Asian Restaurant	Austrian Restaurant	Bar	Beach Bar	...	Tantuni Restaurant	Thai Restaurant
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

217 rows × 81 columns

	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

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

Created clusters (using k-means clustering) of neighborhoods in order to identify general zones / neighborhoods which should be a starting point for final exploration
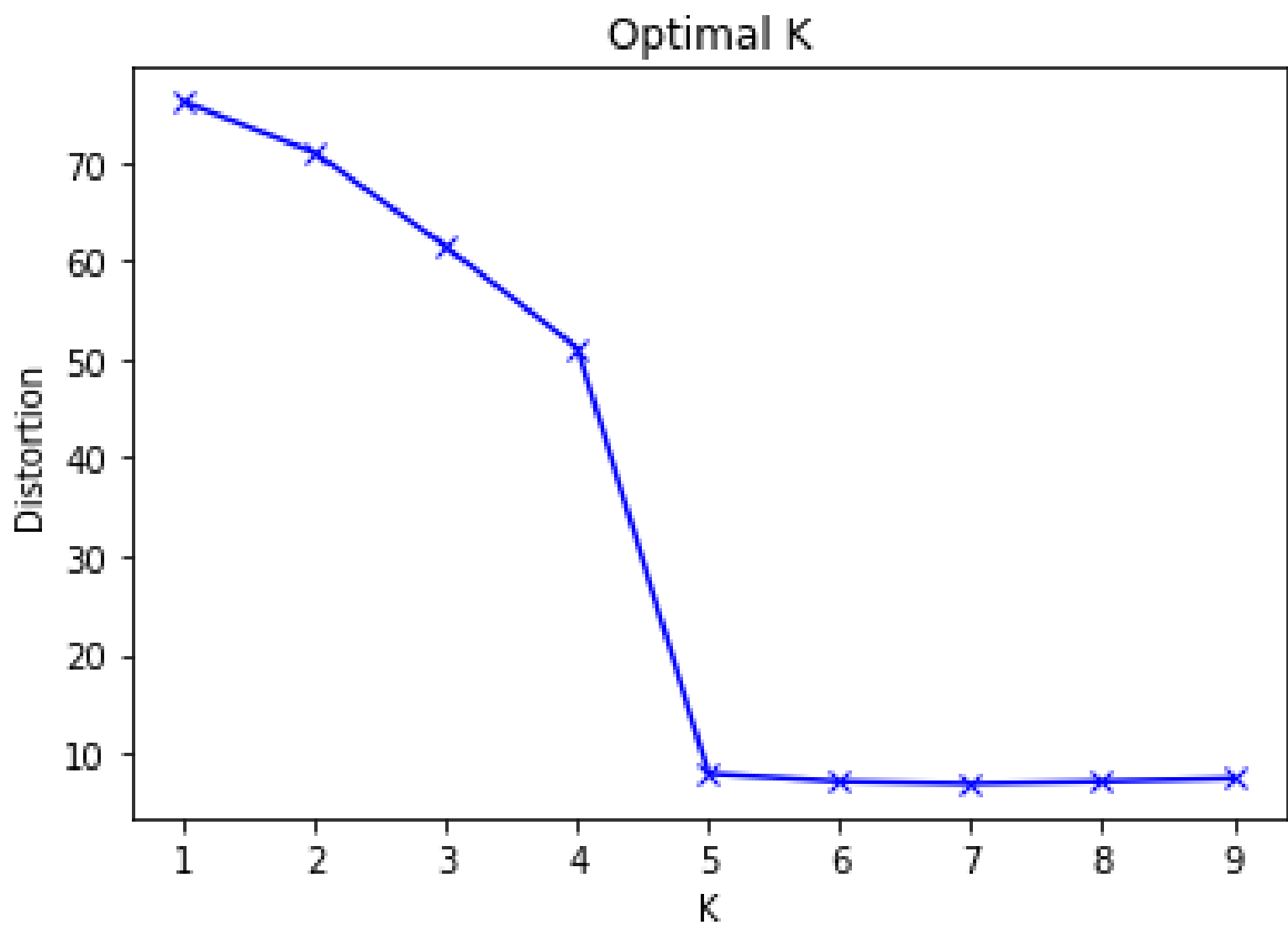
	Postcode	Borough	Neighborhood	Latitude	Longitude	Clusters	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue
1	34970	ADALAR	BUYUKADA	40.855719	29.123110	3.0	Café	Restaurant	Wine Bar	Fondue Restaurant	Dumpling Restaurant	Eastern European Restaurant	Ethiop Restaur
2	34973	ADALAR	HEYBELIADA	40.876150	29.090943	1.0	Café	Wine Bar	Doner Restaurant	Eastern European Restaurant	Ethiopian Restaurant	Falafel Restaurant	Fast Fo Restaur
3	34977	ADALAR	KINALIADA	40.908410	29.048496	0.0	Beach Bar	Café	Kebab Restaurant	Seafood Restaurant	Wine Bar	Eastern European Restaurant	Ethiop Restaur
4	34275	ARNAVUTKOY	ARNAVUTKOY	41.209924	28.743275	0.0	Seafood Restaurant	Restaurant	Café	Coffee Shop	Cocktail Bar	Turkish Restaurant	Hool
9	34281	ARNAVUTKOY	HARACCI	41.184955	28.707398	1.0	Café	Wine Bar	Doner Restaurant	Eastern European Restaurant	Ethiopian Restaurant	Falafel Restaurant	Fast Fo Restaur
...	...	...	...	...	...	...	...	...	...	...	...	...	...
261	34690	USKUDAR	YAVUZTURK	41.043003	29.085005	2.0	Turkish Restaurant	Coffee Shop	Wine Bar	French Restaurant	Dumpling Restaurant	Eastern European Restaurant	Ethiop Restaur
262	34025	ZEYTINBURNU	CIRPICI	40.992650	28.897785	3.0	Café	Restaurant	Turkish Restaurant	Dumpling Restaurant	Coffee Shop	Fast Food Restaurant	Fon Restaur
263	34015	ZEYTINBURNU	SEYITNIZAM	41.010765	28.911920	3.0	Café	Restaurant	Turkish Restaurant	Kebab Restaurant	Indonesian Restaurant	Turkish Home Cooking Restaurant	East Europe Restaur
264	34020	ZEYTINBURNU	TELSIZ	40.992106	28.905205	0.0	Turkish Restaurant	Restaurant	Coffee Shop	Fast Food Restaurant	Dumpling Restaurant	Hotel Bar	Hool
265	34010	ZEYTINBURNU	TOPKAPI	41.019560	28.911448	0.0	Café	Comfort Food Restaurant	Coffee Shop	Turkish Restaurant	Restaurant	Pub	Fast Fo Restaur

Using K means clustering, we cluster neighborhoods based on venues. 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.

Identifying neighborhoods that are rich in terms of venue variety and number of venues

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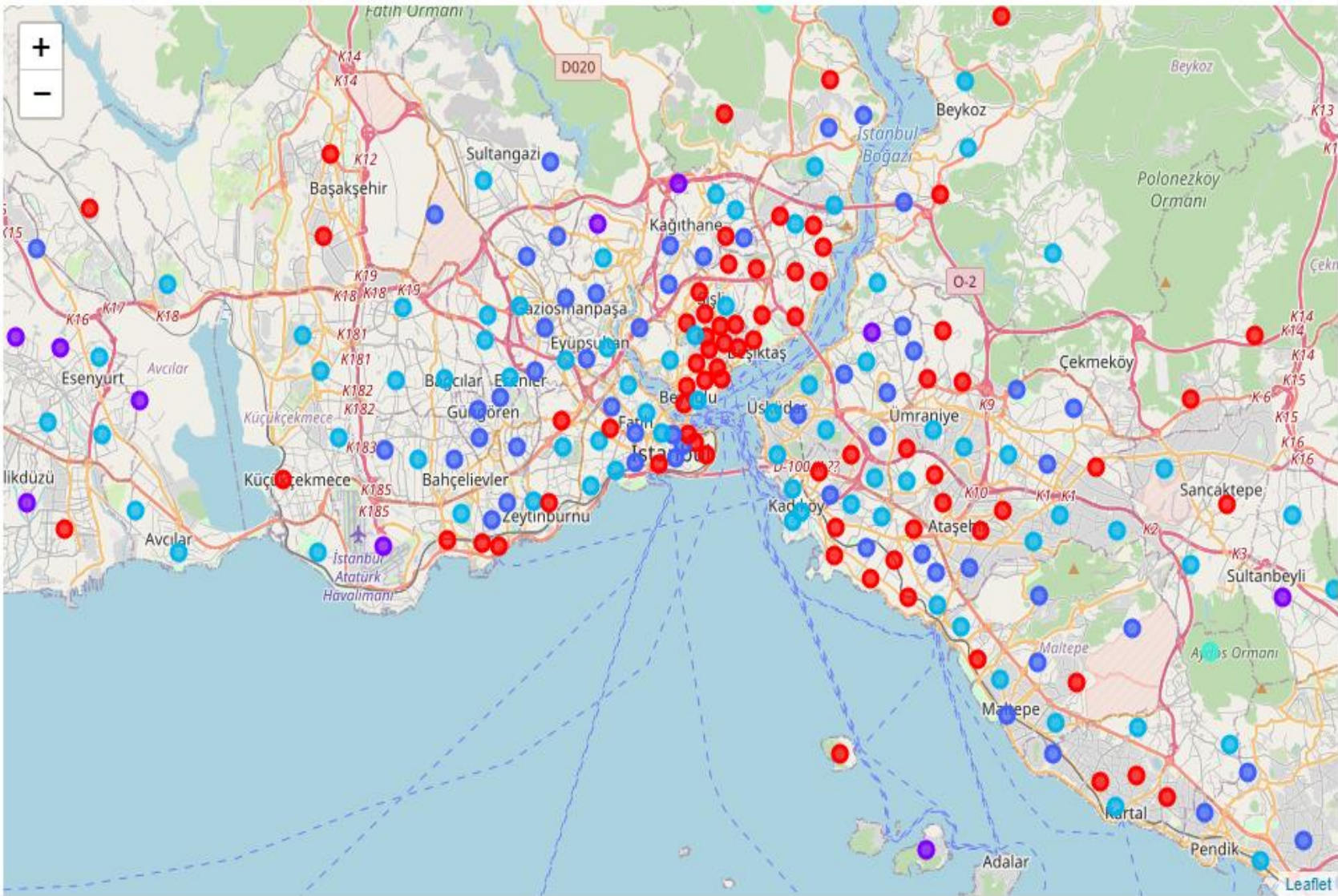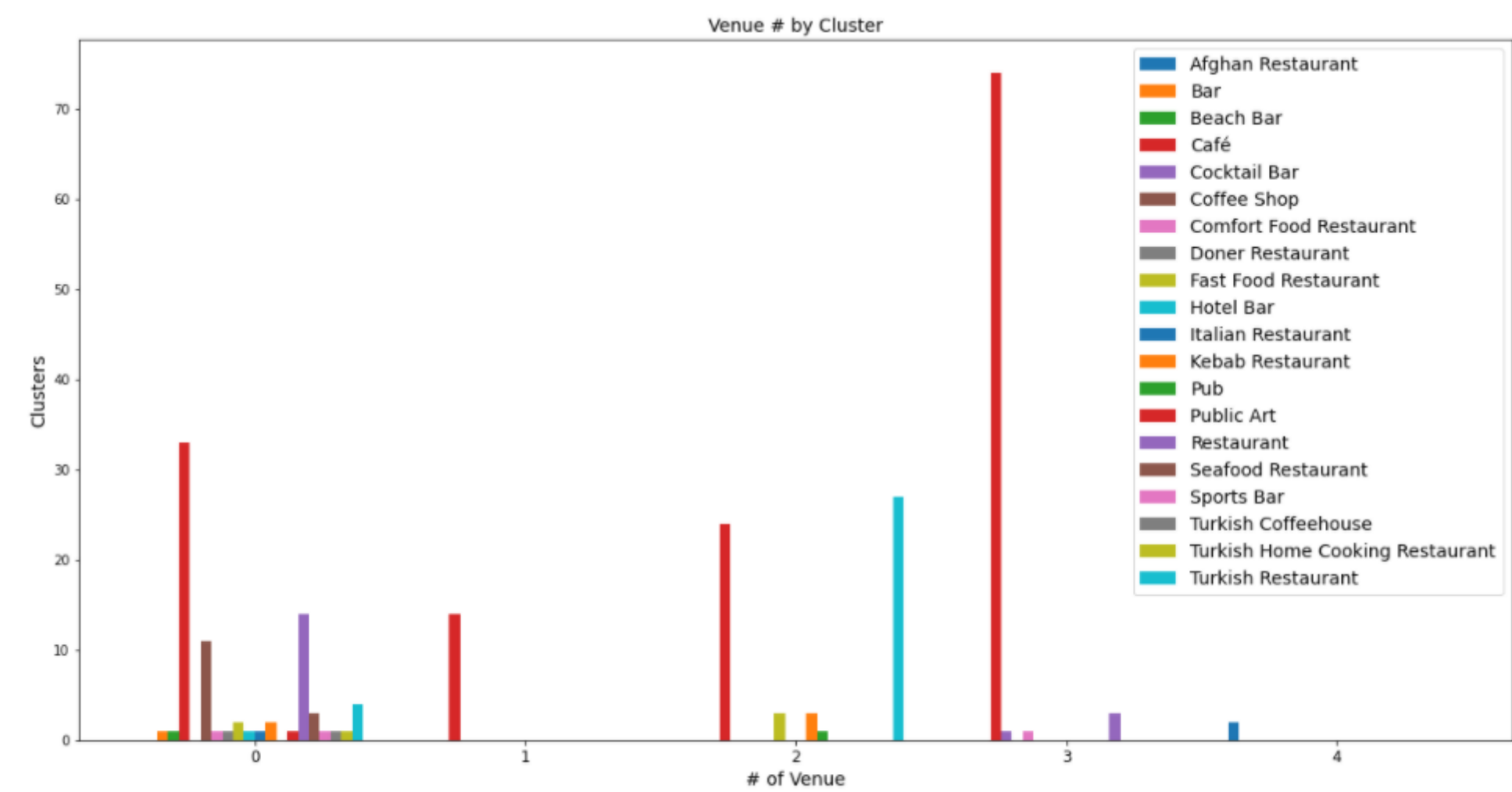
Above figure shows us that, in terms of variety of venue categories, cluster 0 has optimal neighborhoods.

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

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

At last, a table of neighborhoods full of possibilites…

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. Neigborhoods 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

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.

THANK YOU!