Crime Rates Data & Venues Data Analysis of Istanbul

Description & Discussion of the Background

We are a research company that prepares criminological reports to private customers, governments. universities or small and medium sized enterprises (SMEs) in Istanbul, Turkey. One of our customers, Tourism Restaurant Investors and Operators Association approached us for consultancy services. Basically, they request from us to prepare a report on the following topics:

- For a resident, which area has both the safest place and highly dense social places
- For the potential investors, where to open up new venues which are crime free
- For city governance. where to allocate resources efficiently.

PROBLEM:LOCATION

Istanbul is a city with a high population and population density.

Being such a crowded city leads the owners of shops and social sharing places in the city where the population is dense.

- Finding hotspots gains importance to solve the crime problem as well as urban planning.
- Finding a safe location to open up a social venue, which will also have enough customers, requires an extensive research.
- For a resident, which area has both the safest place and highly dense social places
- For the potential investors, where to open up new venues which are crime free
- For city governance. where to allocate resources efficiently.

To shed light on this, a map and information chart was formed on which the crime rate is placed based on the districts of Istanbul and each district is clustered according to the venue density.

Data Description

- The data regarding the coordinates of the city and its districts were obtained from (https://github.com/Srcanyildiz/istanbul/blob/master/istanbul_geo.csv)
- The population data for each district was scraped from Wikipedia (wikipedia.com).
- The crime data of Istanbul per district was collected from an article, published in a journal (https://www.researchgate.net/publication/232913593_Distribution_of_Crime_Rates_in_Different_Districts_in_Istanbul).
    - The crime rate was obtained by dividing the number of total crime to the population of each district.
- I used Foursquare API to get the most common venues of given district of Istanbul .

By using Folium and Foursquare API , a map of Istanbul with boroughs was created and segmented.

get the borough's latitude and longitude values.

To get the information from foursquare api, create get request for url by designing the limit as 100 venues and the radius 750 meter for each district.

clean the json and structure it into a pandas dataframe

[14]:
```
venues = results['response']['groups'][0]['items']

nearby_venues = json_normalize(venues) # flatten JSON

# filter columns
filtered_columns = ['venue.name', 'venue.categories', 'venue.location.lat', 'venue.location.lng']
nearby_venues =nearby_venues.loc[:, filtered_columns]

# filter the category for each row
nearby_venues['venue.categories'] = nearby_venues.apply(get_category_type, axis=1)

# clean columns
nearby_venues.columns = [col.split(".")[-1] for col in nearby_venues.columns]

nearby_venues.head()
```

[14]:

	name	categories	lat	lng
0	Büyükada Tepesi	Mountain	40.861107	29.117418
1	Eski Rum Yetimhanesi	Historic Site	40.861705	29.123323
2	Aşıklar butik by şükrü	Hotel	40.862570	29.118003
3	Büyükada Bisiklet Parkuru	Bike Trail	40.865000	29.116861
4	Büyükada Lale köşkü	Bed & Breakfast	40.865657	29.125223

The number of venues is returned by Foursquare=41

[15]:
```
print('{} venues were returned by Foursquare.'.format(nearby_venues.shape[0]))
```

```
41 venues were returned by Foursquare.
```

Exploring Boroughs in Istanbul

For each district, istanbul_venues was created by using Foursquare API.

```
[18]: print(istanbul_venues.shape)
      istanbul_venues.head()
```

(1575, 7)

[18]:

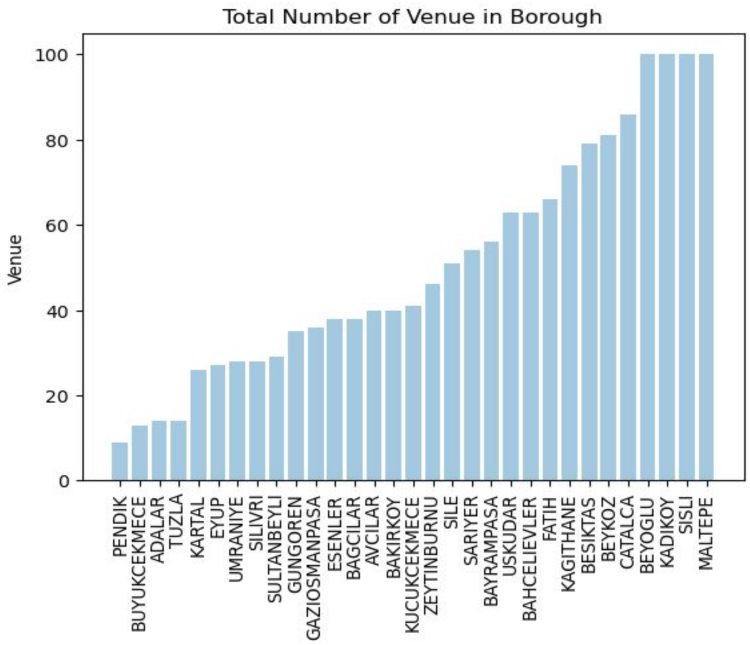
	Borough	Borough Latitude	Borough Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	ADALAR	40.8619	29.1208	Büyükada Tepesi	40.861107	29.117418	Mountain
1	ADALAR	40.8619	29.1208	Eski Rum Yetimhanesi	40.861705	29.123323	Historic Site
2	ADALAR	40.8619	29.1208	Aşıklar butik by şükrü	40.862570	29.118003	Hotel
3	ADALAR	40.8619	29.1208	Büyükada Bisiklet Parkuru	40.865000	29.116861	Bike Trail
4	ADALAR	40.8619	29.1208	Nizam Butik Otel & Bistro	40.863322	29.116257	Bed & Breakfast

Let's check how many venues were returned for each borough and sort them in count

```
[19]: summary = istanbul_venues.groupby('Borough').count().reset_index()
      summary['Count'] = summary['Venue']
      summary = summary.drop(['Borough Latitude', 'Borough Longitude', 'Venue', 'Venue Latitude', 'Venue Longitude','Venue Cate
      summary = summary.sort_values('Count').reset_index(drop=True)
      summary.head()
```

[19]:

	Borough	Count
0	PENDIK	9
1	BUYUKCEKMECE	13
2	ADALAR	14
3	TUZLA	14
4	KARTAL	26



Analyzing Each Borough

- Each borough with venues informations analyzed by applying one hot encoding for each district whether a venue exists or not in a given district.
    - created the new dataframe and display the top 10 venues for each neighborhood.

[25]:

	Borough	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	ADALAR	Café	Hotel	Mountain	Tea Room	Garden Center	Campground	Forest	Road	Bed & Breakfast	Nature Preserve
1	AVCILAR	Café	Hotel	Turkish Restaurant	Ice Cream Shop	Dessert Shop	Mobile Phone Shop	Modern European Restaurant	Fast Food Restaurant	Other Event	Diner
2	BAGCILAR	Café	Turkish Restaurant	Pizza Place	Kebab Restaurant	Gym	Park	Steakhouse	Department Store	Lounge	Market
3	BAHCELIEVLER	Turkish Restaurant	Café	Kebab Restaurant	Furniture / Home Store	Men's Store	Motorcycle Shop	Park	Steakhouse	Ice Cream Shop	Burger Joint
4	BAKIRKOY	Turkish Restaurant	Gym	Spa	Dessert Shop	Bookstore	Park	Tennis Court	Athletics & Sports	Sandwich Place	Tea Room

Cluster of Boroughs

- K-Means algorithm cluster method of **unsupervised learning**,
    - cluster the boroughs into 3 clusters because with elbow method it ensured me the 3 degree

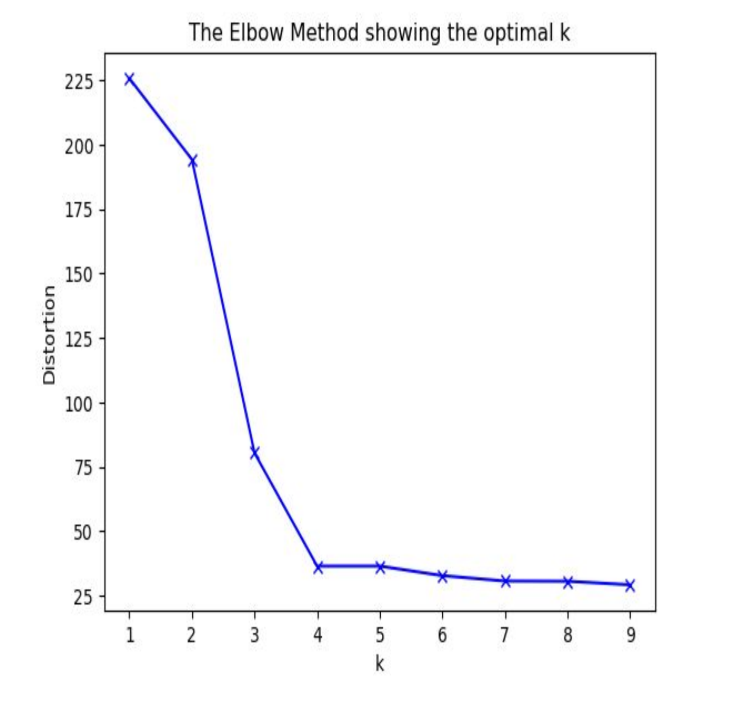
```
[39]: # set number of clusters
      kclusters = 3

      istanbul_grouped_clustering = istanbul_grouped.drop('Borough', 1)

      # run k-means clustering
      kmeans = KMeans(n_clusters=kclusters, random_state=0).fit(istanbul_grouped_clustering)

      # check cluster labels generated for each row in the dataframe
      labels = kmeans.labels_
      labels

[39]: array([2, 0, 2, 2, 2, 1, 2, 0, 2, 1, 0, 0, 0, 2, 0, 2, 0, 1, 1, 2, 2, 0,
             0, 2, 0, 2, 2, 0, 0, 0, 2])
```

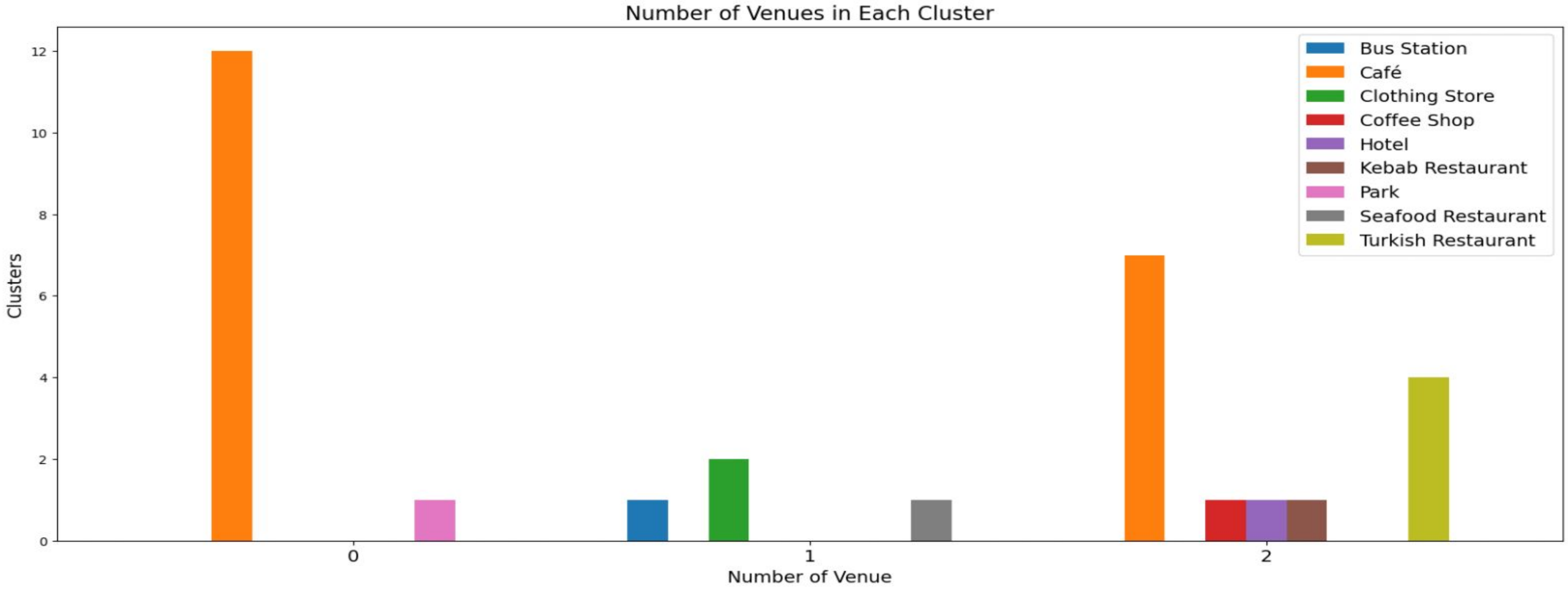


Combining Dataframe

Cluster 0 : "Cafe Venues"

Cluster 1 : "Stores and Station"

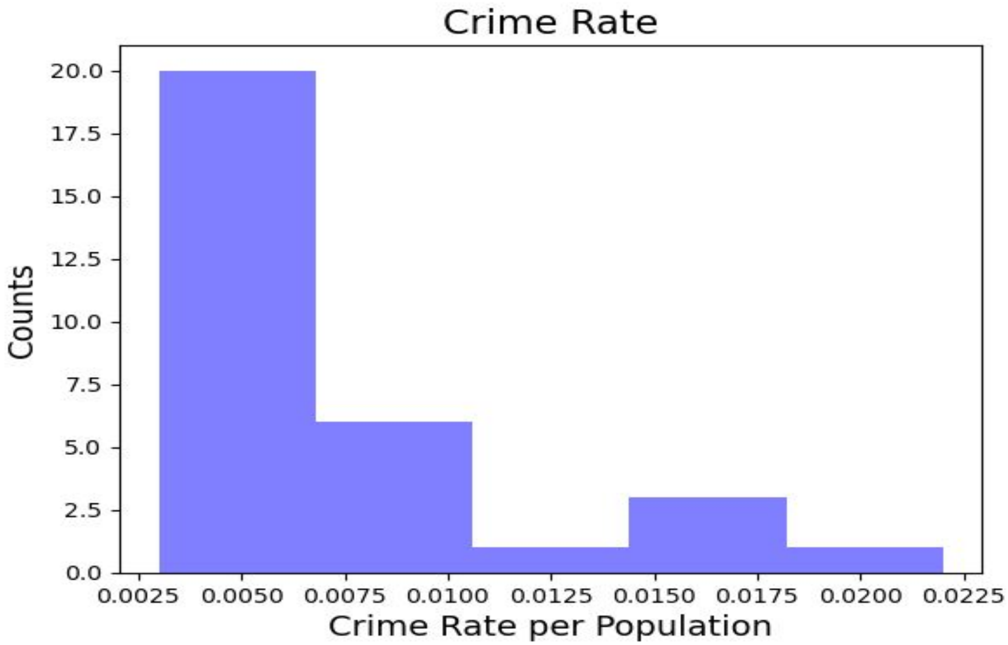
Cluster 2 : "Accommodation & Intensive Cafe Venues & Restaurants"



Results

the number of top 3 venues information for each borough on the map as **Join** column

Level_labels for crime Low level, Mid Level, High Level



	Borough	Crime Rate	Cluster Labels	Level_labels
0	GUNGOREN	0.003	2	Low Level CR
1	BAGCILAR	0.004	2	Low Level CR
2	SULTANBEYLI	0.004	2	Low Level CR
3	PENDIK	0.004	0	Low Level CR
4	KUCUKCEKMECE	0.004	2	Low Level CR

	Borough	Join
0	ADALAR	2 Café, 2 Hotel, 1 Bed & Breakfast
1	AVCILAR	16 Café, 2 Hotel, 2 Turkish Restaurant
2	BAGCILAR	5 Café, 3 Turkish Restaurant, 2 Gym
3	BAHCELIEVLER	8 Turkish Restaurant, 7 Café, 4 Kebab Restaurant
4	BAKIRKOY	3 Gym, 3 Turkish Restaurant, 2 Athletics & Sports

Final Dataframe

- Includes crime Rate, top 10 Most common venue, top 3 venues for each district, venue labels and crime levels

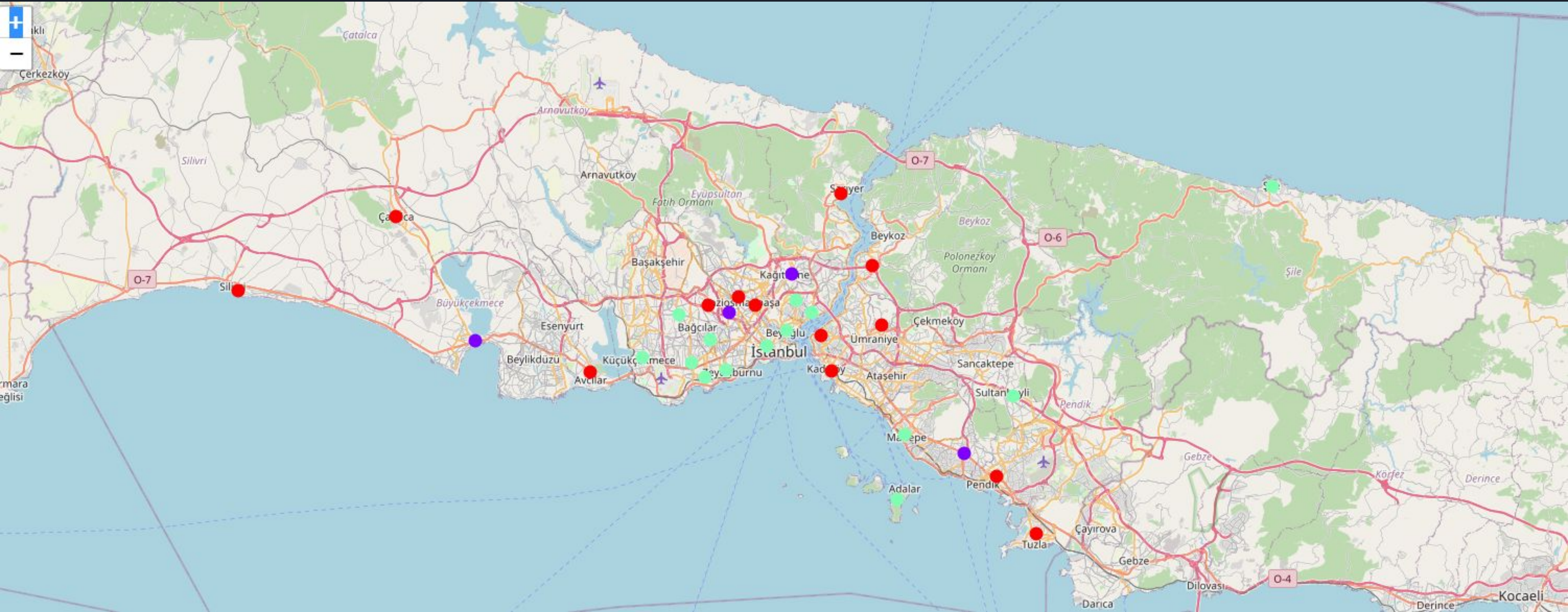
[51]:

	Borough	Crime Rate	Latitude	Longitude	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	Join	Labels	Level_labels
0	ADALAR	0.006	40.8619	29.1208	2	Café	Hotel	Mountain	Tea Room	Garden Center	Campground	Forest	Road	Bed & Breakfast	Nature Preserve	2 Café, 2 Hotel, 1 Bed & Breakfast	Accommodation & Intensive Cafe Venues & Restau...	Low Level CR
1	AVCILAR	0.005	40.9880	28.7170	0	Café	Hotel	Turkish Restaurant	Ice Cream Shop	Dessert Shop	Mobile Phone Shop	Modern European Restaurant	Fast Food Restaurant	Other Event	Diner	16 Café, 2 Hotel, 2 Turkish Restaurant	Cafe Venues	Low Level CR
2	BAGCILAR	0.004	41.0450	28.8338	2	Café	Turkish Restaurant	Pizza Place	Kebab Restaurant	Gym	Park	Steakhouse	Department Store	Lounge	Market	5 Café, 3 Turkish Restaurant, 2 Gym	Accommodation & Intensive Cafe Venues & Restau...	Low Level CR

WE NOW COMBINED THREE DIFFERENT DATAFRAMES AND CREATED A NEW ONE; WHICH INCLUDES NE COLUMNS AS JOIN, LABELS and LEVEL_LABELS columns as the LAST 3 COLUMNS in above table.

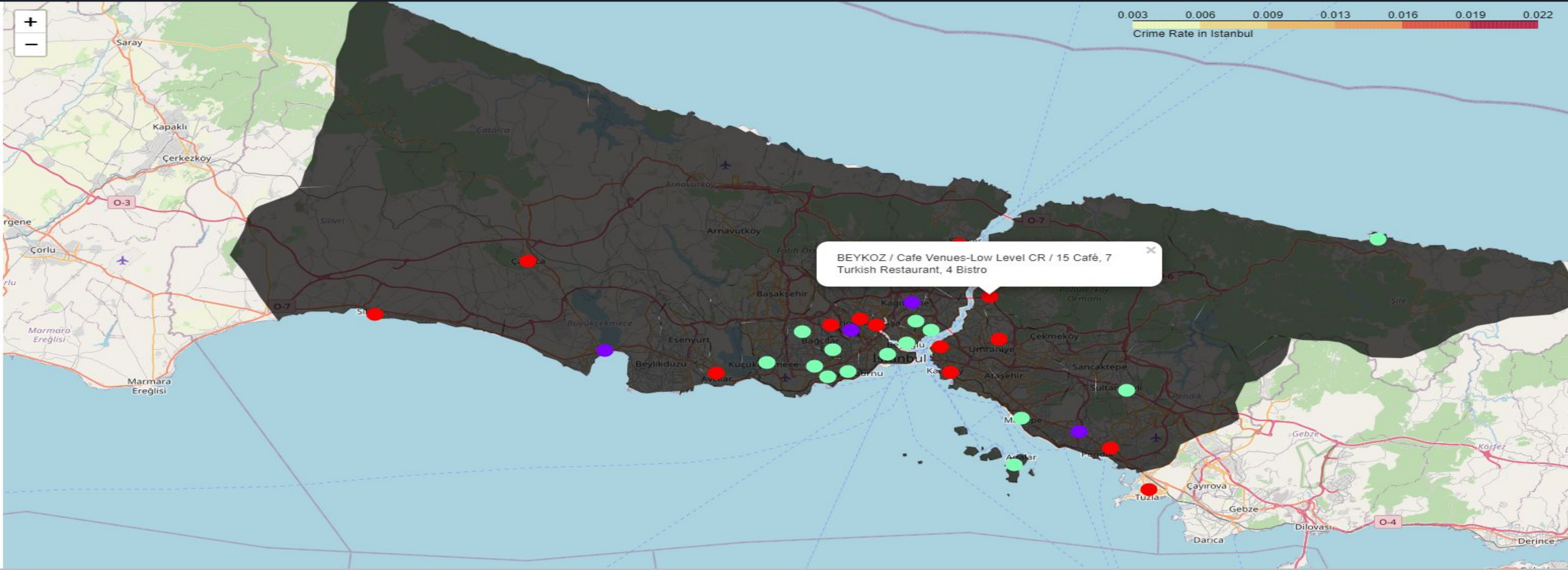
Map of Cluster Results

- Visualization of the resulting clusters



MAP OF CRIME RATES

- Choropleth map has the below informations for each borough:
  - Borough name,
  - Cluster name,
  - Crime Rate (CR) Levels,
  - Top 3 number of venue



Conclusion

As mentioned before, Tourism Restaurant Investors and Operators Association request from us to prepare a report on the following topics:

- For a resident, which area has both the safest place and highly dense social places
- For the potential investors, where to open up new venues which are crime free
- For city governance. where to allocate resources efficiently.

Now we can create a profile for each group, considering the common characteristics of each cluster. For example, the 3 clusters can be:

- LOW CRIME RATE, HIGH DENSITY OF CAFE VENUES
- MIDDLE LEVEL CRIME RATE, HIGH DENSITY OF STORES
- HIGH LEVEL OF CRIME RATE, HIGH DENSITY OF DIFFERENT TYPES OF SOCIAL VENUES (Accommodation & Intensive Cafe Venues & Restaurants)

Conclusion

As the Choropleth-map shows the findings for above problems.

- For example, for a resident Beykoz district can be suitable to spend time because this district has low level crime rates but with high density of social venues.
- On the other hand, for the city governance Fatih district has high level crime and high density of social venues, which is alarming for city management.
- If this kind of analysis is performed at the neighborhood level and deployed as an application with updated data, the relevant parties can achieve better results.
- This kind of platform will make the outcomes dynamic. By this way, multiple stakeholders can benefit from this kind of analysis.