Analyzing part of Kadıköy district in Istanbul

Applied Data Science Capstone Project

Coursera – IBM Data Science Course

Mehmet Aydinel

26.06.2019

Table of Contents

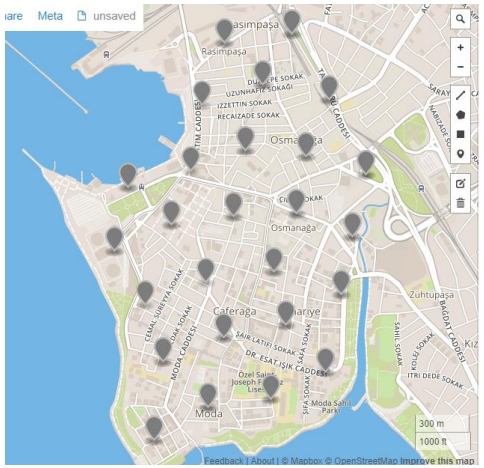
Introduction / Business Problem

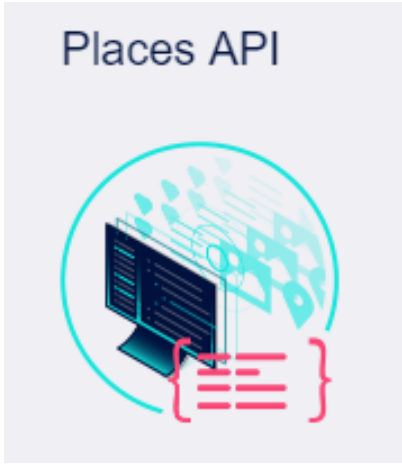
- A certain part of Kadiköy district in Istanbul, Turkey is analyzed
- A dynamic and lively part of the city with lots of Cafes, Restaurants, Bars, Shops and Residential areas
- The study can be interesting for tourists, investors or just curious residents
- Additionally, it can serve as a base for further analysis

Data

- Two main sources of data:
  - Location file created using the geojson.io website
  - Foursquare Places API

- Location data 26 points are created and saved as a .csv file





	lon	lat	marker-color	marker-size	marker-symbol	name
2	29.02231693267822	40.979962866583264	#7e7e7e	medium		A1
3	29.02538537979126	40.9814045825477	#7e7e7e	medium		A2
4	29.028990268707275	40.98174475812385	#7e7e7e	medium		A3
5	29.03214454650879	40.98291106106472	#7e7e7e	medium		A4
6	29.029870033264157	40.984952041582275	#7e7e7e	medium		B3
7	29.026308059692383	40.9844013070302 6	#7e7e7e	medium		B2

Methodology

- Location points are shown on the map created by Folium
- Information for 687 venues were retrieved from Places API
- 143 Unique categories

Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
A1	33	33	33	33	33	33
A2	15	15	15	15	15	15
A3	1	1	1	1	1	1
A4	9	9	9	9	9	9
B1	21	21	21	21	21	21
B2	71	71	71	71	71	71
B3	2	2	2	2	2	2
C1	16	16	16	16	16	16
C2	77	77	77	77	77	77
C3	26	26	26	26	26	26
C4	7	7	7	7	7	7
D1	8	8	8	8	8	8
D2	66	66	66	66	66	66

```
In [25]: kadikoy_venues
```

Out[25]:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	A1	40.979963	29.022317	Cafe MOON	40.980442	29.021340	Café
1	A1	40.979963	29.022317	Cafe Los Manços	40.981140	29.022968	Café
2	A1	40.979963	29.022317	Moda Parkı	40.980748	29.021341	Park
3	A1	40.979963	29.022317	Nefess İstanbul	40.980128	29.022858	Yoga Studio
4	A1	40.979963	29.022317	Tatlı Mesai	40.980285	29.023170	Café
5	A1	40.979963	29.022317	Meze Moda	40.980325	29.023182	Mediterranean Restaurant
6	A1	40.979963	29.022317	MEKAN Cafe	40.980362	29.021181	Restaurant
7	A1	40.979963	29.022317	Pizza Locale	40.980556	29.023853	Pizza Place
8	A1	40.979963	29.022317	MODACTIVE	40.978970	29.023144	Gym

Methodology (continued)

- One hot encoding is applied
- Top 10 venues are obtained and sorted

	Neighborhood	Accessories Store	Antique Shop	Arcade	Art Gallery	Arts & Crafts Store	Asian Restaurant	Athletics & Sports	Bagel Shop	Bakery	...
0	A1	0	0	0	0	0	0	0	0	0	...
1	A1	0	0	0	0	0	0	0	0	0	...
2	A1	0	0	0	0	0	0	0	0	0	...
3	A1	0	0	0	0	0	0	0	0	0	...
4	A1	0	0	0	0	0	0	0	0	0	...

	Neighborhood	Accessories Store	Antique Shop	Arcade	Art Gallery	Arts & Crafts Store	Asi Restaura
0	A1	0.000000	0.000000	0.000000	0.000000	0.000000	0.0000
1	A2	0.000000	0.000000	0.000000	0.000000	0.000000	0.0000
2	A3	0.000000	0.000000	0.000000	0.000000	0.000000	0.0000
3	A4	0.000000	0.000000	0.000000	0.000000	0.000000	0.0000
4	B1	0.000000	0.000000	0.000000	0.000000	0.000000	0.0000
5	B2	0.000000	0.000000	0.000000	0.000000	0.000000	0.0000
6	B3	0.000000	0.000000	0.000000	0.000000	0.000000	0.0000
7	C1	0.000000	0.000000	0.000000	0.062500	0.000000	0.0000
8	C2	0.000000	0.000000	0.000000	0.012987	0.000000	0.0000
9	C3	0.000000	0.000000	0.153846	0.038462	0.000000	0.0000
10	C4	0.000000	0.000000	0.000000	0.142857	0.000000	0.0000
11	D1	0.000000	0.000000	0.000000	0.000000	0.000000	0.0000

	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	A1	Café	Coffee Shop	Pub	Yoga Studio	Tea Room	Diner	Hot Dog Joint	Mediterranean Restaurant	Gym	Fast Food Restaurant
1	A2	Café	Music Venue	Seafood Restaurant	Boutique	Park	Turkish Restaurant	Museum	Hookah Bar	Flower Shop	Thrift / Vintage Store
2	A3	Restaurant	Yoga Studio	Fast Food Restaurant	Furniture / Home Store	Fried Chicken Joint	Food Truck	Food Court	Food & Drink Shop	Flower Shop	Falafel Restaurant
3	A4	Pool	Café	Hotel	Athletics & Sports	Restaurant	Fast Food Restaurant	Health & Beauty Service	Park	Tennis Court	Falafel Restaurant
4	B1	Café	Coffee Shop	Bakery	Dessert Shop	Chocolate Shop	Sausage Shop	Steakhouse	Food & Drink Shop	Music Venue	Sushi Restaurant

Methodology (continued)

The regions are grouped into 5 clusters by k-means Clustering algorithm

	lon	lat	name	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	29.022317	40.979963	A1	2	Café	Coffee Shop	Pub	Yoga Studio	Tea Room	Diner	Hot Dog Joint	Mediterranean Restaurant	Gym	Fast Food Restaurant
1	29.025385	40.981405	A2	2	Café	Music Venue	Seafood Restaurant	Boutique	Park	Turkish Restaurant	Museum	Hookah Bar	Flower Shop	Thrift / Vintage Store
2	29.028990	40.981745	A3	4	Restaurant	Yoga Studio	Fast Food Restaurant	Furniture / Home Store	Fried Chicken Joint	Food Truck	Food Court	Food & Drink Shop	Flower Shop	Falafel Restaurant
3	29.032145	40.982911	A4	2	Pool	Café	Hotel	Athletics & Sports	Restaurant	Fast Food Restaurant	Health & Beauty Service	Park	Tennis Court	Falafel Restaurant
4	29.029870	40.984952	B3	3	Music Venue	Tattoo Parlor	Yoga Studio	Gastropub	Fried Chicken Joint	Food Truck	Food Court	Food & Drink Shop	Flower Shop	Fast Food Restaurant

Results

- The 26 areas are grouped into 5 clusters

Mostly Cafes

Mostly Cafes, but various other venues different from the Blue Cluster

Electronic Stores and Bed&Breakfast as most common venues

Music Venue / Tattoo Parlor

Restaurant

Discussion

- Preprocessing for Venue category can improve the results

  (Falafel Restaurant, Mediterranean Restaurant, Seafood Restaurant can be categorized under Restaurant)

- Recommendation: for a tourist or for someone who is interested in exploring this area, it may be worthwhile to see the smaller clusters, since these are areas which are relatively different from the other parts
- For an investor, depending on the type of investment, he can use this study as a basis to deepen his analysis based on the type of investment he is planning

Conclusion

- 26 location points, each with about 300 meters to each other, are selected
- The data about the venues within the circles, where these points are the centers and the radius is 150 meters is retrieved from the Foursquare Places API
- A total of 687 venues were retrived and this corresponds to 143 unique categories.
- The venues in each region are sorted according to their frequencies, and then they are grouped into 5 clusters using the k-means clustering machine learning algorithm.
- This study shows the powerful potential of using the location information data. The analysis done here can be deepened further down, by taking other venue characteristics such as user ratings into account, or by focusing on a certain venue such as Restaurants, by increasing the number of location points or extending the analyzed area