

CLUSTERING DISTRICTS OF ISTANBUL



DATA SCIENCE PROJECT

INTRODUCTION

Istanbul is a big city with huge population and great places to live. But with many options to open a restaurant, how we can determine the best places to open a restaurant in exactly true location that is supported with data science

In this project, Our stakeholders want to open a restaurant but they do not have any idea where to open it . Their restaurant is a fast-food chain that is really famous over the world. They want to open their restaurants first in Istanbul,since Istanbul is the biggest city in Turkey with their capacity and population and other advantages.

DATA

We will use following data for our project ;

1. Borough's general information that includes population, annual income etc. From Wikipedia(
https://en.wikipedia.org/wiki/List_of_districts_of_Istanbul)
2. Geopy Library to get location information of boroughs.
3. Foursquare Api will be used to get information of venues.

DATASET BEFORE CLEANED

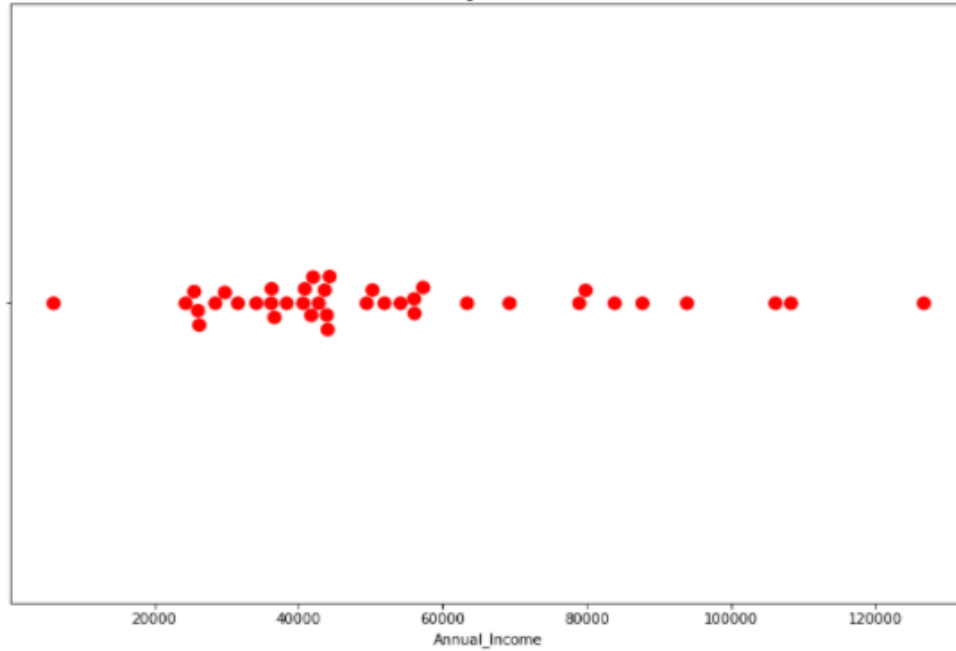
| | District | Population (2020) | Area (km ²) | Density (per km ²) | Mensual household income TL(USD) | Annual household income TL(USD) |
|---|------------|-------------------|-------------------------|--------------------------------|----------------------------------|---------------------------------|
| 0 | Adalar | 16033 | 11.05 | 1451 | 6.652₺ (918\$) | 79.821₺ (10,978\$) |
| 1 | Arnavutköy | 296709 | 450.35 | 659 | 2.030₺ (279\$) | 24.360₺ (3,350\$) |
| 2 | Ataşehir | 422594 | 25.23 | 16750 | 6.577₺ (904\$) | 78.924₺ (10,854\$) |
| 3 | Avcılar | 436897 | 42.01 | 10400 | 3.662₺ (503\$) | 43.938₺ (6,064\$) |
| 4 | Bağcılar | 737206 | 22.36 | 32970 | 3.197₺ (441\$) | 38.367₺ (5,295\$) |

CLEANED DATASET

| | Borough | Population | Area | Density | Mensual_Household_Income | Annual_Income |
|---|------------|------------|------|---------|--------------------------|---------------|
| 0 | Adalar | 16033 | 11 | 1451 | 6652 | 79821 |
| 1 | Arnavutköy | 296709 | 450 | 659 | 2030 | 24360 |
| 2 | Ataşehir | 422594 | 25 | 16750 | 6577 | 78924 |
| 3 | Avcılar | 436897 | 42 | 10400 | 3662 | 43938 |
| 4 | Bağcılar | 737206 | 22 | 32970 | 3197 | 38367 |

- Missing Values Checked
- Deleting TRY and USD Symbol
- Changed column name

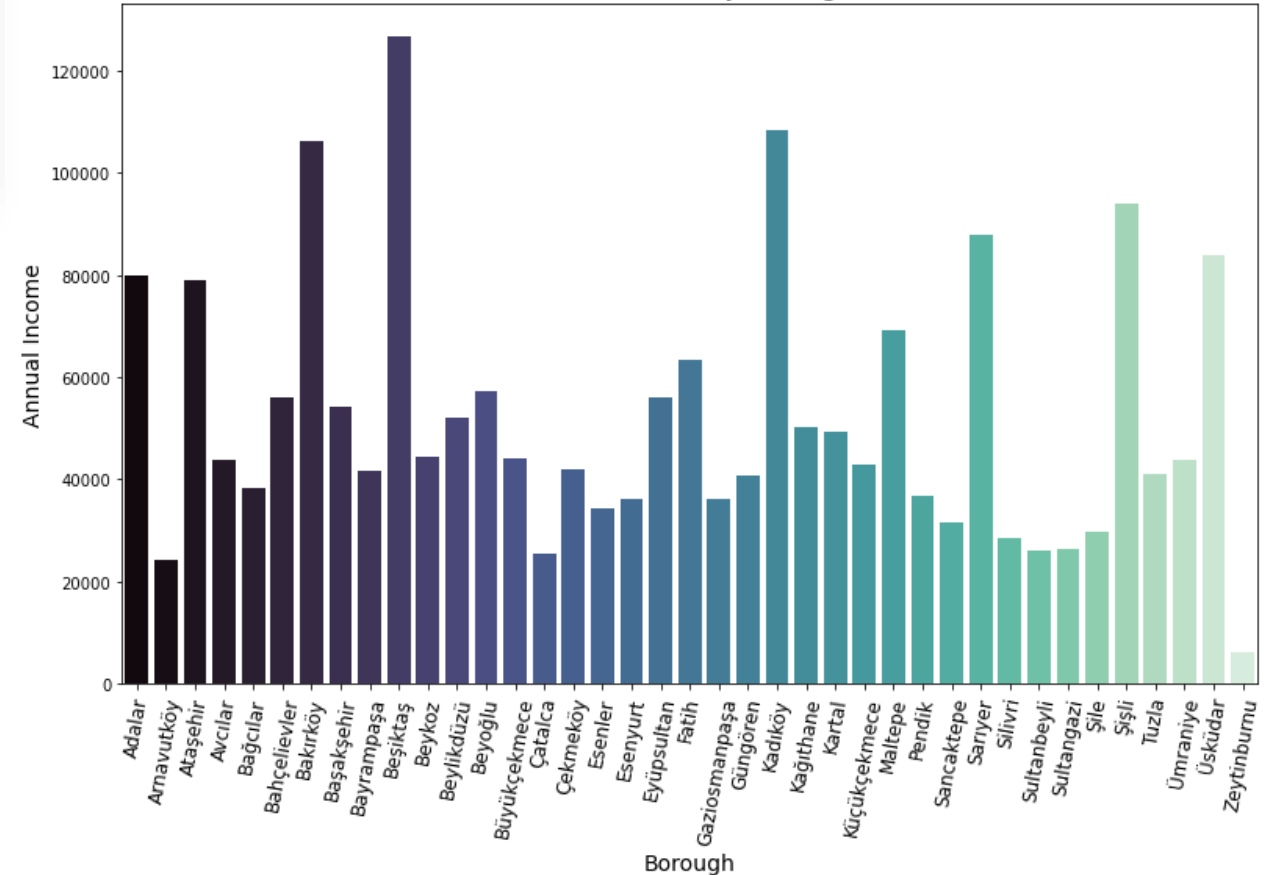
Distribution by Annual Income



- Most Boroughs has annual income between 20000-40000 TRY

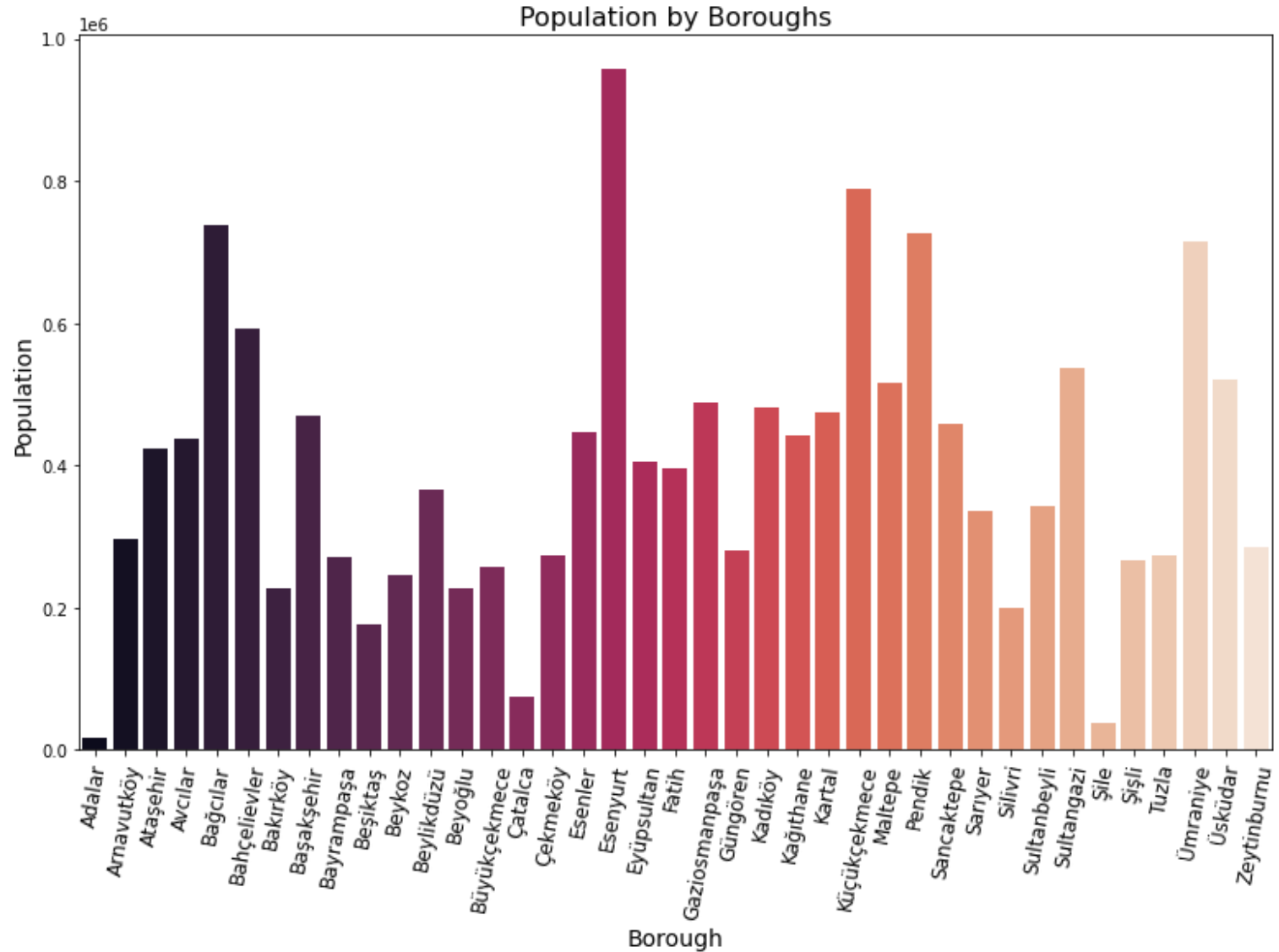
- People in Besiktas has the highest annual income
- Zeytinburnu is the district which has lowest annual income

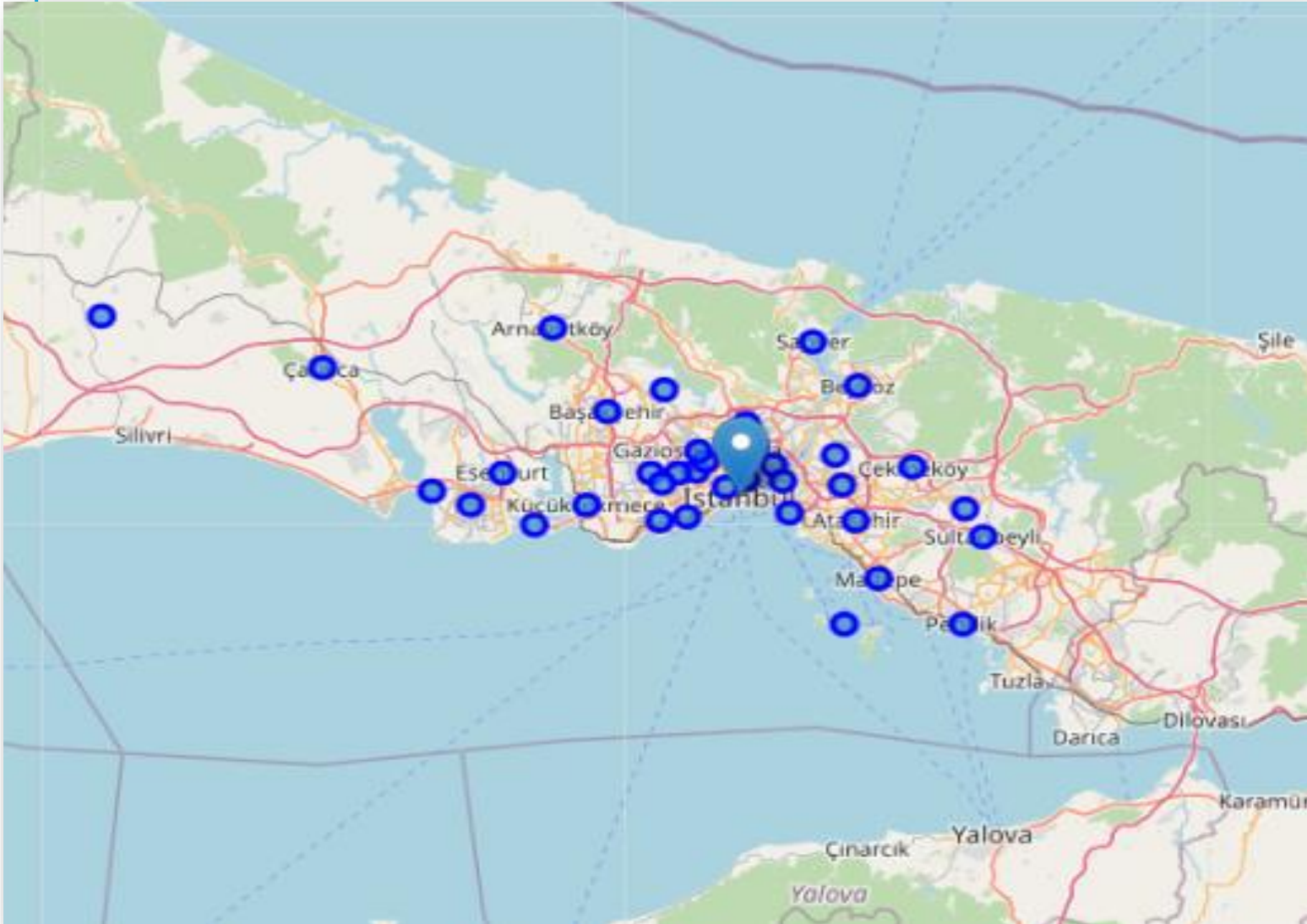
Annual Income by Boroughs



POPULATION BY BOROUGH

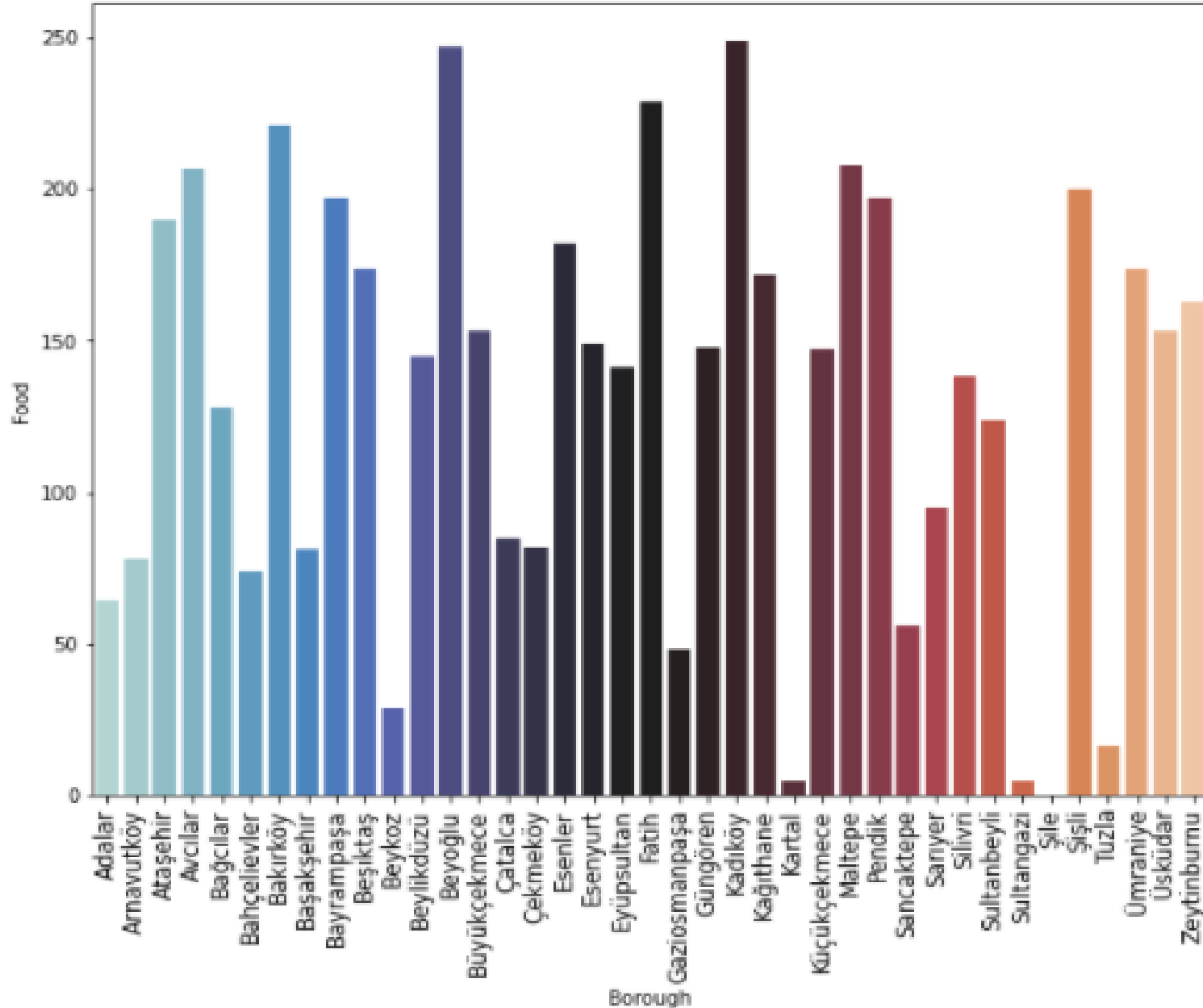
- Esenyurt has the highest population whereas Adalar has the lowest value.





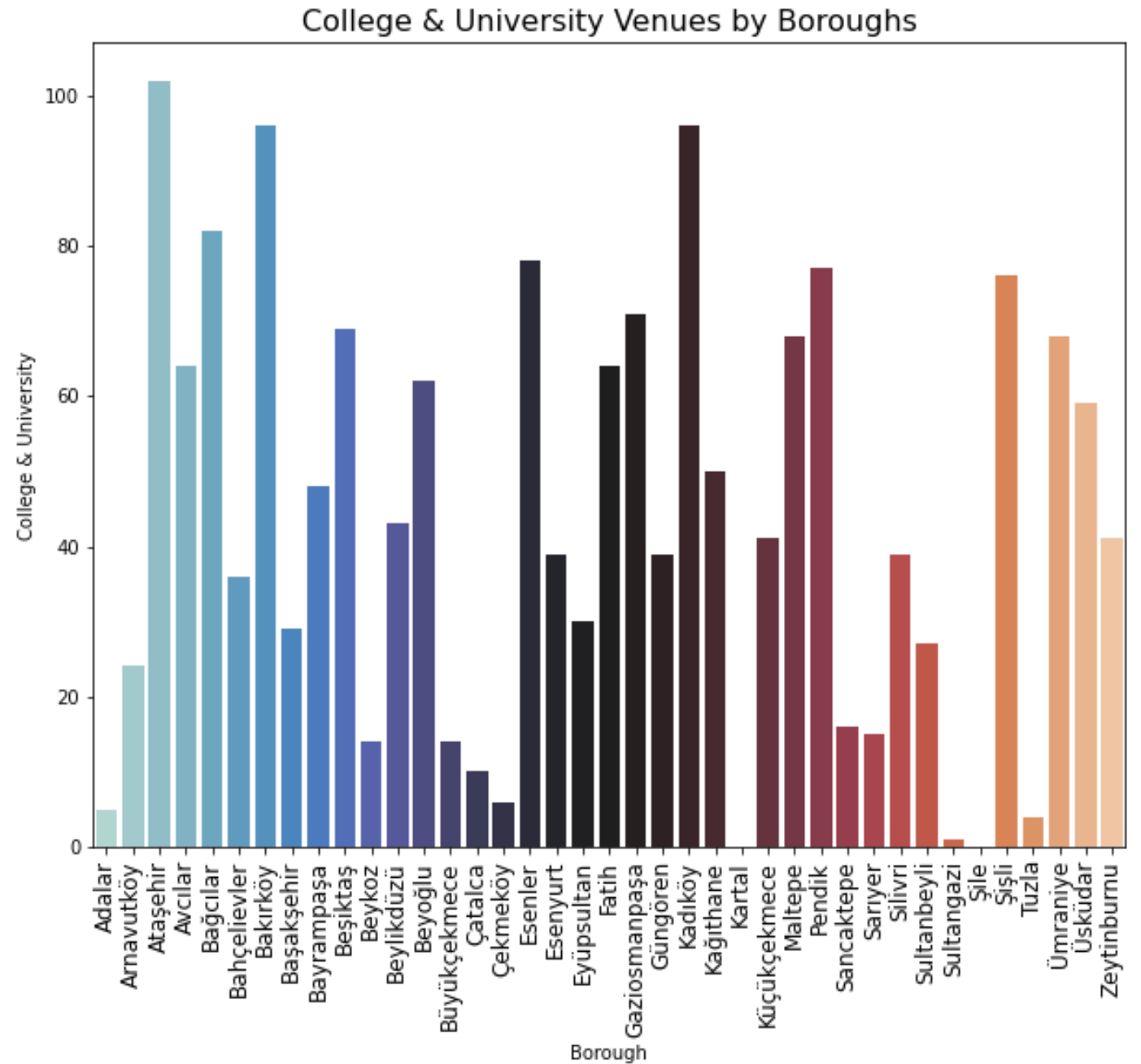
- Most of the boroughs are closer each other .
- A couple of district such as Kadikoy, Besiktas, Maltepe are so close to sea.

Food Venues by Boroughs

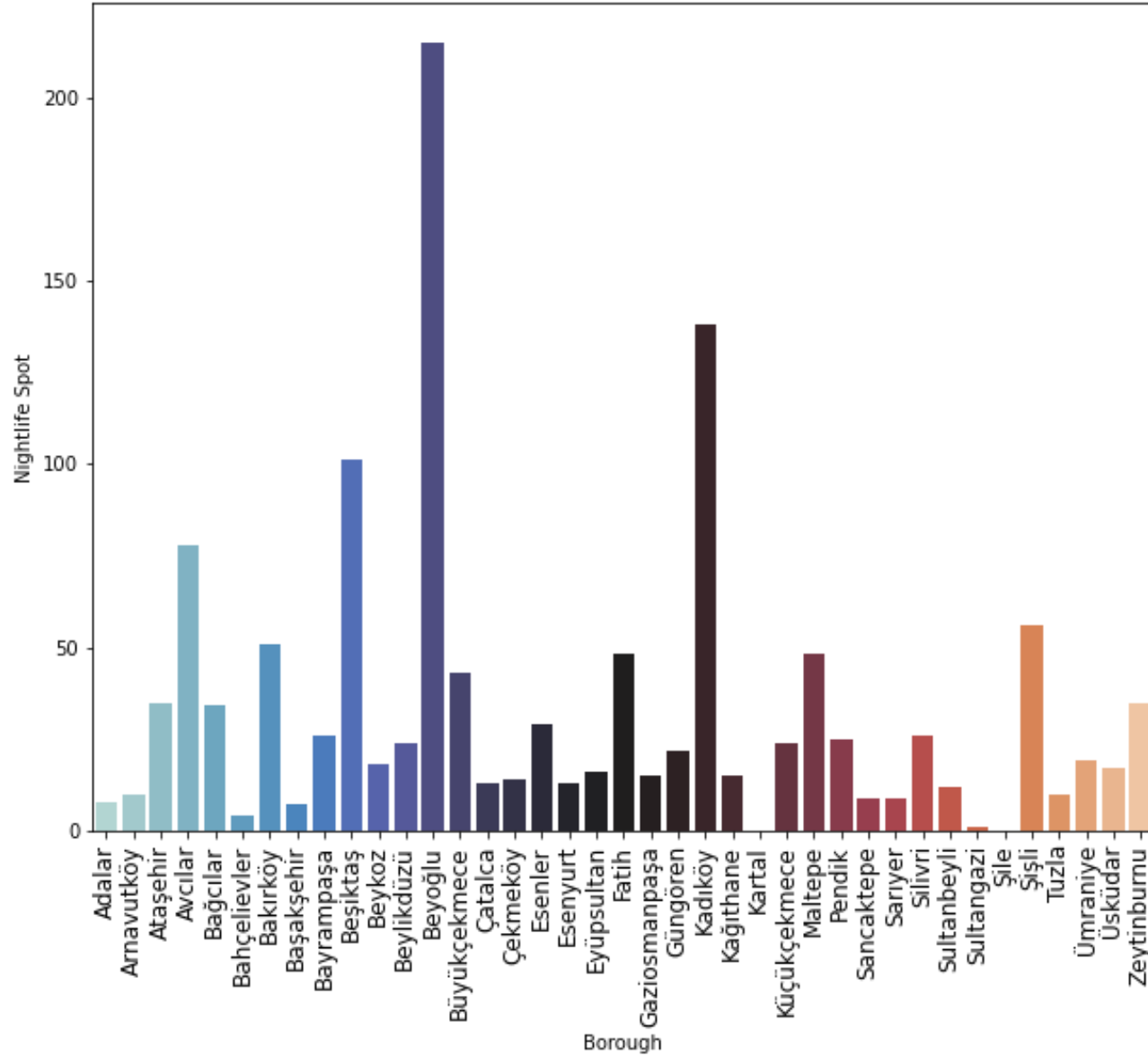


- Beyoğlu and Kadıköy have the higher number of food venues.
- There is only several places in Şile, Sultangazi and Kartal.

Most of the colleges and universities are in
Atasehir,Kadikoy,Bakirkoy
There are several schools in
Adalar,Kartal,Sultangazi

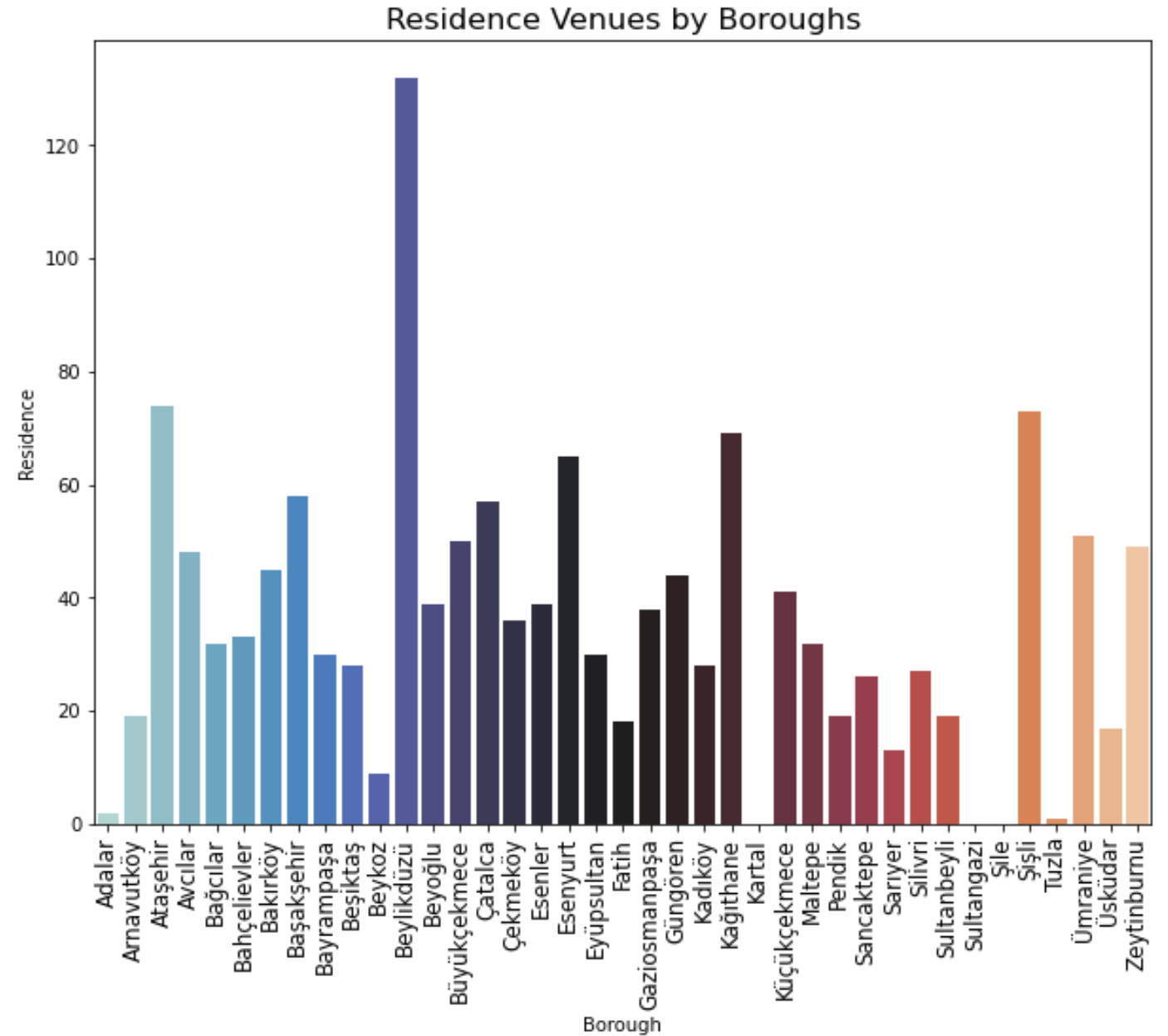


Nightlife Spot Venues by Boroughs



- Most of the Night spot is observed in Beyoglu and Kadikoy.

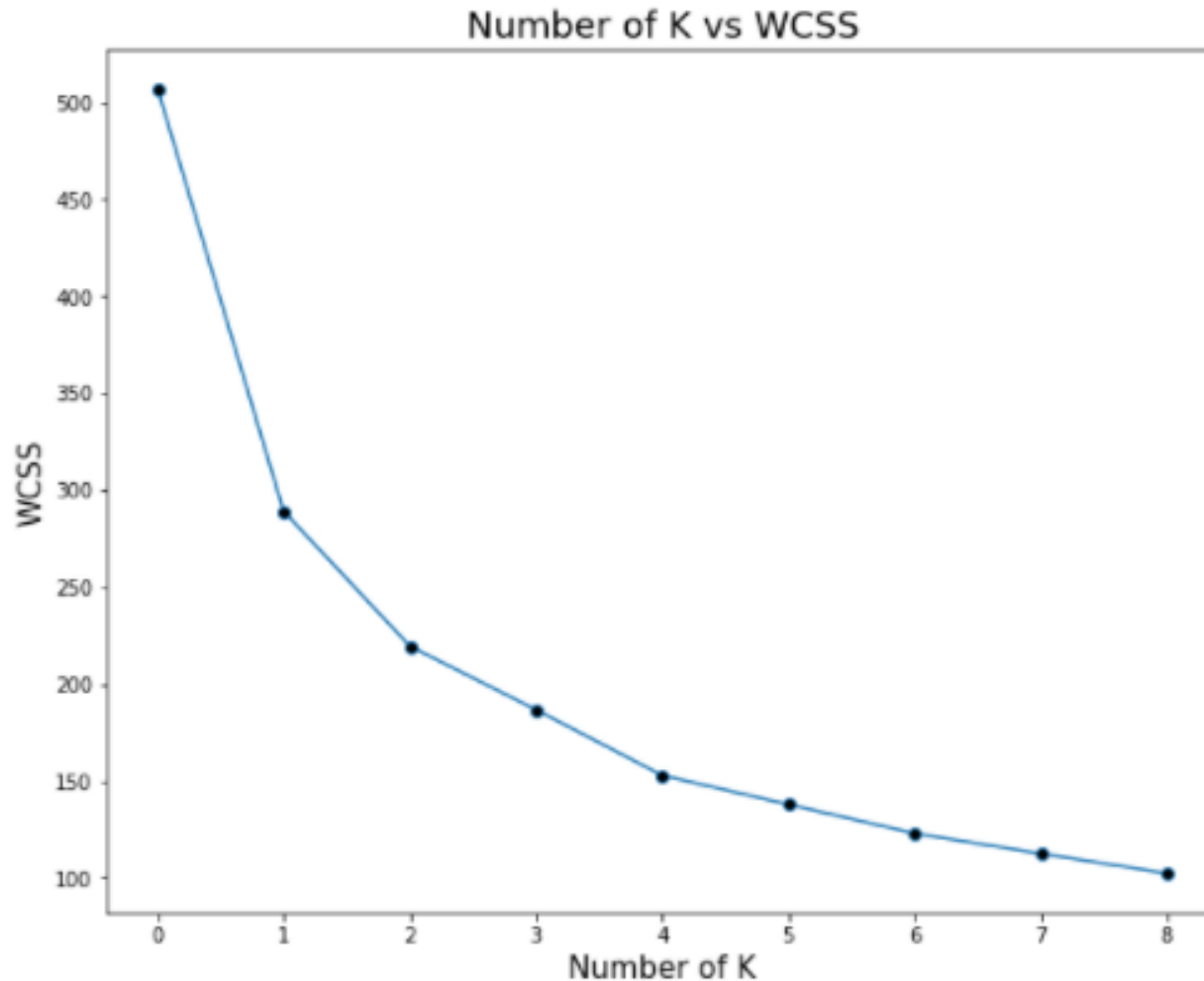
- Number of Residence in Beylikduzu is significantly high.
- There is no residence observed in Sile, Sultangazi and Kartal.



CLUSTERING GROUPING BOROUGHs

We have some common venue categories in boroughs. In this reason I used unsupervised learning algorithm. K-Means algorithm will help us group boroughs with their similar features

CHOOSING NUMBER OF CLUSTER WITH ELBOW METHOD



Before proceed K-Means Algorithm, K value selected. In this study, K value will be 3.

CLUSTER 0

| | Borough | Population | Density | Annual_Income | Arts & Entertainment | College & University | Food | Nightlife Spot | Outdoors & Recreation | Professional & Other Places | Residence | Shop & Service | Travel & Transport | Labels |
|---|----------|------------|---------|---------------|----------------------|----------------------|------|----------------|-----------------------|-----------------------------|-----------|----------------|--------------------|--------|
| 0 | Bakırköy | 226229 | 7633 | 106140 | 53 | 96 | 221 | 51 | 109 | 161 | 45 | 148 | 71 | 0 |
| 1 | Beşiktaş | 176513 | 9801 | 126720 | 73 | 69 | 174 | 101 | 80 | 118 | 28 | 76 | 95 | 0 |
| 2 | Beyoğlu | 226396 | 25409 | 57275 | 150 | 62 | 247 | 215 | 92 | 193 | 39 | 173 | 156 | 0 |
| 3 | Kadıköy | 481983 | 19210 | 108300 | 139 | 96 | 249 | 138 | 107 | 148 | 28 | 160 | 120 | 0 |
| 4 | Şişli | 266793 | 24911 | 93864 | 63 | 76 | 200 | 56 | 100 | 151 | 73 | 116 | 95 | 0 |

CLUSTER 1

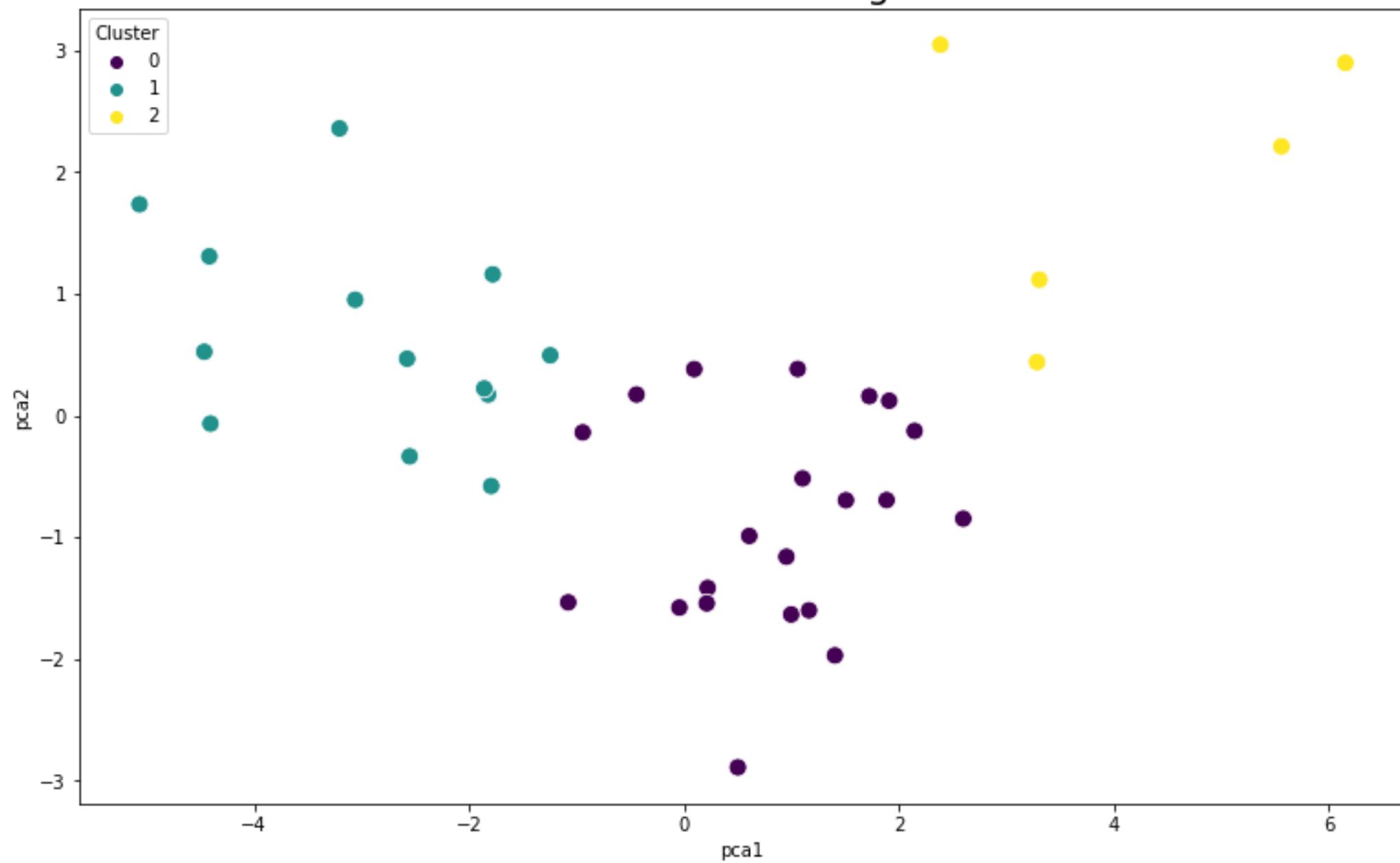
]:

| | Borough | Population | Density | Annual_Income | Arts & Entertainment | College & University | Food | Nightlife Spot | Outdoors & Recreation | Professional & Other Places | Residence | Shop & Service | Travel & Transport | Labels |
|----|---------------|------------|---------|---------------|----------------------|----------------------|------|----------------|-----------------------|-----------------------------|-----------|----------------|--------------------|--------|
| 0 | Ataşehir | 422594 | 16750 | 78924 | 28 | 102 | 190 | 35 | 106 | 162 | 74 | 113 | 54 | 1 |
| 1 | Avcılar | 436897 | 10400 | 43938 | 41 | 64 | 207 | 78 | 101 | 114 | 48 | 105 | 75 | 1 |
| 2 | Bağcılar | 737206 | 32970 | 38367 | 52 | 82 | 128 | 34 | 111 | 107 | 32 | 103 | 26 | 1 |
| 3 | Bahçelievler | 592371 | 35642 | 56088 | 12 | 36 | 74 | 4 | 69 | 72 | 33 | 65 | 11 | 1 |
| 4 | Bayrampaşa | 269950 | 28091 | 41762 | 43 | 48 | 197 | 26 | 109 | 161 | 30 | 149 | 43 | 1 |
| 5 | Beylikdüzü | 365572 | 9676 | 51924 | 27 | 43 | 145 | 24 | 82 | 91 | 132 | 77 | 15 | 1 |
| 6 | Esenler | 446276 | 24215 | 34164 | 30 | 78 | 182 | 29 | 99 | 113 | 39 | 74 | 97 | 1 |
| 7 | Esenyurt | 957398 | 22198 | 36288 | 23 | 39 | 149 | 13 | 81 | 127 | 65 | 115 | 18 | 1 |
| 8 | Eyüpsultan | 405845 | 1777 | 56044 | 11 | 30 | 141 | 16 | 80 | 136 | 30 | 62 | 45 | 1 |
| 9 | Fatih | 396594 | 25439 | 63378 | 35 | 64 | 229 | 48 | 102 | 160 | 18 | 109 | 60 | 1 |
| 10 | Gaziosmanpaşa | 487778 | 41478 | 36228 | 43 | 71 | 48 | 15 | 67 | 125 | 38 | 72 | 37 | 1 |
| 11 | Güngören | 280299 | 38876 | 40656 | 32 | 39 | 148 | 22 | 87 | 144 | 44 | 118 | 35 | 1 |
| 12 | Kağıthane | 442415 | 29752 | 50260 | 23 | 50 | 172 | 15 | 93 | 145 | 69 | 83 | 34 | 1 |
| 13 | Küçükçekmece | 789633 | 21034 | 42804 | 36 | 41 | 147 | 24 | 86 | 83 | 41 | 67 | 24 | 1 |
| 14 | Maltepe | 515021 | 9723 | 69259 | 46 | 68 | 208 | 48 | 97 | 112 | 32 | 114 | 57 | 1 |
| 15 | Pendik | 726481 | 4036 | 36664 | 38 | 77 | 197 | 25 | 81 | 109 | 19 | 102 | 78 | 1 |
| 16 | Silivri | 200215 | 230 | 28464 | 30 | 39 | 138 | 26 | 85 | 134 | 27 | 126 | 39 | 1 |
| 17 | Sultanbeyli | 343318 | 11782 | 26064 | 25 | 27 | 124 | 12 | 56 | 112 | 19 | 91 | 31 | 1 |
| 18 | Ümraniye | 713803 | 15754 | 43641 | 25 | 68 | 174 | 19 | 89 | 122 | 51 | 113 | 52 | 1 |
| 19 | Üsküdar | 520771 | 14740 | 83839 | 43 | 59 | 153 | 17 | 87 | 117 | 17 | 87 | 76 | 1 |
| 20 | Zeytinburnu | 283657 | 24474 | 6036 | 31 | 41 | 163 | 35 | 75 | 115 | 49 | 135 | 47 | 1 |

CLUSTER 2

| | Borough | Population | Density | Annual_Income | Arts & Entertainment | College & University | Food | Nightlife Spot | Outdoors & Recreation | Professional & Other Places | Residence | Shop & Service | Travel & Transport | Labels |
|----|------------|------------|---------|---------------|----------------------|----------------------|------|----------------|-----------------------|-----------------------------|-----------|----------------|--------------------|--------|
| 0 | Adalar | 16033 | 1451 | 79821 | 5 | 5 | 64 | 8 | 46 | 25 | 2 | 7 | 26 | 2 |
| 1 | Arnavutköy | 296709 | 659 | 24360 | 17 | 24 | 78 | 10 | 51 | 101 | 19 | 90 | 13 | 2 |
| 2 | Başakşehir | 469924 | 4506 | 54152 | 3 | 29 | 81 | 7 | 48 | 89 | 58 | 57 | 5 | 2 |
| 3 | Beykoz | 246110 | 793 | 44316 | 6 | 14 | 29 | 18 | 45 | 76 | 9 | 20 | 5 | 2 |
| 4 | Büyükdere | 257362 | 1849 | 44049 | 10 | 14 | 153 | 43 | 76 | 58 | 50 | 55 | 25 | 2 |
| 5 | Çatalca | 74975 | 67 | 25536 | 6 | 10 | 85 | 13 | 53 | 106 | 57 | 85 | 12 | 2 |
| 6 | Çekmeköy | 273658 | 1848 | 42033 | 11 | 6 | 82 | 14 | 42 | 68 | 36 | 43 | 6 | 2 |
| 7 | Kartal | 474514 | 12312 | 49443 | 1 | 0 | 5 | 0 | 1 | 5 | 0 | 4 | 2 | 2 |
| 8 | Sancaktepe | 456861 | 7319 | 31602 | 8 | 16 | 56 | 9 | 42 | 91 | 26 | 35 | 5 | 2 |
| 9 | Sarıyer | 335298 | 1912 | 87696 | 11 | 15 | 95 | 9 | 60 | 75 | 13 | 54 | 21 | 2 |
| 10 | Sultangazi | 537488 | 14807 | 26244 | 4 | 1 | 5 | 1 | 4 | 18 | 0 | 1 | 2 | 2 |
| 11 | Şile | 37904 | 48 | 29789 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 |
| 12 | Tuzla | 273608 | 2213 | 40884 | 8 | 4 | 16 | 10 | 5 | 7 | 1 | 6 | 2 | 2 |

Clusters of Boroughs



RESULTS

Cluster 0 : This group has higher number of venues and annual income. There are many Transportation ,Food Venues,Nightlife Spot in this group.If our menu is expensive. We can choose to start with this group.

Cluster 1 : Boroughs in this group has higher population compared to cluster 0. There are many food venue. Annual income of this group is average of entire boroughs.It is a middle segment group of our analysis.

Cluster 2 : This group has the lowest venue and people who live in these boroughs has average annual.Their population is also lower than others in general.

CONCLUSION

In this project , We used K-Means algorithm to cluster our dataset.K-Means clustering algorithm is moslty used algorithm amongst Unsupervised Learning.

After we explore our dataset with venues information. With this study , We can get important information based on Boroughs.

To get business decision to open a restaurant , We could also expand our dataset with venues name,population details etc. to target more specifically.

This project also can be used for people who have never been to Istanbul but want to know basic information about venues, type of boroughs.