The battle of neighborhoods

Christos Vlachos

Introduction

- We will use data science approaches to make suggestions regarding the best neighborhood to open a new restaurant.
- The same idea could also apply to other types of businesses as well.
- Our suggestions will be based on how many restaurants already exist in one neighborhood.

Materials and Methods: the dataset

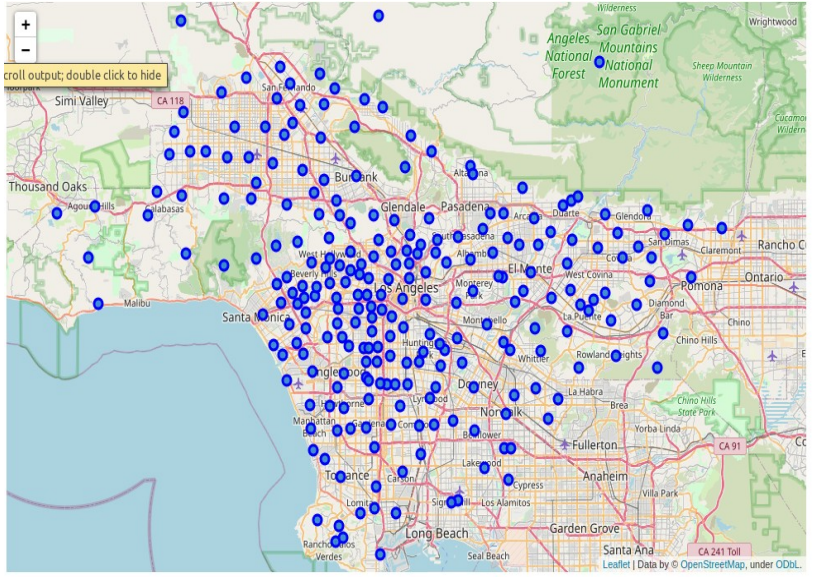
- We will use the geospatial information of Los Angeles.

- The dataset can be downloaded from:

  https://usc.data.socrata.com/dataset/Los-Angeles-Neighborhood-Map/r8qd-yxsr

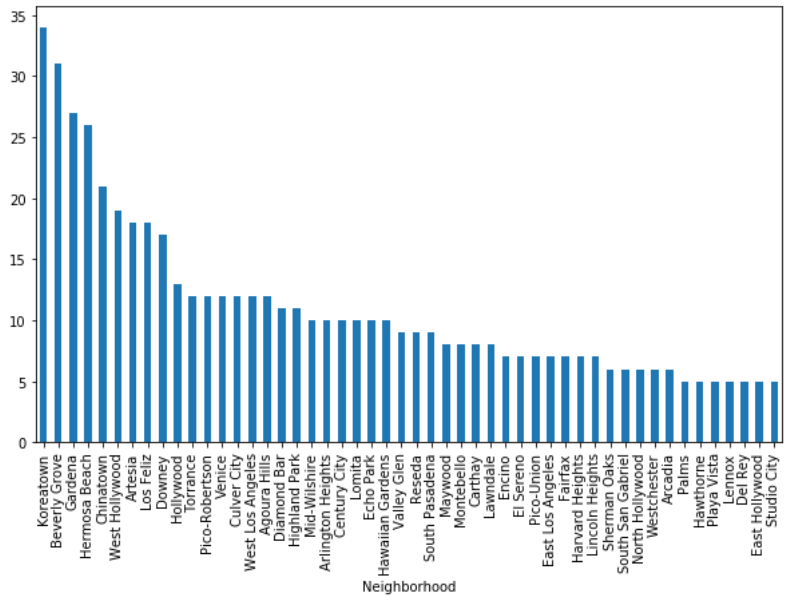
Materials and Methods: the dataset

Each blue dot on the map corresponds to a neighborhood.



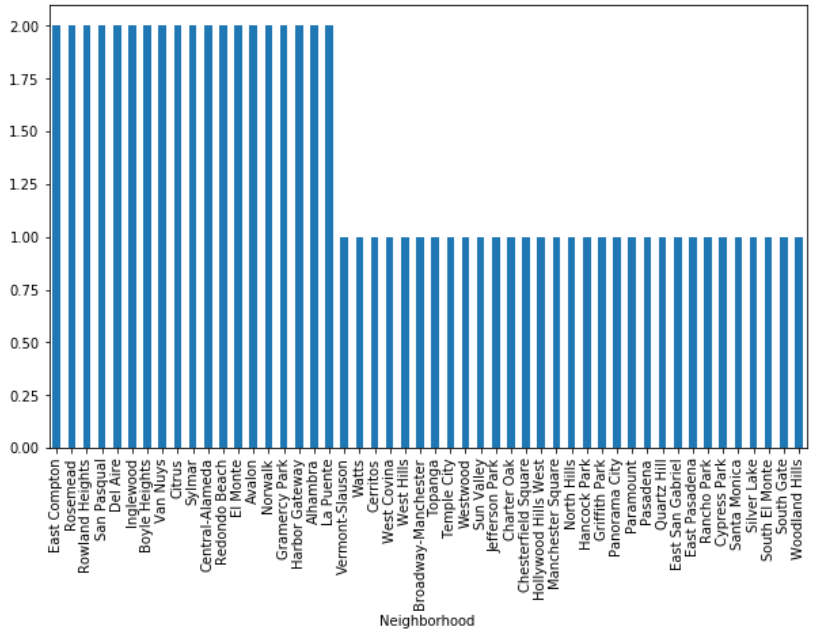
Analysis

- To obtain a first intuition on which neighborhoods might be the best choices for a new restaurant we will use the total number of restaurants.
- The plot shows the 50 neighborhoods with the most restaurants



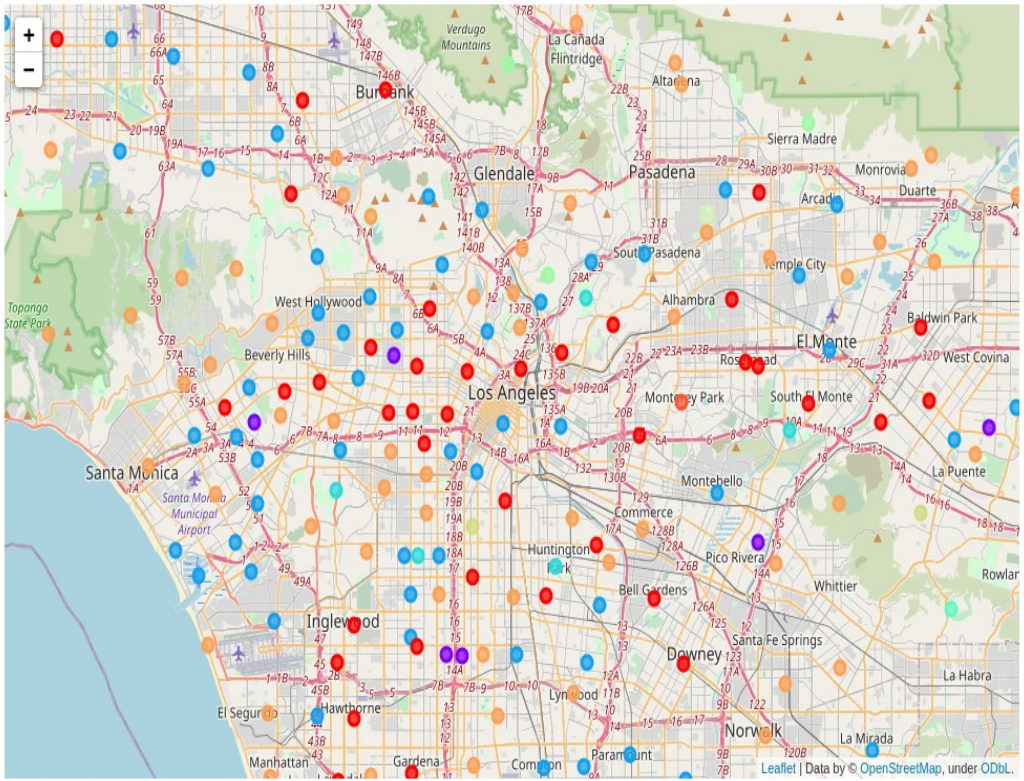
Analysis

- The plot shows the 50 neighborhoods with the least restaurants. Notice that neighborhoods with no restaurants are not inluded.



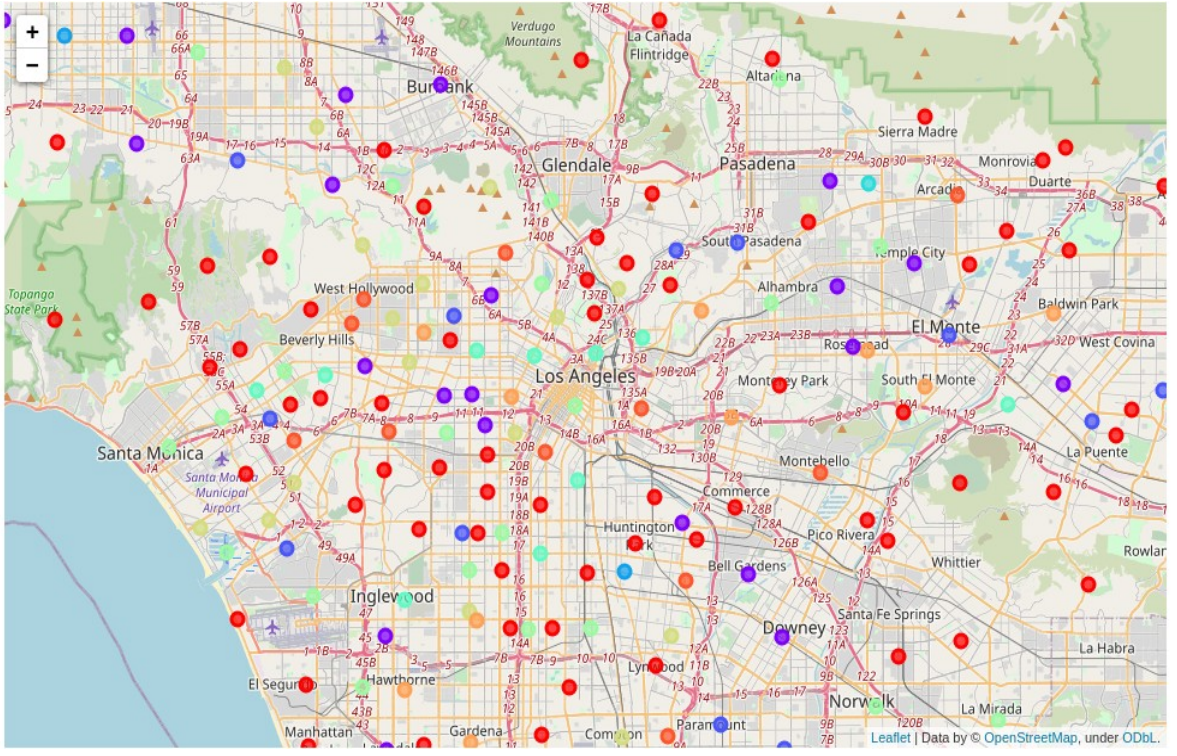
Analysis

- Next we will use the k-means clustering method to make clusters of similar neighborhoods.
- Information for all types of venues are included.



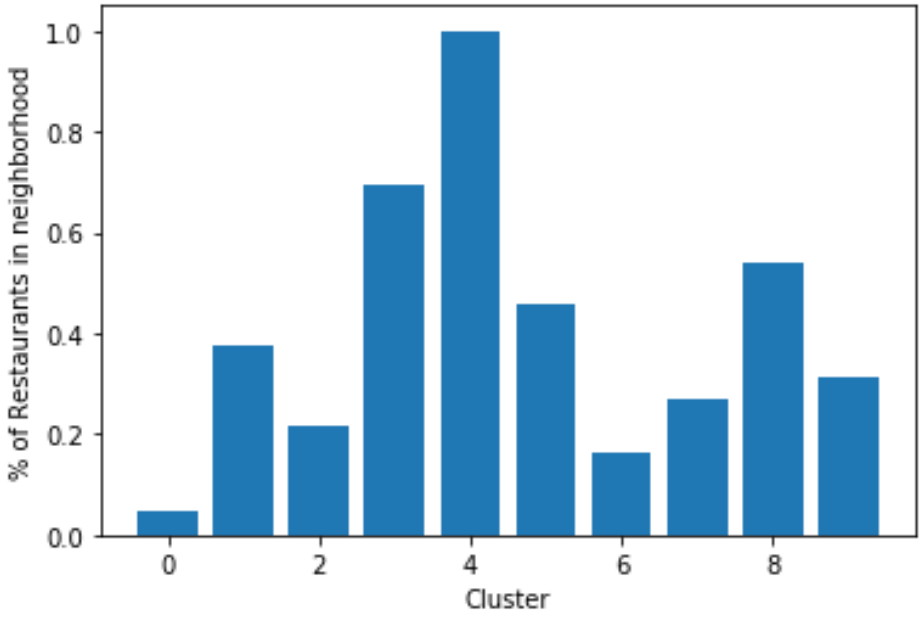
Analysis

- Unfortunately the previous map is not clear enough to make suggestions. Also it contains information that we do nor need. Now let’s repeat the clustering using only information for the restaurants in each neighborhood.



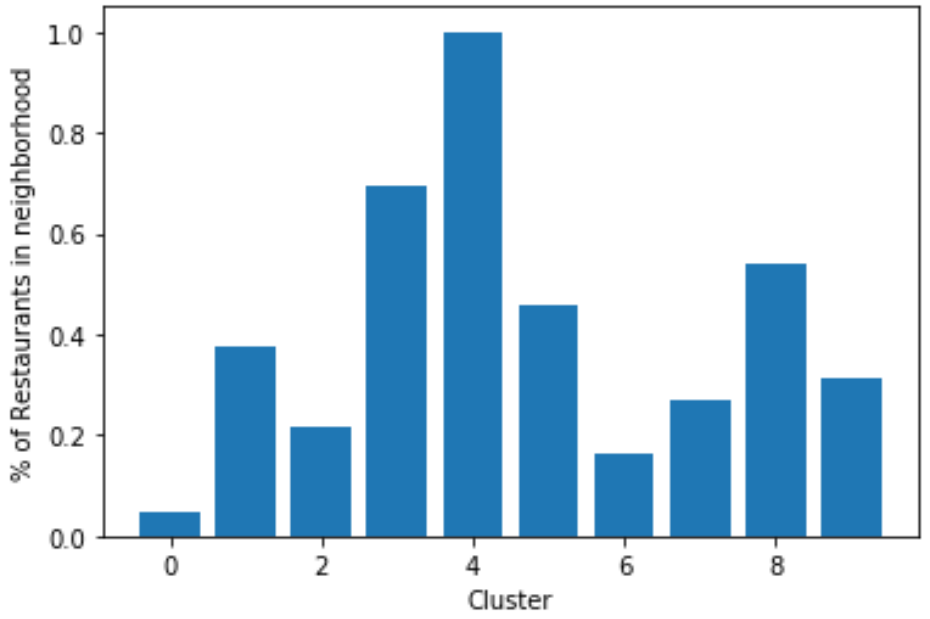
Analysis

- It looks better, but we still cannot make suggestions. Let’s try to investigate our clusters a bit more.
- In this plot we see the percentage of restaurants in each neighborhood as well as how the neighborhoods have been separated into clusters.



Analysis

- We clearly see that neighborhoods from the clusters 0 and 6 do not have many restaurants.



Results

- Finally, we can make a suggestion and say that neighborhoods from the clusters 0 and 6 are the most appropriate to open a new restaurant, since they have the least restaurants.

```
final_dataset.loc[(final_dataset['Cluster Labels']==0) | (final_dataset['Cluster Labels']==6) ,:].head()
```

	TotalRestaurants	Cluster Labels	City	Neighborhood	Latitude	Longitude
0	0.000000	0	L.A.	Acton	34.497355239240846	-118.16981019229348
3	0.000000	0	L.A.	Agua Dulce	34.504926999796837	-118.3171036690717
4	0.153846	6	L.A.	Alhambra	34.085538999123571	-118.13651200000021
5	0.000000	0	L.A.	Alondra Park	33.889617004889644	-118.33515598608159
6	0.000000	0	L.A.	Altadena	34.193870502232173	-118.13623898201556

We already see that in those clusters we have neighborhoods that do not restaurants at all, or the percentage of restaurants is very low.

Discussion

- In this project we used the geolospatial information of Los Angeles to make suggestions regarding which districts are more suitable for opening a new restaurant.
- We identified which districts have less restaurants than the rest, clustered them and made an appropriate suggestion.
- Of cource this is a simplified scenario.
- To address the question at its core, we should also take into account other factors, such as the average income in each neighborhood, the criminality levels, the average age of the citizens etc..