

Finding an optimal location for opening an Indian Restaurant in New York

Author: Shahbaaz Khan

Date: 16th Feb, 2019

Introduction

- Location plays one of the major factor for success of a restaurant
- This project is targeted to the individuals interested in opening an Indian Restaurant in New York
- It will try to answer questions such as: Should I place my restaurant in a place where there are a lot of other Indian Restaurants ?
OR in a place where there is none ?
OR should I try to find some kind of balance between these two extremes ?

Data

Based on definition of our problem, factors that will influence our decision are:

- Find the high rating Indian restaurants in NY and identify them visually on a map
- Create several "clusters" of these Indian Restaurants to identify potential districts to choose from
- We will get the number of restaurants and their type and location in every neighborhood by using **Foursquare API**

Methodology

a. Getting the New York location through Geolocator

```
address = '102 North End Ave, New York, NY'  
  
geolocator = Nominatim()  
location = geolocator.geocode(address)  
latitude = location.latitude  
longitude = location.longitude  
print(latitude, longitude)
```

Methodology

b. Getting the Indian restaurants location details from Foursquare API (Code used)

url =

```
'https://api.foursquare.com/v2/venues/explore?client_id={}&
client_secret={}&ll={},{}&v={}&query={}&radius={}&limit={}'.f
ormat(CLIENT_ID, CLIENT_SECRET, latitude, longitude,
VERSION, search_query, radius, LIMIT)
```

Methodology

c. Storing the Foursquare results (.json) in a file

```
results = requests.get(url).json()  
# results
```

Methodology

d. Transforming the file into a pandas dataframe

```
df=pd.DataFrame()  
df['venue_ID']=dataframe['venue.id']  
df['name']=dataframe['venue.name']  
df['lat']=dataframe['venue.location.lat']  
df['lng']=dataframe['venue.location.lng']  
df.head()
```

]:

	venue_ID	name	lat	lng
0	4bbb9dbded7776b0e1ad3e51	Tamarind TriBeCa	40.719211	-74.008727
1	5a1e961c1987ec47beed877d	Baar Baar	40.724534	-73.991624
2	5b770657c0cacb002c89bc63	The Kati Roll Company	40.709114	-74.009091
3	575dea4c498e2739e43a27e2	Aahar Indian Cuisine	40.713307	-74.007994
4	4593ed04f964a52050401fe3	The Kati Roll Company	40.729570	-74.000861

Methodology

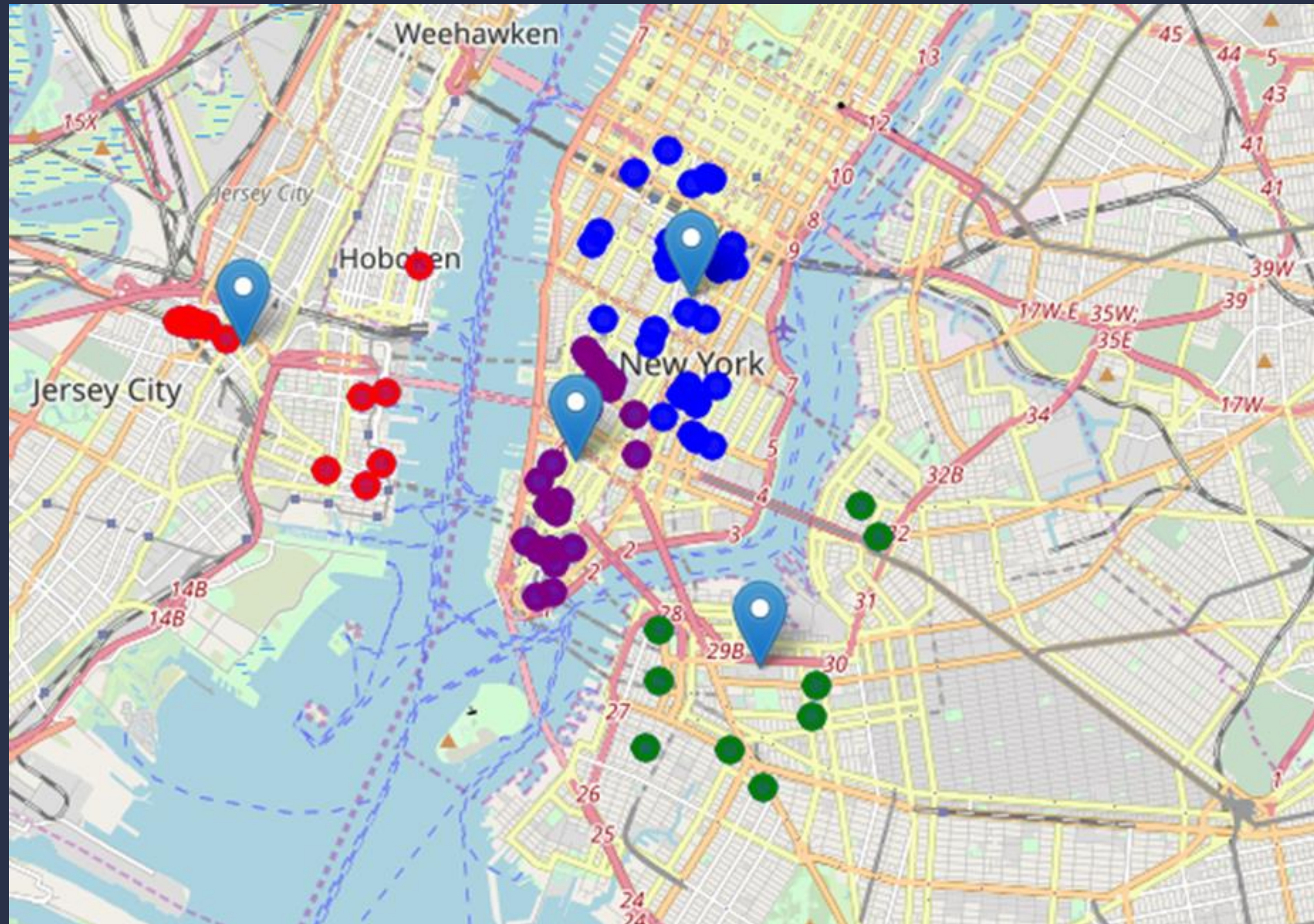
e. Using K-means clustering to cluster the locations

```
k_means = KMeans(init = "k-means++", n_clusters = 4, n_init = 12)
df_array=np.array(df[['lat','lng']])
k_means.fit(df_array)
k_labels=pd.DataFrame(k_means.labels_)
k_labels.info()
df['cluster']=k_labels
df.groupby('cluster').count()
```

	venue_ID	name	lat	lng
cluster				
0	44	44	44	44
1	21	21	21	21
2	9	9	9	9
3	25	25	25	25

Results

Markers shows the recommended location to open an Indian Restaurant



Conclusion

- The clustering is very useful and intuitive to solve the problem related to geographic data
- If we can identify the reason that causes the clustering, it might be a valuable business insight
- Final decision on optimal restaurant location will be after taking into consideration additional factors like proximity to major roads, real estate availability, prices, social and economic dynamics of every neighborhood etc.