Comparison of Neighbourhoods of Cairo, Chicago and Sheffield

Contents

Introduction

The problem I am considering is to compare the neighborhoods of three cities in three different countries. The cities I have selected are Cairo in Egypt, Chicago in USA and Sheffield in England, the reasons of select Cairo is to deal with data around us, Chicago and Sheffield due to the pictures below, Attempt will be made to check which neighborhoods of the 3 cities are similar.

20 Most Profitable Cities In Europe
to Buy an Investment Property

City	Country	Renting Price/m²	Buying Price/m²	RoI Score
Sheffield	UK	3.33	1,027	22.7
Crete	Greece	2.49	807	21.6
Antalya	Turkey	1.70	557	21.4
Derby	UK	3.64	1,412	18.0
Swansea	UK	3.37	1,387	17.0
Coventry	UK	2.82	1,353	14.6
Budapest	Hungary	2.24	1,159	13.5
Cardiff	UK	3.06	1,626	13.2
St. Petersburg	Russia	2.49	1,325	13.1
Athens	Greece	2.43	1,322	12.9
Bath	UK	3.69	2,037	12.7
Sofia	Bulgaria	1.49	862	12.1
Liverpool	UK	3.76	2,183	12.1
Granada	Spain	2.18	1,269	12.0
Manchester	UK	3.55	2,092	11.9
Leipzig	Germany	2.64	1,558	11.9
Istanbul	Turkey	2.39	1,414	11.8
Porto	Portugal	2.08	1,285	11.3
Riga	Latvia	1.85	1,146	11.3
Glasgow	UK	2.61	1,623	11.3

10 Best Cities to Invest in Rental Properties in 2022

Author: Marc Rapport | December 03, 2021

Source: Getty Images

5 of 12
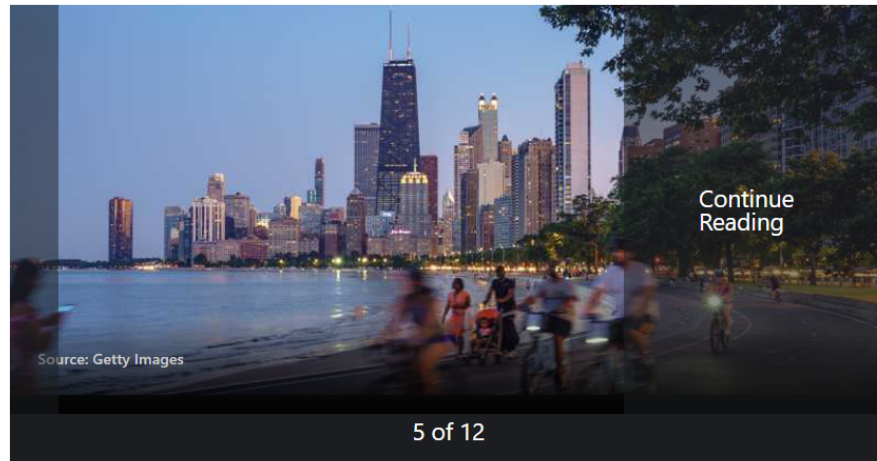
4. Chicago, Illinois

With about 2.75 million people in the city proper and 9.62 million in the metro area as a whole, Chicago is the nation's third-largest market. The Windy City is home to hundreds of major corporations and a wide range of housing options, from the most affordable to the most expensive.

Older neighborhoods in the downtown area are good places to start looking for single-family homes and small apartment buildings for $200,000 or less. The rental demand from students and young professionals -- and for Airbnb-type rentals that enable people to visit the city to experience its sports, culture, and food -- should be strong.

The target audience for this project is the owners of a restaurant chain which might already have their franchises set up in Cairo and who want to enter new markets in Europe. They might consider other

prominent tech cities in Europe since their target customer is the tech community. Since Chicago and Sheffield have many MNCs and have a large tech community, they are the target of this project.

Data Sources

The neighborhood data of the four cities is taken from Wikipedia pages.

https://en.wikipedia.org/wiki/Category:Districts_of_Cairo

https://en.wikipedia.org/wiki/List_of_neighborhoods_in_Chicago

https://en.wikipedia.org/wiki/Category:Suburbs_of_Sheffield

Beautiful Soup library will be used to web-scrap from the Wikipedia pages. The geocode library to be used is geopy. Care is taken to limit the calls to be less than 1call/sec to meet the term of use of the library. Folium library will be used to represent the data on maps. And scikit-learn will be used to utilize machine learning. And Foursquare API will be used to gather neighborhood data.

Cairo has 34 neighborhoods, Chicago has 247 neighborhoods and Sheffield has 59 neighborhoods

Note that Cairo has more neighborhoods than which exists in Wikipedia, but we assume those are the whole neighborhoods as a Wikipedia reference.

Importing Data

Importing data is divided into 3 stages. The first stage is getting list of neighbourhoods of the four cities from the above Wikipedia links. The second stage is getting location of neighbourhoods. The final stage is getting the venues in the neighbourhoods from Foursquare.

Neighbourhood lists of Cities

We already have the links of the Wikipedia pages from which we can get the list of neighbourhoods in each city. Beautiful Soup library is used to extract the information from the wikitables in the pages. This data is stored in a pandas dataframe. Along with the neighbourhoods, the city name, the state name and the country name are stored.

Geolocation of the Neighbourhoods

The geopy library is used to get the location data of the Neighbourhoods. Now in geopy library, Nominatim service is used. For using the free service of Nominatim, there is a restriction of 1call per sec to the service. To avoid 'timeout' error, there needs to be at least 1 sec gap in each call even if a for loop is used. To provide a sufficient gap to accommodate network delay, a gap of 2 sec is provided. The gap is provided by calling the sleep function.

Timeout Errors

Even after providing a 2 sec gap in calls, there are timeout errors. So, to handle these errors is simple. Simply call the Nominatim service again for these locations after checking for network connectivity.

Missing/No Coordinates

Some locations will not resolve into coordinates. This can happen because some locations may have different spellings. These can be rectified by using different spellings. Some locations will not resolve despite that. Then that data is procured manually searching on Google Maps.

Maps

Maps are generated for each city with neighbourhoods shown as markers. But before that a world map is created with the cites as markers. Below is the world map.
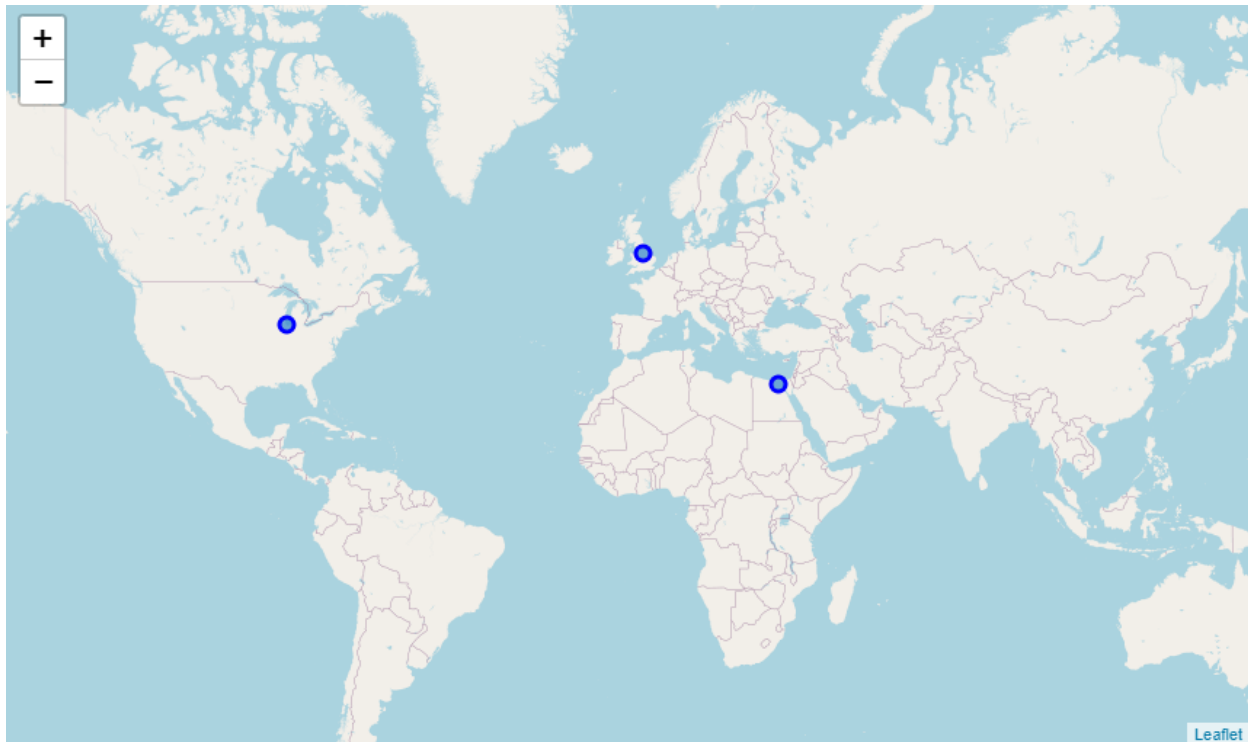
Maps of the cities are below.
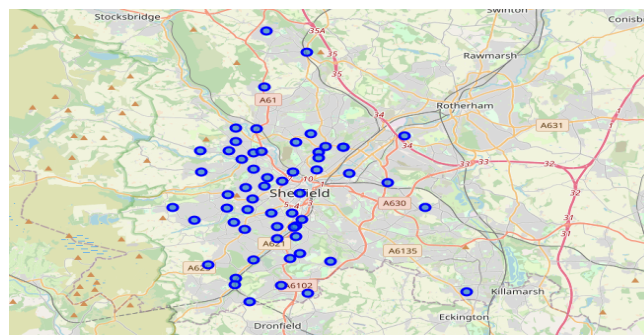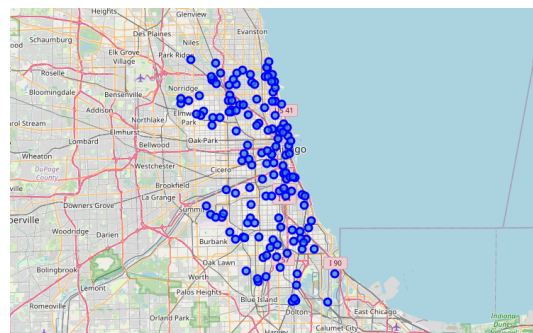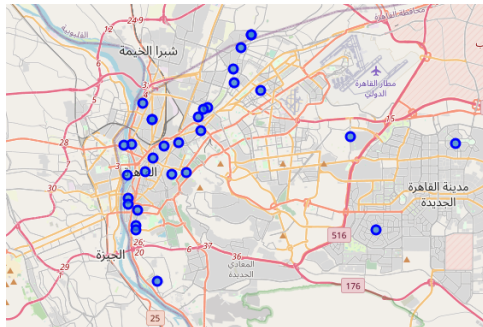
Venues

Getting places

Getting places located in the neighbourhoods like restaurants, hotels, cafes, parks etc. Foursquare API was used get the list of places for a neighbourhood. Since free version of the Foursquare API is used, a maximum of 50 places can be retrieved.

Studying Venues

To study the places, the dataframes containing the venues are grouped by neighbourhoods and summed up.

For Cairo, some of places don't have the category type in the returned data of foursquare API.

For individual clustering, one hot encoding is done for neighbourhoods of each location. While for complete clustering, all venues are combined into the same dataframe and then one hot encoding is done. This makes sure that all the types of venues are considered.

Individual Clustering Results

Individual Clustering will help understand how the individual locations can be clustered. To be consistent with all the individual location clustering and the complete clustering, there are going to be 8 clusters.

It must be noted that cluster labels are not the same across different locations. So, Cluster 0 in Bengaluru is not the same as Cluster 0 in San Francisco.

Cairo

Looking at the clustering of the neighbourhoods in Cairo, the majority of the neighbourhoods in cluster 5 . The neighbourhoods of all clusters do not look equally distributed on the map.
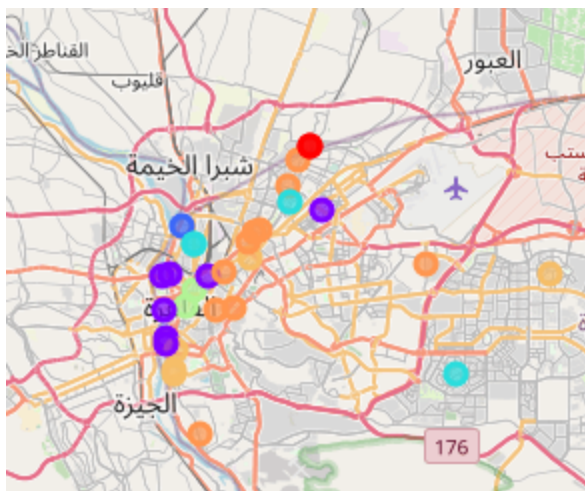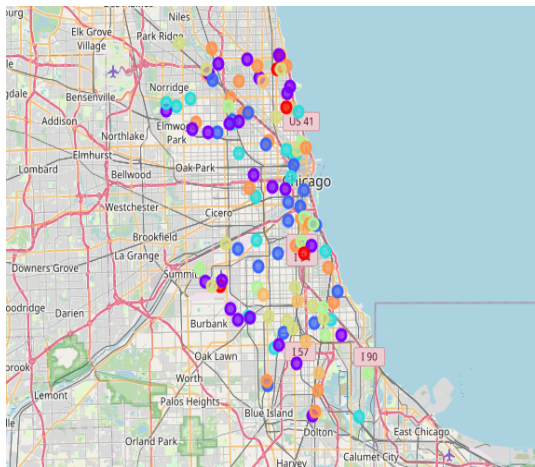

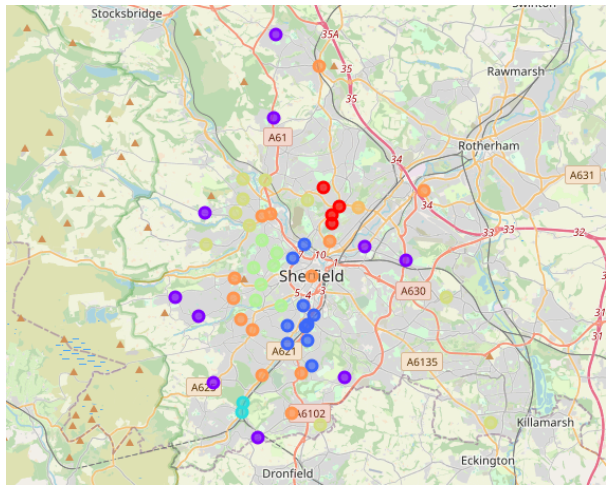| Cluster Labels	
0	2
1	7
2	1
3	1
4	5
5	10
6	4
7	2

Chicago

In Chicago, The neighbourhoods are equally distributed, Most of the neighborhoods are in cluster 1 .

Cluster Labels	
0	10
1	47
2	33
3	27
4	22
5	39
6	38
7	27

Sheffield

In Sheffield map we notice that the neighborhoods on the edges belongs to the same cluster and those which in the middle look like groups which make since, the edges neighborhoods have same features and every group of neighborhoods near of each others have same features.

Cluster Labels	
0	4
1	10
2	10
3	10
4	1
5	13
6	2
7	7

Complete Clustering Results

The complete clustering gives some interesting results. The most common cluster between the 3 cities is cluster 4,

The big takeaway from this is that there are three clusters with Neighbourhoods from all the locations (0,1,3,4,6). So, these Neighbourhoods can be considered similar based on the venues present in them.

Discussion

The objective of this analysis was that if there is a restaurant franchise in Cairo and we want to open a new franchise in Chicago and Sheffield then in which Neighbourhoods of the cities they should open. Based on Complete Clustering Neighbourhoods in clusters 0,1,3 and 6 are similar Neighbourhoods. So, if the restaurant in the Neighbourhoods of these clusters in Chicago and Sheffield then a new franchise can be opened in the Neighbourhoods of the same clusters in Cairo. Since majority of the Neighbourhoods of all the locations are in these three clusters then there is a good probability of finding a match.

Conclusion

The result showed that the restaurant franchise can be opened in Chicago and Sheffield in places which in cluster 0, 1, 3 and 6, though more data and analysis is needed. More data like the customer rating and pricing details will help but with the free Foursquare API there is limited access to the required data.

Cluster Labels	City	
0	Sheffield	5
	cairo	5
	chicago	40
1	Sheffield	8
	cairo	4
	chicago	22
2	Sheffield	4
	chicago	18
3	Sheffield	3
	cairo	5
	chicago	21
4	Sheffield	18
	cairo	11
	chicago	54
5	Sheffield	3
	chicago	19
6	Sheffield	8
	cairo	6
	chicago	44
7	Sheffield	8
	cairo	1
	chicago	25