Grouping the world's most populous Metropolitan cities

By G.Mukkesh

Introduction

With the advent of technology, the world has now become a global village. People living in different metropolitan cities of the world are being exposed to the cultures of various nations in the forms of movies, books, tv shows and restaurants, and are able to experience them without actually visiting the country. The converse also holds good, i.e., a person who has immigrated to another country can still experience his own country's culture in the major metropolitan cities.

The purpose of this study is to use the FourSquare location data to group different metropolitan cities into clusters, depending on how similar/dissimilar they are to each other. By doing so, we will be able to find out the cities which offer a similar lifestyle in terms of the amenities present. This would be useful to people who are moving to different countries, since they would know how similar/different living in the new city would be, compared to their previous city.

Data Set

The 100 most populous metropolitan cities of the world were scraped from Wikipedia (https://en.wikipedia.org/wiki/List_of_metropolitan_areas_by_population).

Using the geocoder module in python, we get the latitudes and longitudes of the 100 cities that were scraped. Feeding these coordinates into the FourSquare API, we can get the top 10 venues for the cities. After getting the top 10 locations for every city, we can get the set of unique venue categories and then one-hot-vector encode them, after grouping all the venues by their city. After the above-mentioned transformations are done to the collected data, it can be used for further analysis, since it is now clean and pre-processed.

	Rank	Metropolitan	Country	lat	long
0	1	Tokyo	Japan	35.676192	139.650311
1	2	Delhi	India	28.704059	77.102490
2	3	Shanghai	China	31.230416	121.473701
3	4	Jakarta	Indonesia	-6.208763	106.845599
4	5	Seoul	South Korea	37.566535	126.977969

```
df.shape
```

```
(100, 5)
```

[47]:
```python
city_grouped = city_onehot.groupby('City').mean().reset_index()
city_grouped
```

[47]:

	City	Accessories Store	Afghan Restaurant	African Restaurant	Airport Lounge	Airport Terminal	American Restaurant	Amphitheater	Antique Shop	Aquarium	...	Wine Bar	Wine Shop	Winery	Wings Joint	Women's Store	Y Rest
0	Ahmedabad	0.00	0.00	0.00	0.000000	0.0	0.000000	0.00	0.00	0.00	...	0.00	0.00	0.0	0.00	0.00	
1	Alexandria	0.00	0.00	0.00	0.000000	0.0	0.014085	0.00	0.00	0.00	...	0.00	0.00	0.0	0.00	0.00	
2	Ankara	0.00	0.00	0.00	0.000000	0.0	0.000000	0.00	0.04	0.00	...	0.00	0.00	0.0	0.00	0.00	
3	Atlanta	0.00	0.00	0.00	0.000000	0.0	0.030000	0.00	0.00	0.00	...	0.00	0.01	0.0	0.00	0.00	
4	Bandung	0.00	0.00	0.00	0.000000	0.0	0.000000	0.00	0.00	0.00	...	0.00	0.00	0.0	0.01	0.00	
5	Bangalore	0.00	0.01	0.00	0.000000	0.0	0.010000	0.00	0.00	0.00	...	0.01	0.00	0.0	0.00	0.00	
6	Bangkok	0.00	0.00	0.00	0.000000	0.0	0.000000	0.00	0.00	0.00	...	0.00	0.00	0.0	0.00	0.00	
7	Barcelona	0.00	0.00	0.00	0.000000	0.0	0.000000	0.00	0.00	0.00	...	0.04	0.01	0.0	0.00	0.02	
8	Beijing	0.00	0.00	0.00	0.000000	0.0	0.010000	0.00	0.00	0.00	...	0.00	0.00	0.0	0.00	0.00	
9	Belo Horizonte	0.01	0.00	0.00	0.000000	0.0	0.000000	0.00	0.00	0.00	...	0.00	0.00	0.0	0.00	0.00	
10	Berlin/Brandenburg	0.00	0.00	0.00	0.000000	0.0	0.000000	0.00	0.00	0.00	...	0.01	0.01	0.0	0.00	0.00	
11	Bogotá	0.00	0.00	0.00	0.000000	0.0	0.010000	0.00	0.00	0.00	...	0.00	0.00	0.0	0.02	0.00	