Exploring Neighborhoods near Airports across India

Apoorva S

July 24, 2020

Data Section

The airport dataset is collected from the website "data.world". The dataset has information about all the airports in India. The data set consists of nearly 344 airport names, location, type of airport, region code, official website link, wikipedia link, score, etc.

Using the latitude and longitude information of an airport we try to fetch the neighborhood locations using foursquare api.

Approach

As a database, I have used IBM watson studio to store the csv file and notebook as an editor.

1. Collect the airport dataset

```
df.columns

]: Index(['id', 'ident', 'type', 'name', 'latitude_deg', 'longitude_deg',
         'elevation_ft', 'continent', 'iso_country', 'iso_region',
         'municipality', 'scheduled_service', 'gps_code', 'iata_code',
         'local_code', 'home_link', 'wikipedia_link', 'keywords', 'score',
         'last_updated'],
       dtype='object')
```

2. Preprocess the dataset as there are junk values, duplicate values and missing values.

3. Using Foursquare API we will get all venues for each neighborhood and filter out the required fields

In [13]: nearby.head()

Out[13]:

	Name	Municipality	Latitude	Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Indira Gandhi International Airport	New Delhi	28.5665	77.103104	Arrival Terminal	28.564559	77.107258	Airport Terminal
1	Chhatrapati Shivaji International Airport	Mumbai	19.0886993408	72.8678970337	Chhatrapati Shivaji International Airport	19.090509	72.865148	Airport
2	Chhatrapati Shivaji International Airport	Mumbai	19.0886993408	72.8678970337	Starbucks	19.091774	72.868675	Coffee Shop
3	Chhatrapati Shivaji International Airport	Mumbai	19.0886993408	72.8678970337	BOM/VABB Runway 09/27	19.088076	72.864418	Airport Service
4	Chhatrapati Shivaji International Airport	Mumbai	19.0886993408	72.8678970337	Apron Control	19.091559	72.865813	Airport

4. Analyse using K-Means Clustering

i) Find the optimal k value using elbow method
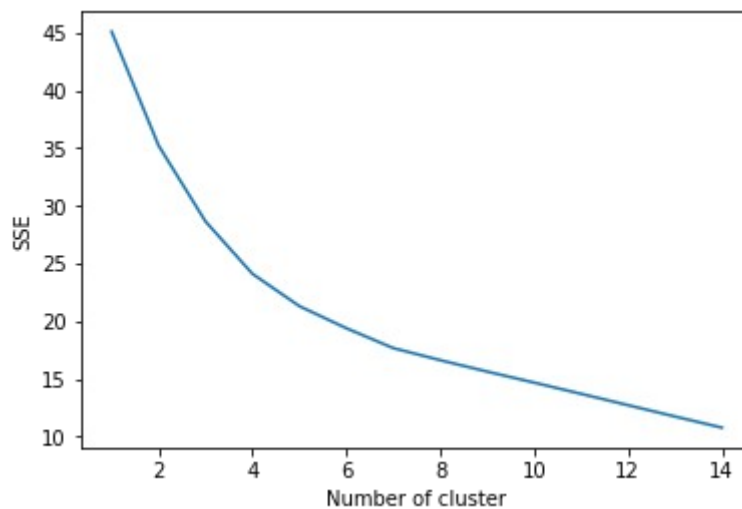
ii) Find k clusters (I have taken k value as 4)
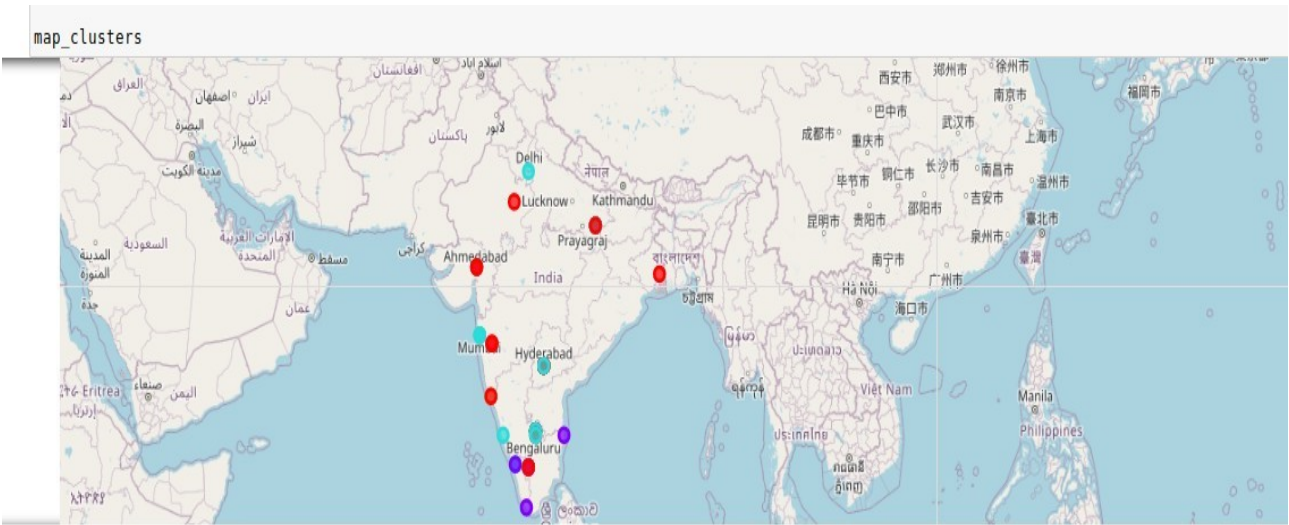
```
In [55]: nearby_grouped.groupby('Cluster')['Municipality'].count()

Out[55]: Cluster
         0    23
         1     7
         2    32
         3     8
         Name: Municipality, dtype: int64

In [56]: nearby_grouped.shape

Out[56]: (70, 73)
```

iii) Plot the clusters using folium maps

5. Visualise the neighborhoods
I have used the categories: Bar, Bakery, Hotel, Gift shops which are nearby airport locations.

Upper chart (Bar):

State	Value
Kohima	1.0
Upper Tadong	1.0
Lucknow	1.0
Jaipur	1.0
Bilaspur	1.0
Bengaluru	1.0
Mangalore	1.0
Vijay Nagar	0.50
Mumbai	0.42
Pelling City	0.40

Lower chart (Shops):

State	Value
Mathura	1.0
Guwahati	0.20
Patna	0.20
Mumbai	0.05
Bangalore	0.04
Port Blair	0.0
Porbandar	0.0
Poonch	0.0
Pondicherry	0.0
Pelling City	0.0