Project Report

on

Exploring Neighborhoods near Airports across India

Apoorva S

July 24, 2020

Introduction

This project focuses on the popular neighborhoods around international airports located in India. There are may different categories explored in the dataset. The main categories of business focused in this project are hotels, bakery, gift shops and bars nearby airports within the radius of 500 meters. Foursquare API is used to fetch the nearby places of airport.

This project tries to find the answer for the question: what kind of business can be opened near airports in India?

Target audience

This project is for any business personnel who wishes to open new shops nearby airports. As different categories of venues are explored, any sector of business personnel can refer to this application. This is not restricted to any particular type of business.
It also covers airport data of whole India, which gives more option for everyone to refer to this project.

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
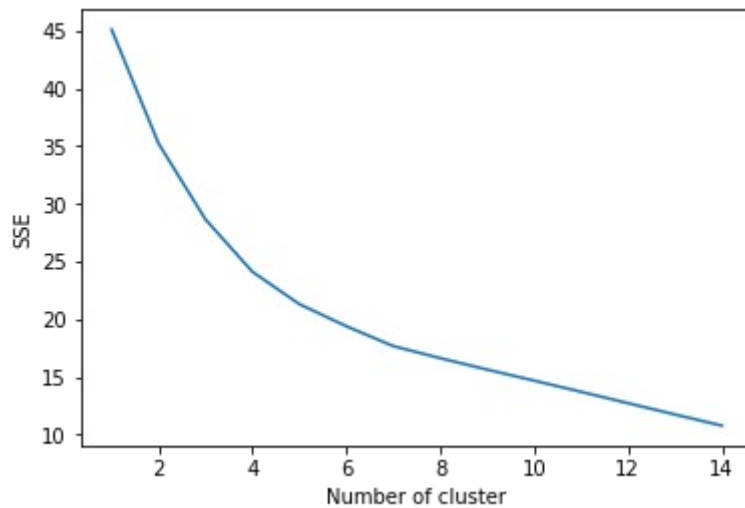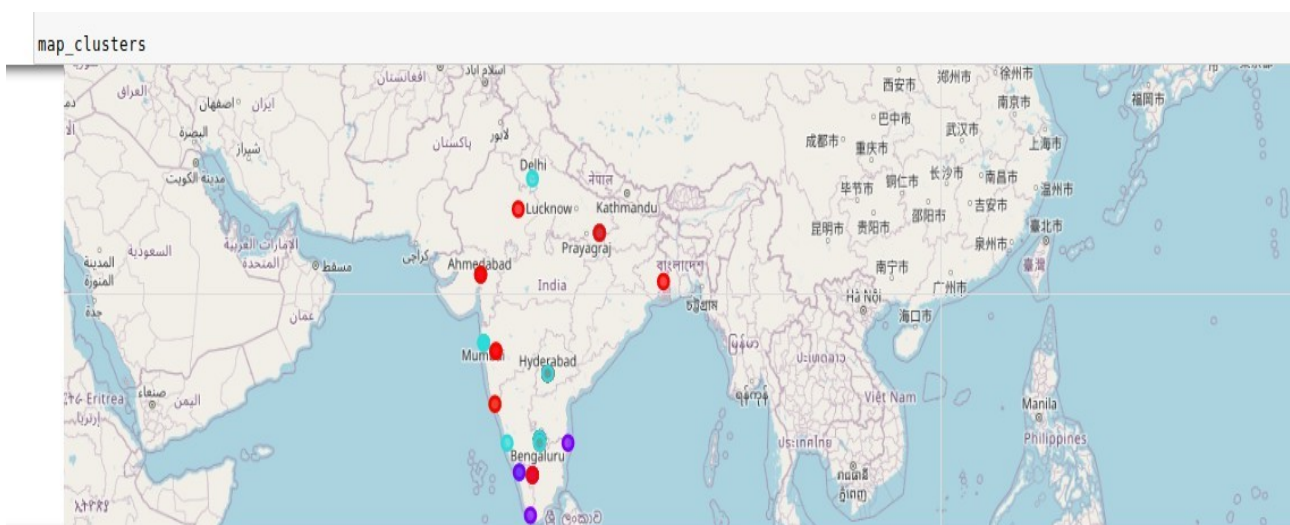
ii) Find k clusters (I have taken k value as 4)

```
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)
```
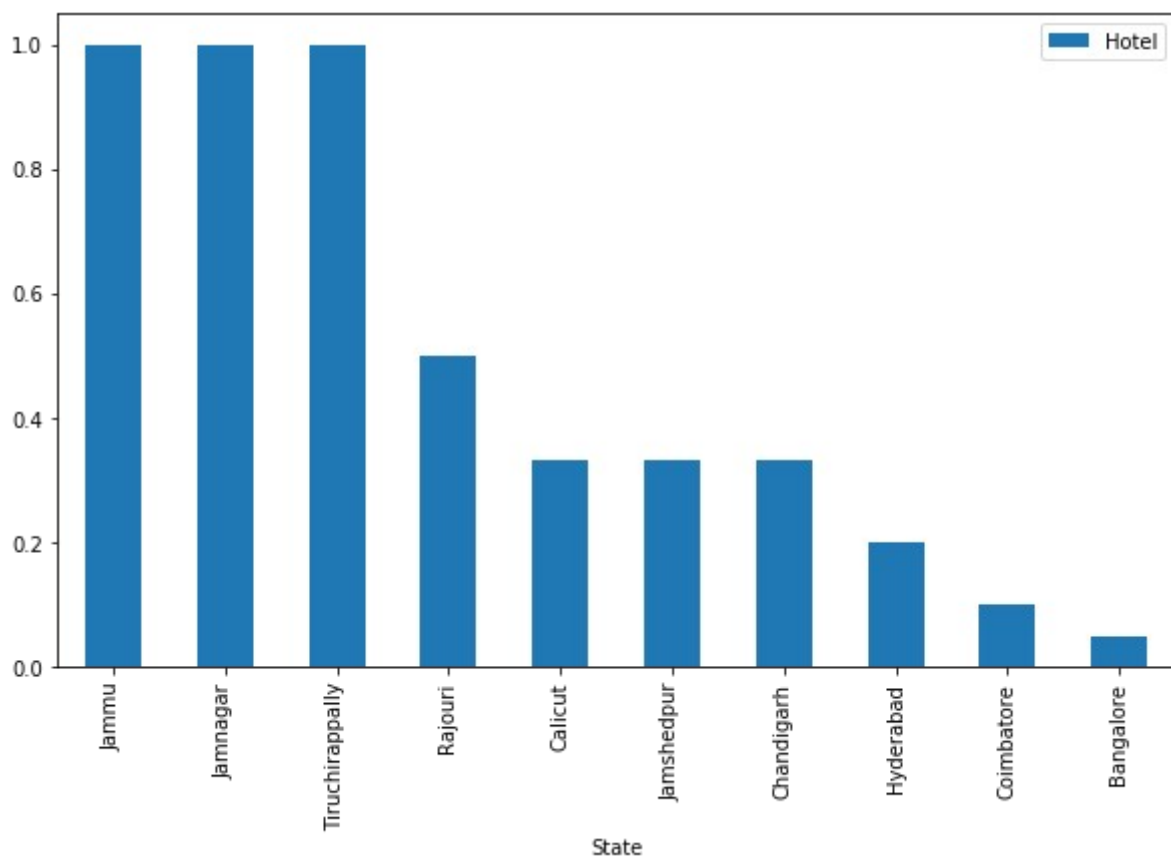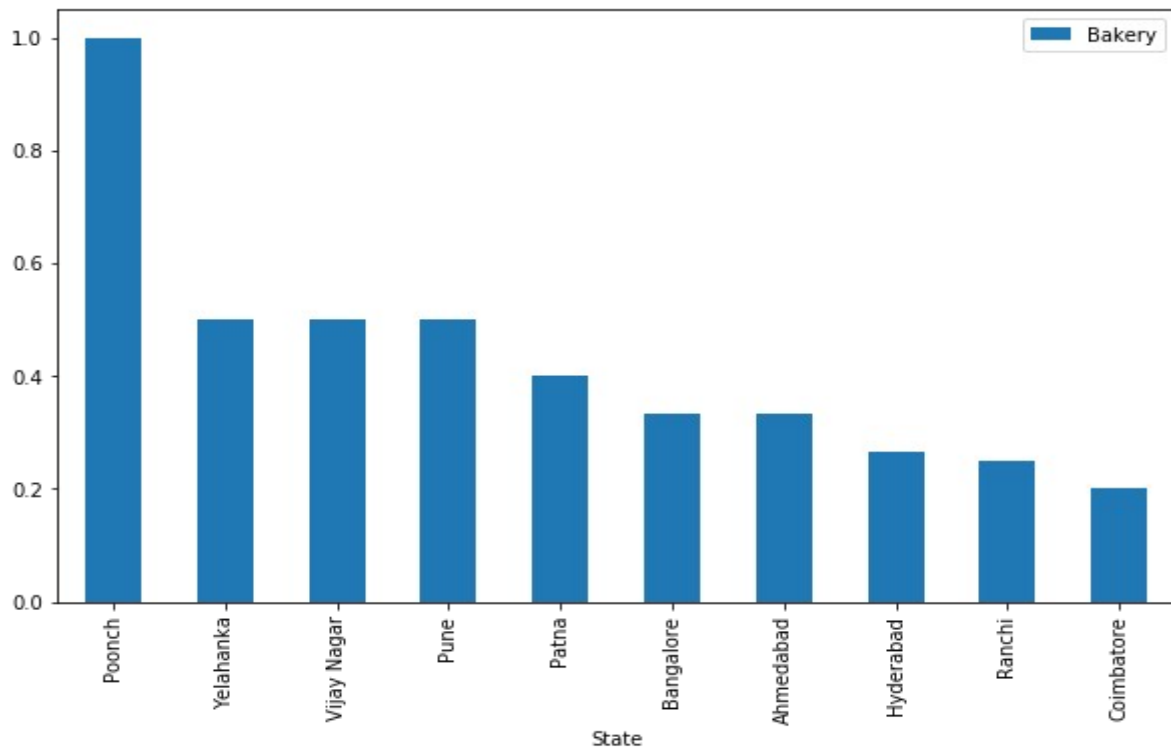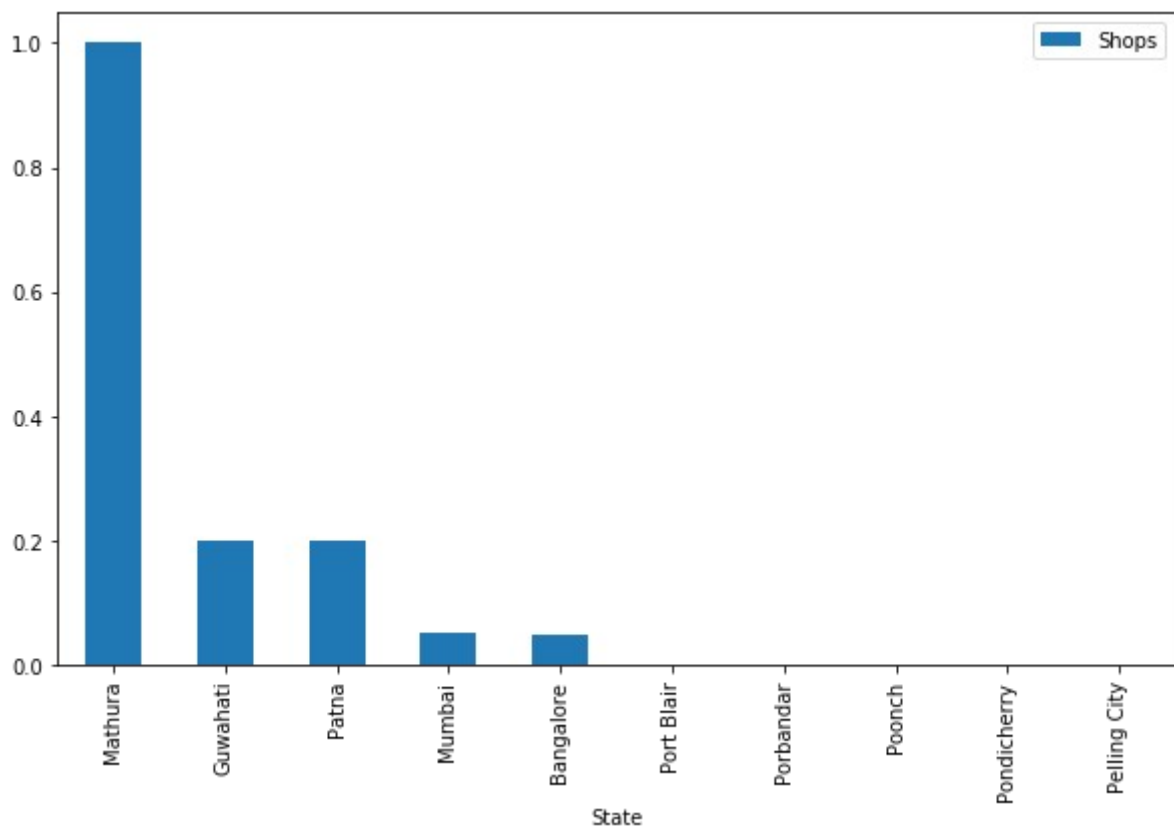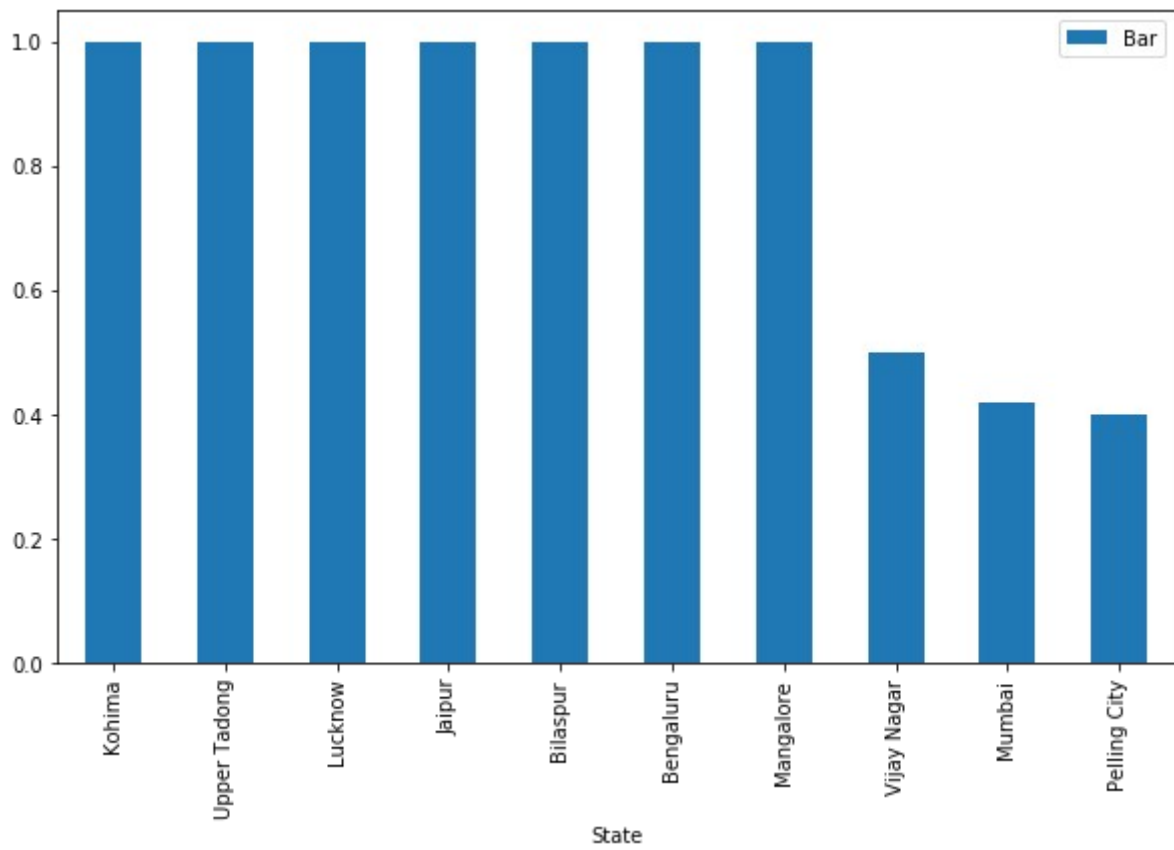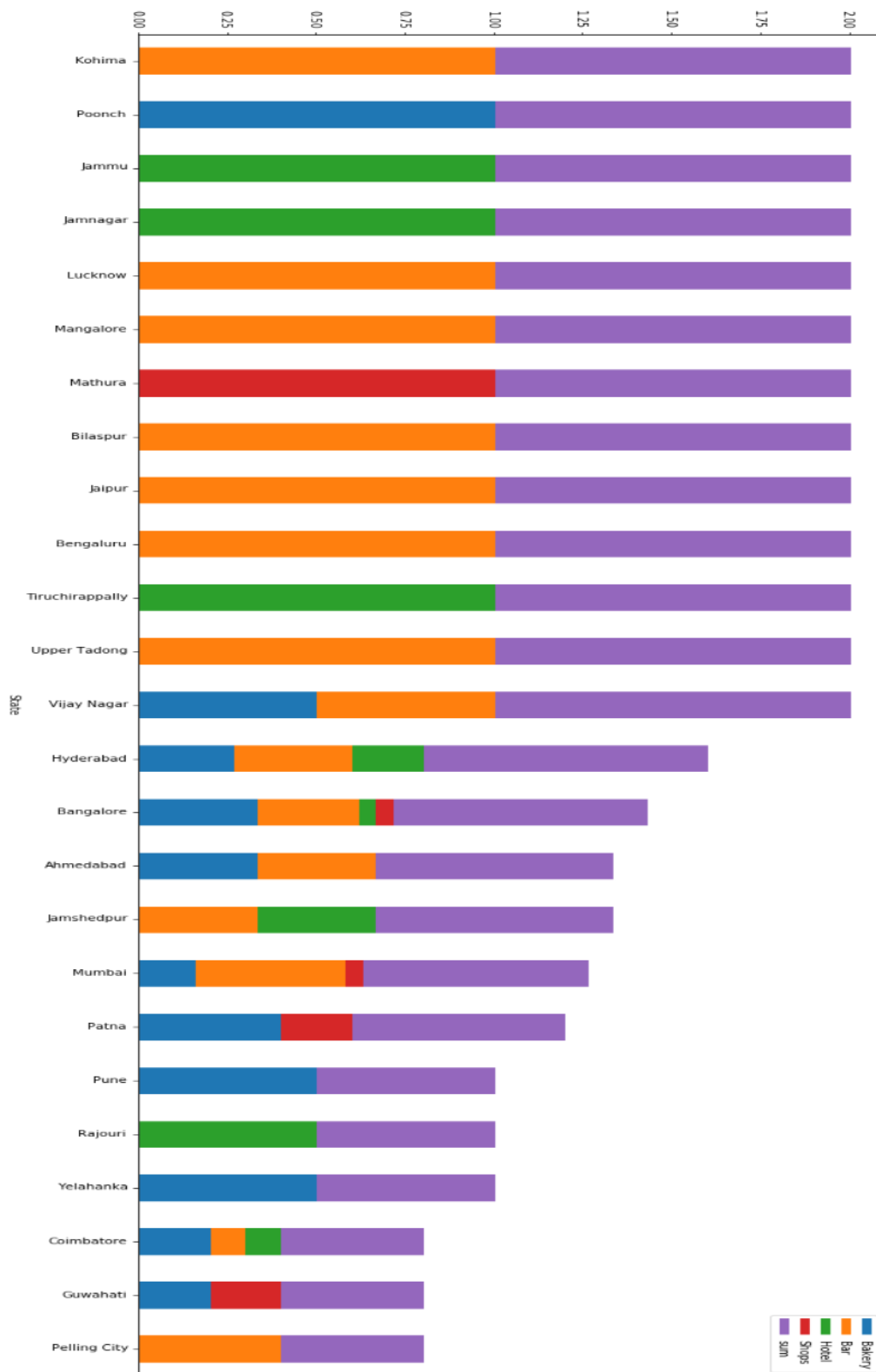
iii) Plot the clusters using folium maps

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

Result

The result of data analysis and clustering are summarised as below:
i) Kohima has more number of bars near airport in India
ii) Poonch has more number of bakery near airport in India
iii) Jammu and Jamna Nagar has more number of hotels near airport in India
iv) Mathura has more gift shops near airport in India
v) Bangalore airports has all kinds of facilities available near to it in India
vi) Cluster 0 marked in red on map has less number of facilities near airports.

Discussion

Even though I have tried to implement some of data analysis techniques, there is still more scope to improve this project. The study can be extended for other places and more clarity can be given to the venue category which helps the business planning process in different sectors.

Conclusion

Finally, to conclude this project, we saw how real-life Data science project looks like. I have used some frequently used python libraries like pandas for data analysis, matplotlib to plot graphs, seaborn for data visualization, folium to plot maps and sklearn library for k-means clustering. I have also uses Foursquare API to neighborhoods given the airport locations. This kind of analysis in a real-life business problem is discussed.