COURSERA CAPSTONE

IBM Applied Data Science Capstone

Opening a new Chinese Restaurant in India

By: Venkata Sai Pavan Kagitha

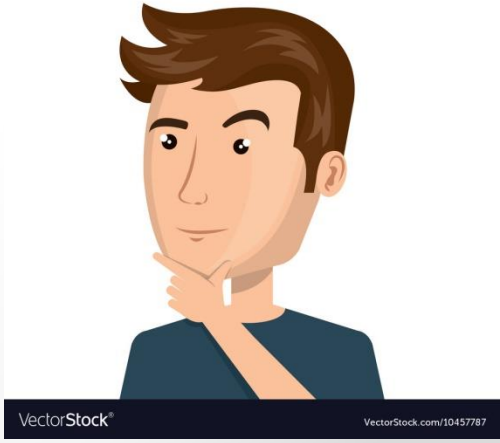
April 2020

Business Problem

- Objective: To analyze the best locations in India to open a new Chinese Restaurant.

- Business Question:

In India, if a Property developer and investor are looking to open a new Chinese Restaurant, where would you recommend that they open it?



Data

- Data required
- ❑ List of cities in India
- ❑ Latitude and longitude coordinates of these cities
- ❑ Venue data, particularly data related to Chinese restaurants

- Sources of Data
- ❑ Wikipedia page for cities in India (https://en.wikipedia.org/wiki/List_of_cities_in_India_by_population)
- ❑ Geo-coder package for latitudes and longitudes
- ❑ Foursquare API for venue data.

Methodology

- Web scrapping Wikipedia page for cities list
- Get latitude and longitude coordinates

Out[11]:

	Neighborhood	Latitude	Longitude
0	Mumbai	18.940170	72.834830
1	Delhi	28.634100	77.216890
2	Bangalore	12.966180	77.586900
3	Hyderabad	17.394870	78.470760
4	Ahmedabad	23.027760	72.600270
5	Chennai	13.083620	80.282520
6	Kolkata	22.570530	88.371240
7	Surat	21.185780	72.836790
8	Pune	18.504220	73.853020
9	Jaipur	26.925730	75.806590
10	Visakhapatnam[a][5]	17.728654	83.304541
11	Kanpur	26.435620	80.329860
12	Nagpur	21.157050	79.082170
13	Lucknow	26.854710	80.921350
14	Thane	19.200000	72.966670
15	Bhopal	23.264660	77.405180
16	Indore	22.716220	75.865120

Methodology

- Use Foursquare API to get venue data.
- Group data by neighborhood taking the mean of the frequency of each venue category

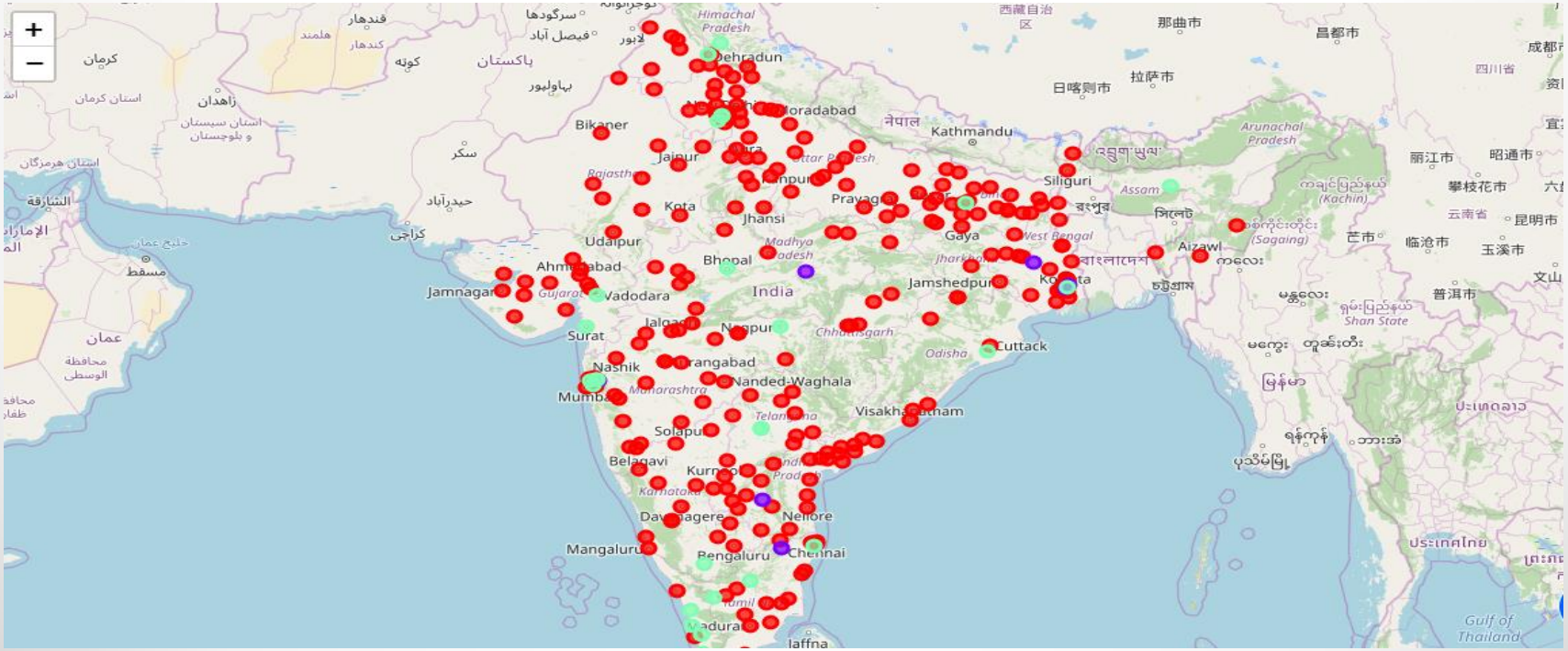
```
In [154]: ind_grouped = ind_onehot.groupby(["Neighborhoods"]).mean().reset_index()

          print(ind_grouped.shape)
          ind_grouped
```

Out[154]:

	Neighborhoods	ATM	Accessories Store	Afghan Restaurant	Airport	Airport Lounge	Airport Service	Airport Terminal	American Restaurant	Andhra Restaurant	Arcade	Art Gallery	Art Museum	Arts & Crafts Store	Asian Restaurant	Athletic & Sport
0	Adoni	0.000000	0.000000	0.000000	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.00000
1	Agartala	0.125000	0.000000	0.000000	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.00000
2	Agra	0.000000	0.000000	0.000000	0.025000	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.00000
3	Ahmedabad	0.000000	0.000000	0.000000	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.010000	0.000000	0.000000	0.010000	0.010000	0.00000
4	Ahmednagar	0.000000	0.000000	0.000000	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.00000
5	Aizawl	0.000000	0.000000	0.000000	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.00000
6	Ajmer	0.000000	0.000000	0.000000	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.00000
7	Akola	0.100000	0.000000	0.000000	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.00000
8	Alappuzha	0.000000	0.000000	0.000000	0.000000	0.000000	0.00	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.068966	0.00000

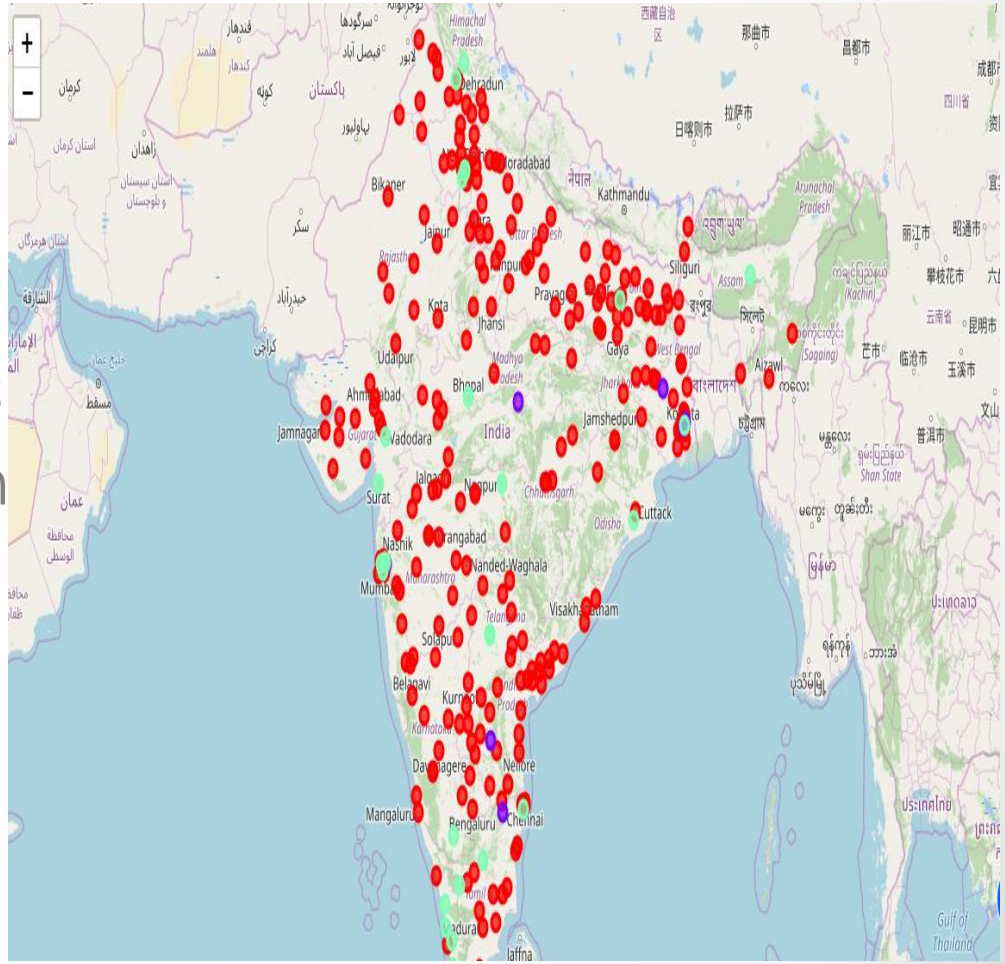
Methodology

- Filter venue category by Chinese restaurants
- Perform clustering on the data by using K-means clustering
- Visualize the clusters in a map using Folium



Results

- Categorized the neighborhoods into 3 clusters:
  - ❑ Cluster 0 – Red dots
  - ❑ Cluster 1 – Violet dots
  - ❑ Cluster 2 – Mild green



Results

- Cluster 0 – Cities with moderate number of Chinese Restaurants.

Cluster 0

```
In [165]: ind_merged.loc[ind_merged['Cluster Labels'] == 0]
```

Out[165]:

	Neighborhood	Chinese Restaurant	Cluster Labels	Latitude	Longitude
0	Adoni	0.000000	0	15.633330	77.283330
178	Miryalaguda	0.000000	0	16.861550	79.553490
179	Mirzapur	0.000000	0	25.134320	82.567070
180	Moradabad	0.000000	0	28.838930	78.776840
181	Morbi	0.000000	0	22.807770	70.823560
182	Morena	0.000000	0	26.490520	77.985720
183	Motihari[30]	0.000000	0	26.657380	84.919220
184	Mumbai	0.010000	0	18.940170	72.834830
185	Munger	0.000000	0	25.365560	86.487670
186	Muzaffarnagar	0.000000	0	29.470290	77.707610
187	Muzaffarpur	0.000000	0	26.123990	85.386690

Results

- Cluster 1 – Cities with low number of Chinese restaurants.

Cluster 1

```
In [166]: ind_merged.loc[ind_merged['Cluster Labels'] == 1]
```

Out[166]:

	Neighborhood	Chinese Restaurant	Cluster Labels	Latitude	Longitude
166	Madhyamgram	0.100000	1	22.67097	88.44321
14	Ambarnath	0.100000	1	19.19574	73.19404
120	Jabalpur	0.166667	1	23.17418	79.93136
219	Proddatur	0.111111	1	14.73333	78.55000
290	Vellore	0.076923	1	12.91356	79.13251
151	Kolkata	0.083333	1	22.57053	88.37124
85	Durgapur	0.166667	1	23.48333	87.31667

Results

- Cluster 2 – Cities with high concentration of Chinese restaurants

Cluster 2

```
In [167]: ind_merged.loc[ind_merged['Cluster Labels'] == 2]
```

Out[167]:

	Neighborhood	Chinese Restaurant	Cluster Labels	Latitude	Longitude
274	Thrissur	0.028571	2	10.51082	76.21121
287	Vadodara	0.020408	2	22.30946	73.17993
284	Ulhasnagar	0.047619	2	19.22907	73.16580
136	Kalyan-Dombivli	0.022222	2	19.23246	73.13409
264	Surat	0.058824	2	21.18578	72.83679
149	Kochi	0.021978	2	9.93601	76.26142
152	Kollam	0.029412	2	8.88689	76.59228
155	Kottayam	0.032258	2	9.59477	76.53291
116	Hyderabad	0.020000	2	17.39487	78.47076
106	Guwahati	0.045455	2	26.17724	91.75419
105	Gurgaon	0.020000	2	28.47762	77.06952
188	Mysore[8][9][10]	0.040000	2	12.30906	76.65303
191	Nagpur	0.031915	2	21.15705	79.08217
198	Navi Mumbai	0.030000	2	19.03262	73.02962
200	New Delhi	0.020000	2	28.63095	77.21721
77	Delhi	0.020000	2	28.63410	77.21689
208	Pallavaram	0.020000	2	12.97611	80.18361
70	Coimbatore	0.030000	2	10.99416	76.96629

Discussion

- Most of the Chinese restaurants are located in capital cities of the states.
- Highest number of restaurants in Cluster 2 and moderate to no restaurants in Cluster 0.
- Cluster 1 has very low number of Chinese restaurants in those cities.

Recommendation

- Can open new Chinese restaurants in the capital cities present in cluster 1 which has moderate number of restaurants.
- Can also open in other cities of cluster 1.
- Avoid cities in cluster 2, already high concentration of restaurants and intense competition.

Conclusion

- Answer to Business Question:

The cities in Cluster 1 are the most preferred locations to open a new Chinese restaurant.

- Findings of the project will help the investors and business owners to capitalize on the opportunities on high potential locations while avoiding over crowded areas in their decisions to open a new Chinese restaurant.

Thank you