

Shanghai Data Analysis

Coursera Capstone Project
by Sadasivam Natanakumar

Problem Statement/ Background Discussion

- Travel websites focus on the features within hotels while making suggestions to users.
- However, people when travelling want to explore the local culture, their cuisine, while also maintaining a certain comfort zone and keeping logistics in mind.
- This is the area of focus for my project.

Sample Use Cases

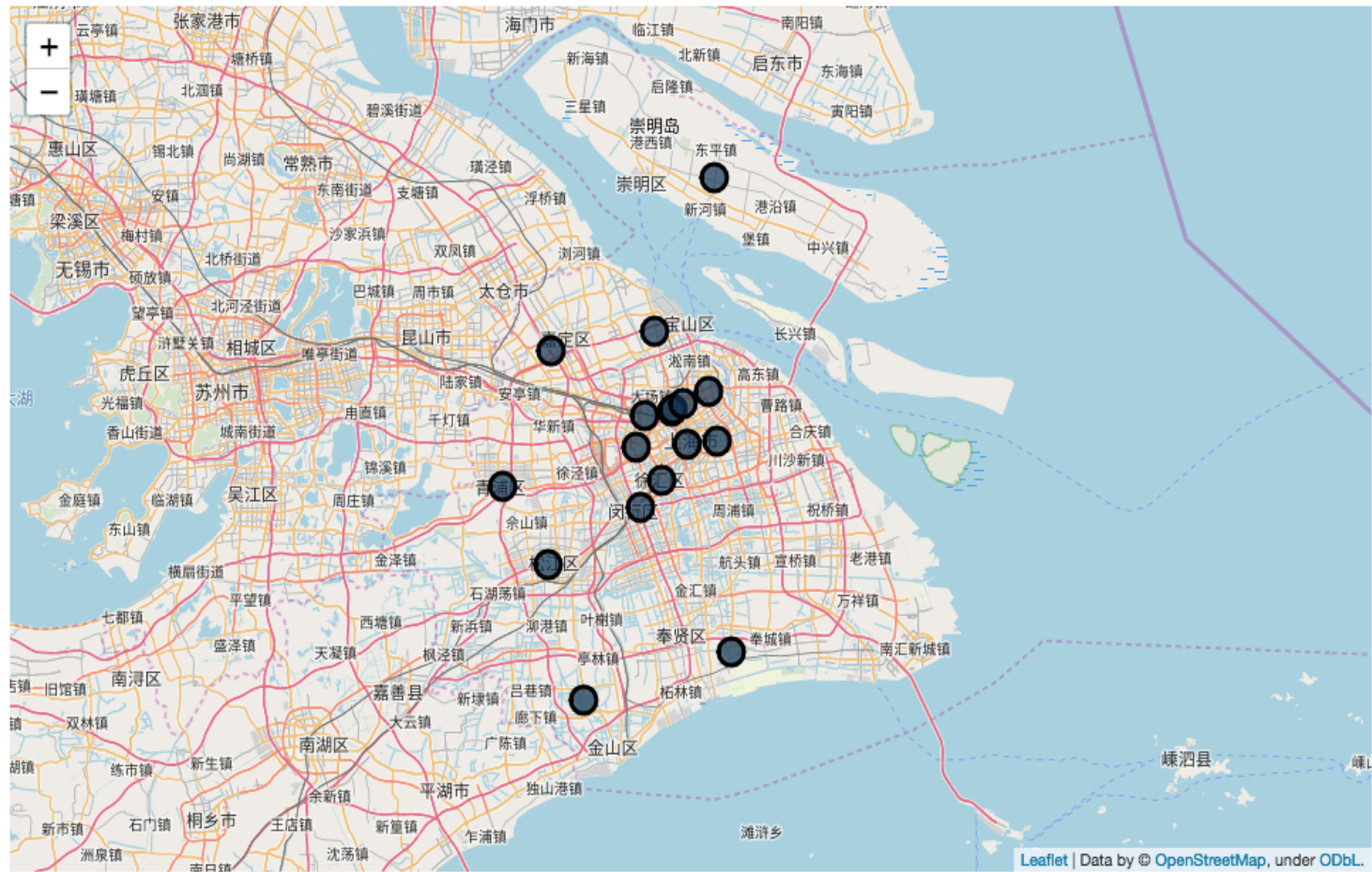
- A person planning to visit Shanghai as a tourist or on a business trip and looking for accommodation.
- For people who want ease of accessibility to the venues flocked by locals.
- To also combine the location of the hotel (distance from airport and metros) and its proximity to venues such as cafes, restaurants and spas.

Data Sources

- Hotels acquired from Foursquare data.
- Popular venues in the vicinity of the hotels, also acquired from Foursquare data.
- Distance from Airport, using data from open sources.

Postcodes and coordinates of the districts

- This was done using data from the travelchinaguide.com website which housed a table with the details of the names of districts in Shanghai along with their postcodes.
- By using simple web scraping techniques with the help of `requests` and `BeautifulSoup` libraries the data from the table was converted to a dataframe.



Folium map with all the districts indicated using black markers.

Details about the hotels and venues in Shanghai City

- This was acquired using the (foursquare.com). The API allowed me to use its 'explore' feature to get the relevant information required.
- I first made an API call to get details regarding the hotels, then made another API call to get details about venues near the hotels.
- With the first call, details regarding the hotels' name, location, coordinates, etc. was made, using the details of the districts. Each hotel was grouped by the district it was present in.

	District	Dist_Latitude	Dist_Longitude	Hotel	Hotel_Lat	Hotel_Long
1	Huangpu District	31.218228	121.480302	Fraser Residence Shanghai	31.225639	121.476177
2	Huangpu District	31.218228	121.480302	Andaz Xintiandi, Shanghai (上海新天地安达仕酒店)	31.221941	121.475431
3	Xuhui District	31.162210	121.432119	Pullman Shanghai South (中星铂尔曼酒店)	31.161600	121.425152
4	Jing'an District	31.269891	121.450685	Four Points by Sheraton Shanghai, Daning	31.273794	121.451831
5	Jing'an District	31.269891	121.450685	Shanghai Holand Hotel	31.265858	121.455452
6	Yangpu District	31.299157	121.521357	Hyatt Regency Shanghai, Wujiaochang (上海五角场凯悦酒店)	31.302408	121.514755
7	Yangpu District	31.299157	121.521357	WH Ming Hotel (小南国花园大酒店)	31.291290	121.525227
8	Pudong New Area	31.221783	121.538740	Renaissance Shanghai Pudong Hotel (上海淳大万丽酒店)	31.225783	121.548058
9	Pudong New Area	31.221783	121.538740	Parkview Hotel	31.225966	121.538728
10	Pudong New Area	31.221783	121.538740	Crowne Plaza Century Park Shanghai (上海世纪皇冠假日酒店)	31.225281	121.547203
11	Songjiang District	31.029593	121.210838	Neo-Sunshine Hotel	31.028412	121.206445
12	Songjiang District	31.029593	121.210838	Shanghai Vienna Hotel	31.036080	121.216409
13	Songjiang District	31.029593	121.210838	BaoLong Home Hotel Shanghai	31.030738	121.200523
14	Changning District	31.213015	121.382477	Ruitai Hongqiao Hotel (瑞泰虹桥酒店)	31.212176	121.386333
15	Putuo District	31.262501	121.400107	Radisson Blu Hotel Shanghai Hong Quan	31.260797	121.397849
16	Hongkou District	31.280883	121.474403	SISU Guest House (SISU Guest House Hotel Shang...	31.279578	121.477662
17	Qingpu District	31.150700	121.124200	Crowne Plaza	31.153690	121.132946

Details of the best hotels in Shanghai, grouped by their districts.

	Hotel	Hotel_Lat	Hotel_Long	Venue	Venue_Lat	Venue_Long	Venue_Category
1	Fraser Residence Shanghai	31.225639	121.476177	Mayita	31.226250	121.476113	Mexican Restaurant
2	Fraser Residence Shanghai	31.225639	121.476177	Green Massage (青籜养身)	31.223859	121.472873	Massage Studio
3	Fraser Residence Shanghai	31.225639	121.476177	Starbucks Reserve (星巴克臻选)	31.222805	121.474812	Coffee Shop
4	Fraser Residence Shanghai	31.225639	121.476177	city'super	31.226773	121.473824	Grocery Store
5	Fraser Residence Shanghai	31.225639	121.476177	Open Kitchen by Hunter Gatherer (Hunter Gather...	31.222577	121.474145	Restaurant

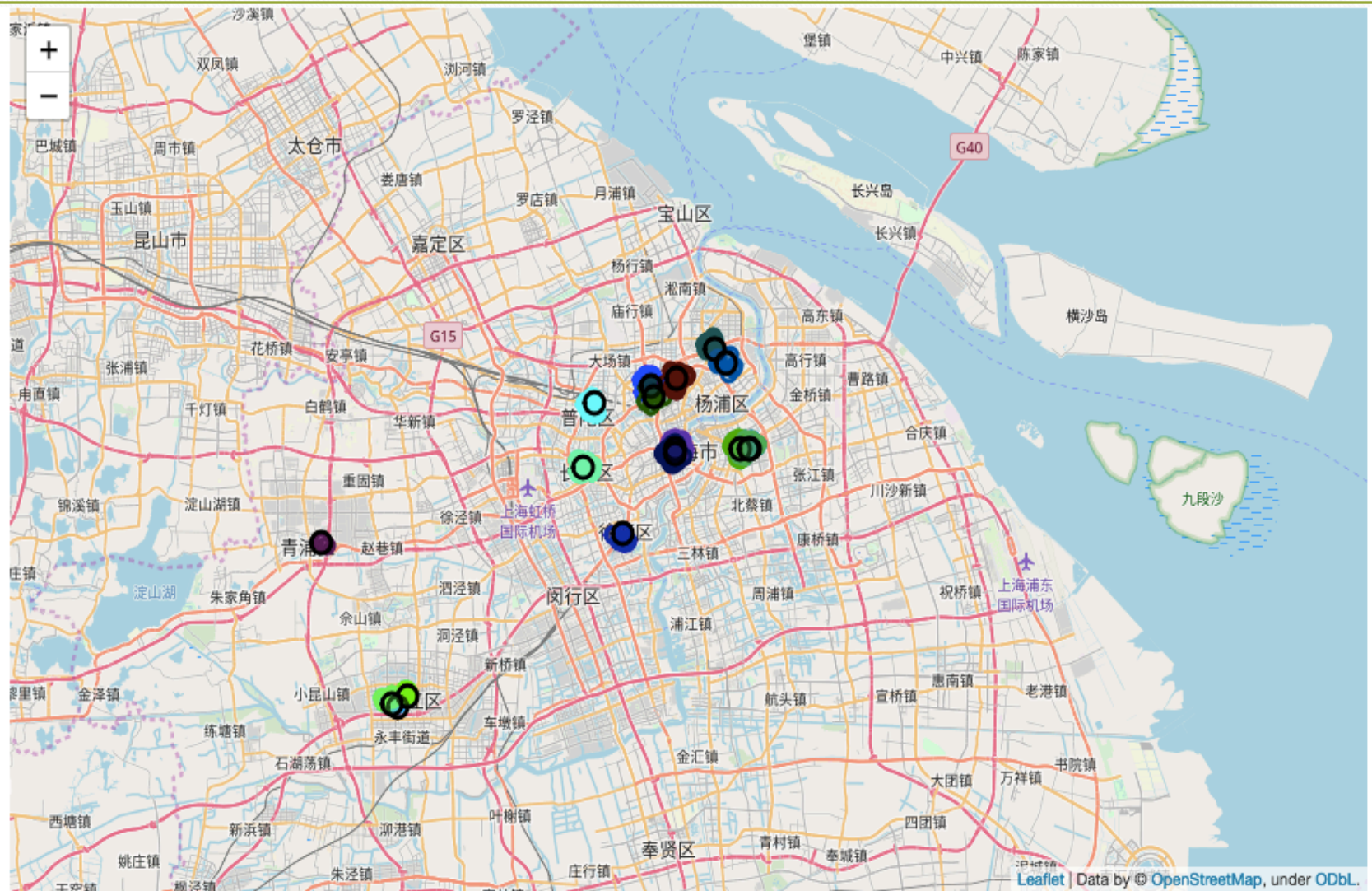
A sample from the dataframe, showing the venue details for places next to Fraser Residence hotel

Methodology

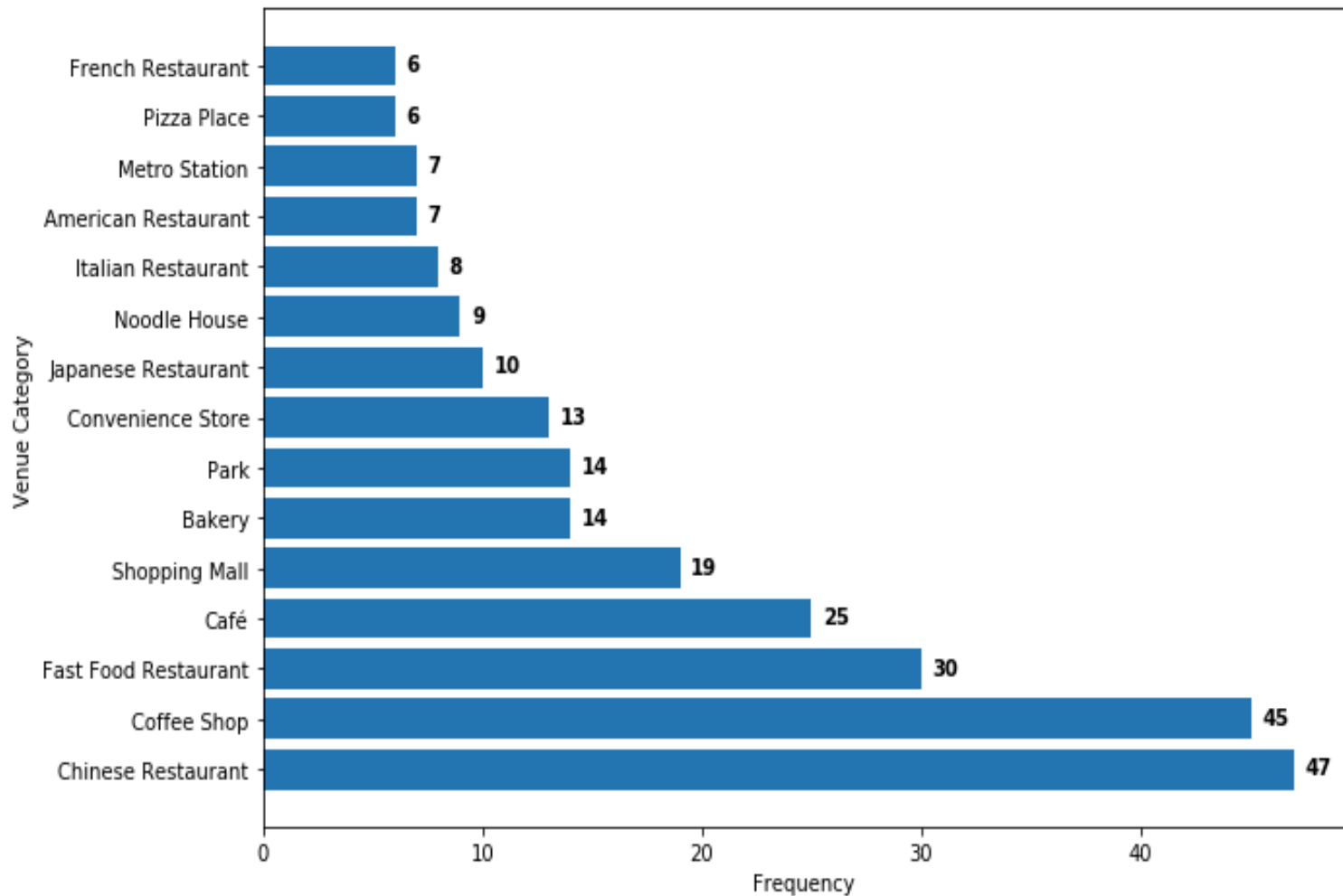
- The first task was to acquire details regarding the different postcodes of different districts using the aforementioned website. The data was scraped and saved to a dataframe.
- The second task was to get the coordinates of the district using library and then displaying them on a map.
- Following this, the first Foursquare API call was made, to get the hotels in Shanghai. This was done by first finding out all the venues, and then filtering out the ones that weren't a hotel.

Methodology

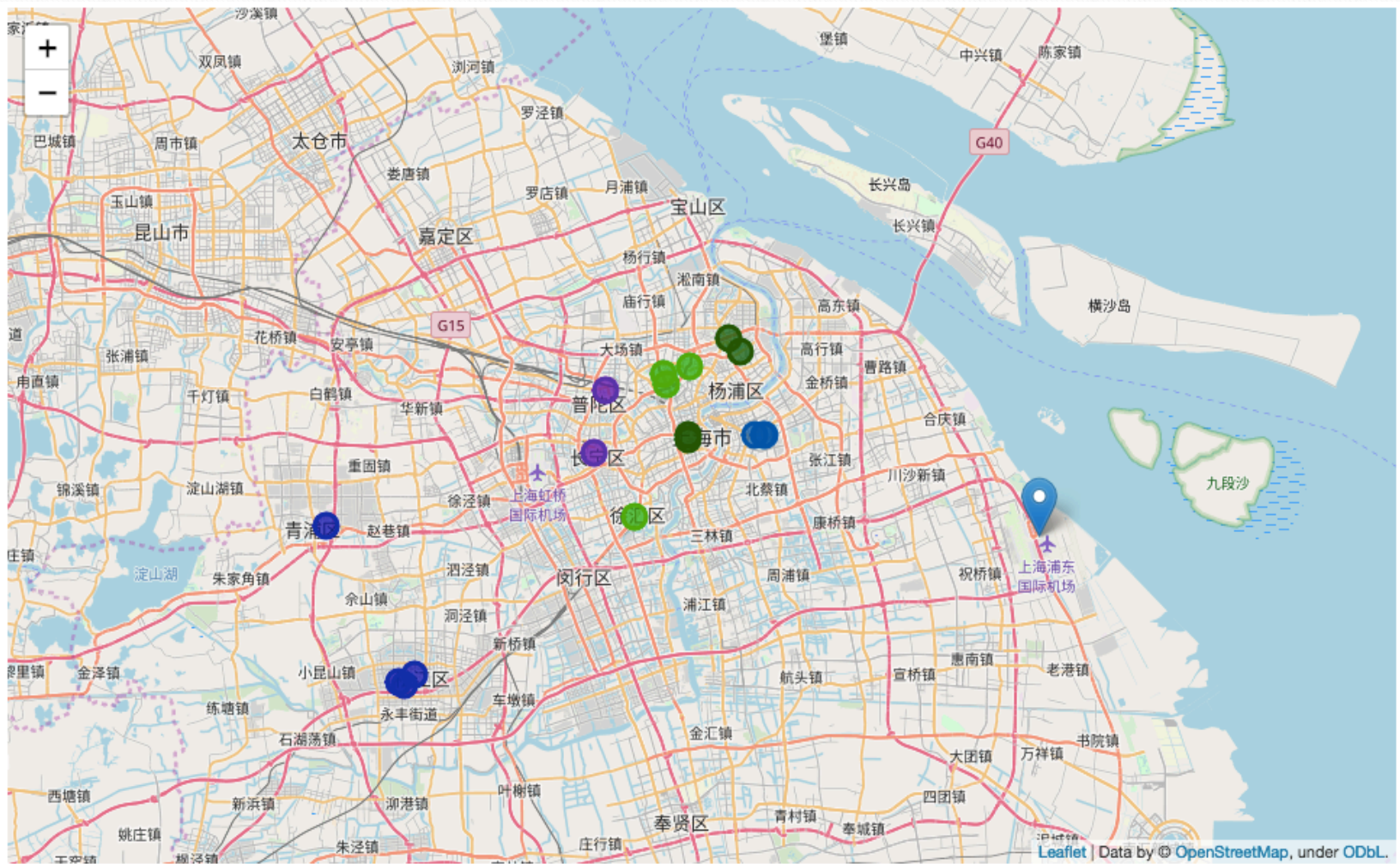
- The relevant details of the hotel were stored in a dataframe, from which the coordinate details were used to make a second API call.
- In the second API call, venues nearby the hotels were found and stored in a new dataframe. Using this data, a list of the top 15 most popular places was created.
- Using this list, onehot encoded dataframe was created to compare which hotels had more popular places nearby and the number was totalled and added to a new 'Sum' column.



A folium map showing all the venues (in multicolour) near the hotels (in black rings)



A bar graph of the 'Top 15 Most Popular Venues' in Shanghai



A folium map rendered with the 5 clusters and the Shanghai Intl Airport

Results

- The distance from the airport influences clusters. If we followed an imaginary radius with the airport as a concurrent centre, we can identify that all the hotels fit in ideal clusters that can be imagined to be on a circle.
- The venues that are nearby hotels equally influence the cluster analysis. No two hotels in a cluster (bar the first one, which indicates an outlier) have a difference of more than 0.25 in terms of their sums.
- Only 4 hotels have a sum of less than 0.5, indicating that popular venues in Shanghai are also located near the hotels.

Discussion/Conclusion

- The generated clusters indicate that there are a lot of popular venues in close proximity of the Hotels. However, this doesn't mean that one influences the other (**Correlation doesn't imply causation**).
- Travellers who are on a decent budget and want to explore the local culture will find this catalogue the most useful.
- Since this also includes a graph of most popular venues, people can identify those places and manually choose to visit them.
- The most popular venues are Chinese restaurants, Cafes in general and fast food outlets.