Final Project: Find Ideal Startup Location

Table of Contents

Introduction/Business Problem

Pakistan has recently experienced an immense number of startups in the recent years within its most popular cities such as Islamabad, Lahore, Rawalpindi and many more. These staggering numbers will only continue to grow more in the coming years which makes it difficult for someone starting their own business, to thrive in these times and make their business stand out from the rest. Consider that you are planning to initiate a startup, specifically a '**Fast Food Restaurant**' in **Pakistan** but you have no idea which location would be ideal to start with. Also considering that you only have enough investment amount to open one restaurant so you want to find a location which will generate maximum income as well as interest.

For this problem, the [Foursquare](#) API will help us identify all the locations in Pakistan where Fast-Food Restaurants are the most popular so that the investor can strategically choose a location that best fits them.

Data

Since we want to find an ideal startup spot in one of the major cities of Pakistan, we will need to have a dataset that contains information about all the cities of Pakistan including their geographical information such as longitudes and latitudes. For this, I will be using the Pakistan Cities Dataset which provides all the crucial information that will be needed to carry out this project such as city names, longitudes and latitudes. For visual purposes, a screenshot of what the dataset looks like is attached below.

Download Data in Spreadsheet (146 cities) ▾			Get all cities in Pakistan (140,909 cities)		Looking for an Pakistan map?			
city	lat	lng	country	iso2	admin_name	capital	population	population_proper
Karachi	24.8600	67.0100	Pakistan	PK	Sindh	admin	14835000	14835000
Lahore	31.5497	74.3436	Pakistan	PK	Punjab	admin	11021000	11021000
Sialkot City	32.5000	74.5333	Pakistan	PK	Punjab	minor	3893672	3893672
Faisalabad	31.4180	73.0790	Pakistan	PK	Punjab	minor	3203846	3203846
Rawalpindi	33.6007	73.0679	Pakistan	PK	Punjab	minor	2098231	2098231
Peshawar	34.0000	71.5000	Pakistan	PK	Khyber Pakhtunkh	admin	1970042	1970042
Saidu Sharif	34.7500	72.3572	Pakistan	PK	Khyber Pakhtunkh	minor	1860310	1860310
Multan	30.1978	71.4711	Pakistan	PK	Punjab	minor	1606481	1606481
Gujranwala	32.1500	74.1833	Pakistan	PK	Punjab	minor	1569090	1569090
Islamabad	33.6989	73.0369	Pakistan	PK	Islāmābād	primary	1365000	1365000
Quetta	30.1920	67.0070	Pakistan	PK	Balochistān	admin	1140000	1140000
Bahawalpur	29.3956	71.6722	Pakistan	PK	Punjab	minor	681696	681696
Sargodha	32.0836	72.6711	Pakistan	PK	Punjab	minor	602631	602631
New Mirpur	33.1333	73.7500	Pakistan	PK	Azad Kashmir	minor	523500	523500
Chiniot	31.7167	72.9833	Pakistan	PK	Punjab	minor	477781	477781
Sukkur	27.6995	68.8673	Pakistan	PK	Sindh	minor	476776	476776
Larkana	27.5600	68.2264	Pakistan	PK	Sindh	minor	435817	435817
Shekhupura	31.7083	74.0000	Pakistan	PK	Punjab	minor	411834	411834
Jhang City	31.2681	72.3181	Pakistan	PK	Punjab	minor	365198	365198
Rahimyar Khan	28.4202	70.2952	Pakistan	PK	Punjab		353203	353203
Gujrat	32.5736	74.0789	Pakistan	PK	Punjab	minor	328512	328512
Kasur	31.1167	74.4500	Pakistan	PK	Punjab	minor	314617	314617
Mardan	34.1958	72.0447	Pakistan	PK	Khyber Pakhtunkh	minor	300424	300424
Mingaora	34.7717	72.3600	Pakistan	PK	Khyber Pakhtunkh		279914	279914

Table 1: Pakistan Cities Dataset

Once I have gathered all the information that I need about the investor's cities of interest, I will leverage the Foursquare API and use it to explore those cities using the geographical data obtained from the Pakistan Cities Dataset and provide visuals highlighting all the similar Fast-Food Restaurants within the cities of interest so that the investor can choose which ever location is the most suitable for their business.

Methodology

Data Cleaning:

Since the dataset which was going to be used wasn't going to be very large, because we only needed the data for every city of Pakistan from which we were going to choose the top cities of interest, there was barely any data cleaning involved. Though we did reduce the total number of cities which we were going to explore (initially 146), to only the few major cities of Pakistan which are displayed in the table below.

	city	lat	lng	country	iso2	admin_name	capital	population	population_proper
0	Karachi	24.8600	67.0100	Pakistan	PK	Sindh	admin	14835000.0	14835000.0
1	Lahore	31.5497	74.3436	Pakistan	PK	Punjab	admin	11021000.0	11021000.0
3	Faisalabad	31.4180	73.0790	Pakistan	PK	Punjab	minor	3203846.0	3203846.0
4	Rawalpindi	33.6007	73.0679	Pakistan	PK	Punjab	minor	2098231.0	2098231.0
5	Peshawar	34.0000	71.5000	Pakistan	PK	Khyber Pakhtunkhwa	admin	1970042.0	1970042.0
7	Multan	30.1978	71.4711	Pakistan	PK	Punjab	minor	1606481.0	1606481.0
9	Islamabad	33.6989	73.0369	Pakistan	PK	Islāmābād	primary	1365000.0	1365000.0
10	Quetta	30.1920	67.0070	Pakistan	PK	Balochistān	admin	1140000.0	1140000.0

Table 2: Cities of Interest

Now that we have the data of all the cities that we are interested in, we are going to perform 'Feature Selection' to extract only the necessary columns that we will need for our exploration.

Feature Selection:

The dataset contained a few features we weren't going to need so I dropped those features. The following table shows the feature selection along with the reasoning behind it.

Kept Features	Dropped Features	Reason for Dropping
• City Name • City Longitude • City Latitude	• Country, iso2 • Admin • Capital, Population	• Country and country code are already known • Province is not needed • Not needed

Table 3: Simple Feature Selection

The following table is what the data looked like after we selected the features that we needed.

	city	lat	lng
0	Karachi	24.8600	67.0100
1	Lahore	31.5497	74.3436
2	Faisalabad	31.4180	73.0790
3	Rawalpindi	33.6007	73.0679
4	Peshawar	34.0000	71.5000
5	Multan	30.1978	71.4711
6	Islamabad	33.6989	73.0369
7	Quetta	30.1920	67.0070

Table 4: Simple Feature Selection Dataset

Exploring Cities using Foursquare API:

After I performed Feature Selection, the next thing to do was to explore venues in each city one by one through the Foursquare API and to filter the results to only venues that are 'Fast Food Restaurants' so that the investor would get an idea of where each restaurant is located. The locations returned by each city were stored in a separate dataset as shown in the table below.

City_Name	City_Lat	City_Lng	Venue_Name	Venue_Category	Venue_Lat	Venue_Lng	Venue_Address
Karachi	24.8600	67.0100	Karachi Broast	Fast Food Restaurant	24.826819	67.026328	(Boat Basin, Kh-e-Roomi,)
Karachi	24.8600	67.0100	Roasters	Fast Food Restaurant	24.862709	67.054954	(Sindhi Muslim Cooperative Housing Society,)
Karachi	24.8600	67.0100	Hanifia	Fast Food Restaurant	24.881117	67.042152	(Binori Town,)
Karachi	24.8600	67.0100	Tooso	Fast Food Restaurant	24.882845	67.066238	(Bahadur Shah Zafar Road,)
Karachi	24.8600	67.0100	Milano	Fast Food Restaurant	24.816153	67.040003	(Zamzama Blvd,)
Karachi	24.8600	67.0100	Bilal Broast	Fast Food Restaurant	24.881119	67.042395	(12 New Town, Jama Masjid,)
Karachi	24.8600	67.0100	KFC	Fast Food Restaurant	24.826852	67.027390	(Boat Basin,)
Karachi	24.8600	67.0100	KBC Restaurant	Fast Food Restaurant	24.856892	67.030592	Karachi
Karachi	24.8600	67.0100	Subway	Fast Food Restaurant	24.827225	67.030201	(Shop #39, Hashoo Terrace, Boat Basin,)
Karachi	24.8600	67.0100	KFC	Fast Food Restaurant	24.849552	67.004348	(I. I. Chundrigar Rd,)
Karachi	24.8600	67.0100	OPTP	Fast Food Restaurant	24.813815	67.048572	(Khayaban-e-Shamsheer,)
Karachi	24.8600	67.0100	Subway	Fast Food Restaurant	24.883871	67.064696	(Shop #2, Al-Haram Tower 1, B.M.C.H.S., Bahadu...)
Karachi	24.8600	67.0100	McDonald's	Fast Food Restaurant	24.854860	67.026075	(Lakson Square,)
Karachi	24.8600	67.0100	McDonald's	Fast Food Restaurant	24.798660	67.034456	(Sea View Rd,)

Table 5: City Venue Dataset

Once I got these results, I plotted the number of locations returned by each city so I can choose the top 3 cities which had the maximum number of Fast-Food Restaurants.
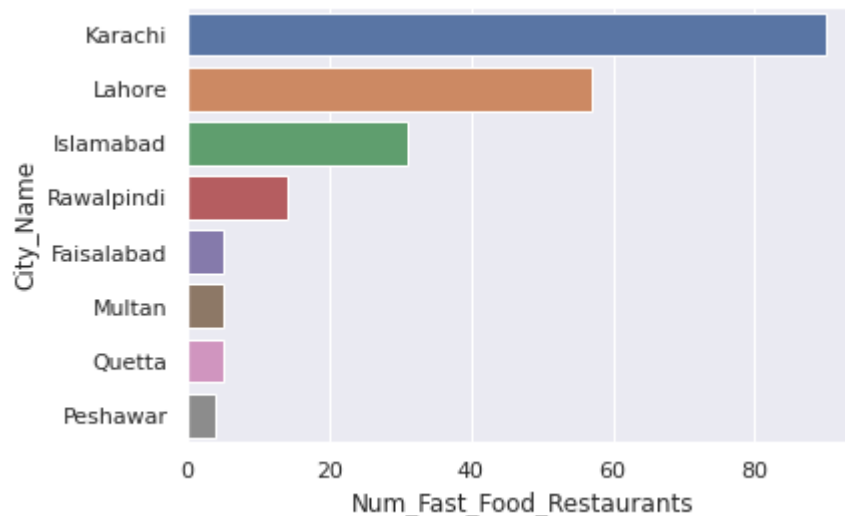

Figure 1: City Venue Amount

Results/Discussion

From the figure above, I found out that the top 3 locations which would serve as an ideal startup location are **Karachi, Lahore** and **Islamabad**. Now the only thing which was left was to plot these locations on the map so that the investor can easily see where similar restaurants are located. The results are shown in the images below.

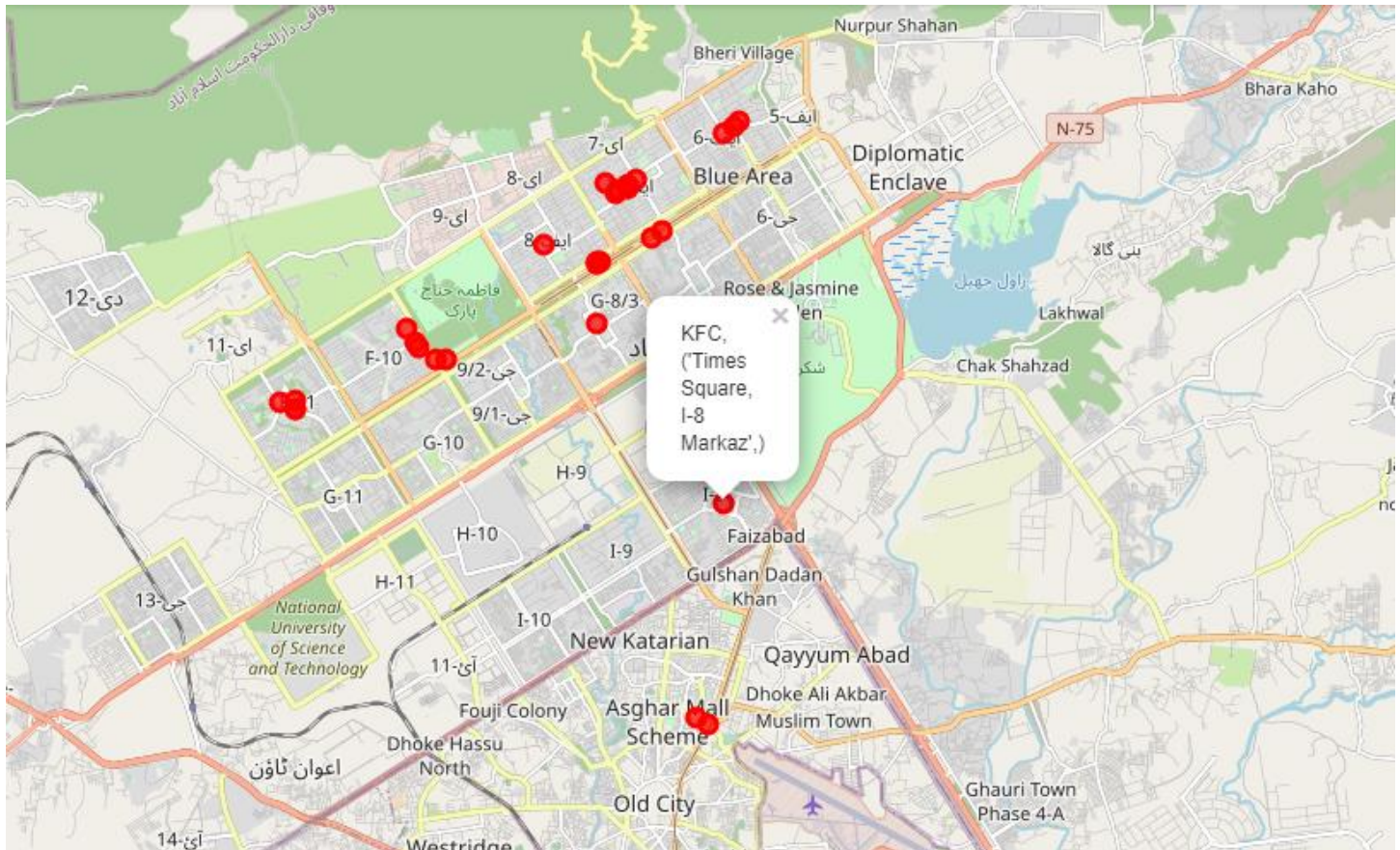The first figure shows the restaurants located in Islamabad highlighted with a **Red** marker.

Figure 2: Islamabad Map

Observations:

We can see that there are a few locations returned by the API, it shows that there are a bunch of Fast-Food restaurants within some sectors, some sectors don't have any Fast-Food spots, and some spots are located too far from all the other spots. A good idea would be to open a restaurant in a location where there is maximum distance from every other Fast-Food restaurant so that people visit the nearest restaurant which the investor is opening and they wouldn't have to travel far. This is also an idea location as there isn't enough competition and there is demand in less populated places in terms of Fast-Food venues.

The second figure shows the restaurants located in Lahore highlighted with a **Yellow/Gold** marker.
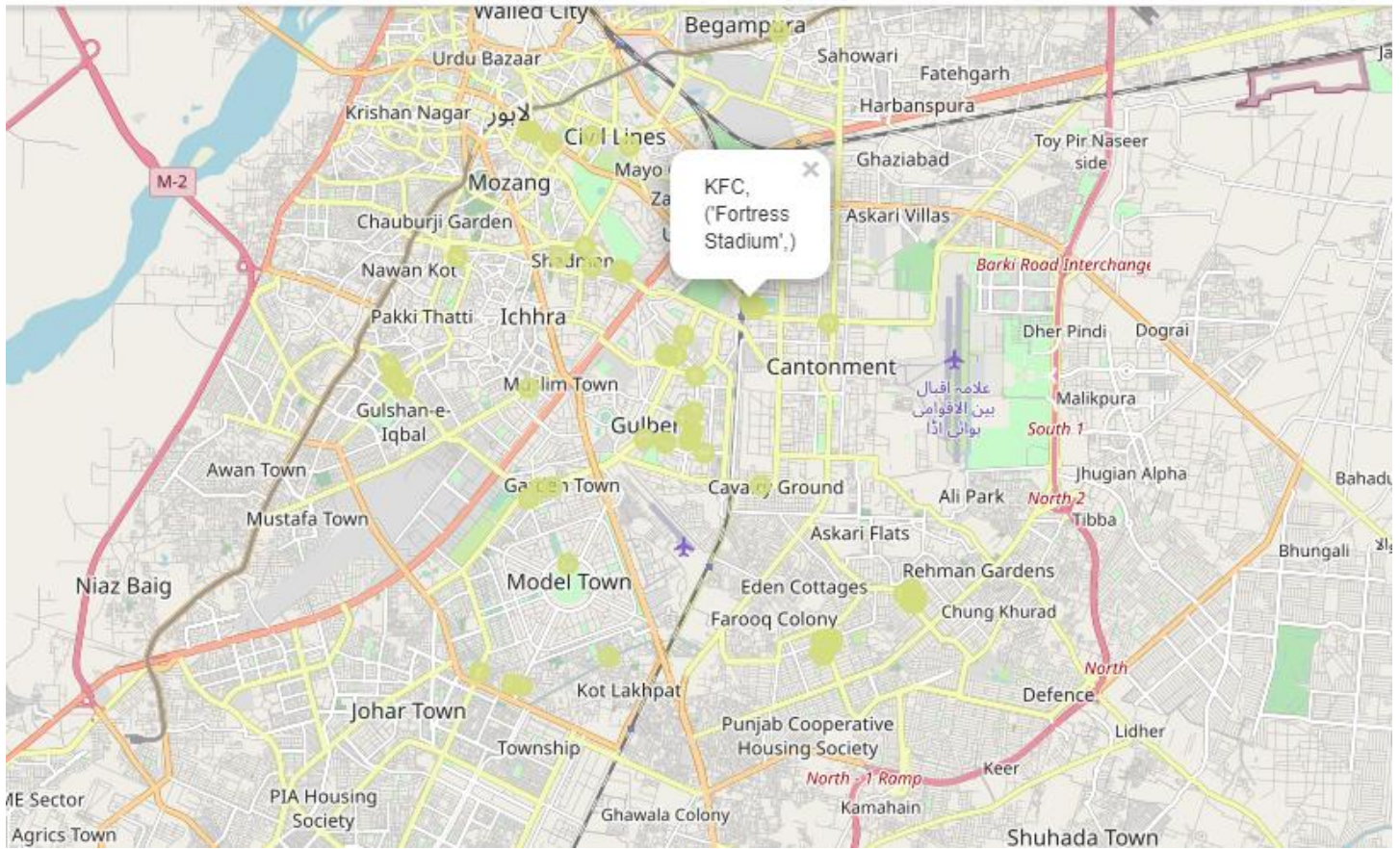

Figure 3: Lahore Map

Observations:

From the Lahore map we observe that there are a lot more venues as compared to Islamabad and they are more evenly spread out as well. As popular as this city is when it comes to Fast-Food spots, it will be difficult for a startup to be successful in its initial stages in such a high competition environment.

The last figure shows the restaurants located in Karachi highlighted with a **Purple** marker.

Figure 4: Karachi Map

Observations:

From this last map, we can see that this map has a similar spread to the Lahore map in terms of Fast-Food Locations, but its number of restaurants is significantly larger than that of Lahore which makes it extremely tough for a small startup to be noticed.

Conclusion

In this study, I analyzed all the major cities of Pakistan where Fast-Food locations are the most popular and upon mapping them on to a geographical map, saw that Karachi and Lahore were the most popular places in terms of Fast-Food locations per area and their spread was almost even. Ideally, for a startup, it is better to start in a location where there is high demand and low availability and, in our case, Islamabad was the best option out of them all as though it was not as popular in terms of Fast-Food locations as compared to Lahore and Karachi, it was still popular but the reason why it is the best option is because the venues aren't as evenly spread out as was the case with Karachi and Lahore. The spread was rather odd and there were a lot of places which barely had any Fast-Food venue in a large radius.

Therefore, it would be ideal for an investor to open a startup in Islamabad and once the demand increases, and the business is popular, their business could be expanded to the remaining big cities. Similarly this analysis will help out any other investor is finding out the ideal location for their first business startup.