

IBM
Coursera Data Science Specialization

Delhi Neighbourhood and Restaurants Analysis

Capstone Project

By

Simran Verma

May 2020

Introduction

Food is something that has brought communities, cultures, countries and people together for as long as we have existed as a human civilization. Now with the rise in globalization, restaurants from all over the world can be found in all major metropolitan cities of the world. From United States to Japan, Russia to Mexico, you can get all type of cuisines in the **capital of India, New Delhi**, which is the choice of city I have made. I have had the opportunity to spend multiple summers in Delhi with my grandparents throughout my childhood, and I speak from experience, it is one of the most wholesome and beautiful cities I have been to.

Coming back to food, Delivery, Dine-out, Pubs, Bars, Drinks, Buffet, Desserts you name it and Delhi will have a place. Delhi is a great place for foodies and also for businesses around dining and food. The number of restaurants is increasing day by day. Being one of the most densely populated cities in the world, there are customers for all kind of food joints, after all food is a basic necessity. The industry hasn't been saturated yet. Moreover, restaurants in Delhi are not just a sign for food businesses but also give the most prime locations for other businesses in Delhi too.

However, it has become difficult for them to compete with already established restaurants and hangout spots. The key issues that continue to pose a challenge to them include high real estate costs, rising food costs, shortage of quality manpower, fragmented supply chain and over-licensing.

With this project I aim at analysing demography, its impact on food culture and finding the most popular places in Delhi to open up a new consumer centric business around dining. Most importantly it will help new restaurants in deciding the theme, menus, cuisine, cost etc for a particular location, based on the observed trends.

It also aims at finding similarity between neighbourhoods of Delhi on the basis of food. With the analysis the project also will help people in choosing the restaurant based on several other factors. The project will mainly try to answer the question based on restaurants. And what factors should be kept in mind if someone wants to open new restaurant.

- Does the demography of an area matters?
- Does location of a particular type of restaurant also depends on the people living in that area?
- Does the theme of the restaurant matter?
- Are any neighbourhoods similar based on the type of food?

- Is a particular neighbourhood famous for certain kind of food?
- If two neighbourhoods are similar does that mean these are related or particular group of people live in the neighbourhood or these are the places to it?
- What kind of a food is more popular in a locality?

Data Description

The main data used for this project were from two sources:

- A Kaggle repository with neighbourhood and restaurants data of Delhi used for preliminary analysis

<https://www.kaggle.com/shaswatd673/delhi-neighborhood-data/version/1>

- Explore trending venues in a neighbourhood particularly restaurant (Foursquare API).

Other additional data:

- Coordinates (Geocoder Python)
- Population data for Delhi neighbourhoods to support analysis

<https://www.census2011.co.in/census/state/districtlist/delhi.html>

Data Collection Process

The Open dataset from Kaggle was downloaded as csv files and cleaned, and for reliability purposes verified against data from Wikipedia.

Foursquare API was used to obtain the nearby venues for each of the locations shortlisted by the preliminary analysis and plotting done on the Kaggle dataset. The location coordinates were obtained for those neighbourhoods using geopy module and then restaurants were filtered from the venues returned by the Foursquare API.

Please note that the preliminary analysis was not sufficient as the data obtained from that was limited and incomplete. However, it was essential to reduce the load on data obtained using Foursquare API and hence the data from the Preliminary stage, speeded up the entire process.

Data Cleaning

The most important thing while cleaning data was to make sure that none of the required values are null or nan. Longitudes and Latitudes specially. The restaurants data consisted of a little extra data which had to be filtered out. There were a handful of skew values in the longitudes on Delhi, (-90), which were removed as well.

Methodology

Let's go step by step through the notebook and try to understand each and every process. Firstly, by now, we know what are the libraries essential for performing data analysis involving geolocation and neighbourhoods. So, we go on to import all the necessary libraries. Now we will also be requiring the Foursquare API at various points throughout the project for detailed analysis, so, we initialize the credentials Client ID and Client Secret for it in the beginning.

Now let's get to the interesting parts, We have a file of Delhi neighbourhoods, in CSV format, containing Borough, Neighbourhood, latitude and longitude, we read this data into a pandas data frame to get a basic overlook of the city we want to explore.

```
In [11]: neighborhoods=pd.read_csv('delhi_dataSet.csv')
neighborhoods=neighborhoods[pd.notnull(neighborhoods['latitude'])]
neighborhoods=neighborhoods[pd.notnull(neighborhoods['longitude'])]
neighborhoods.head(10)
```

```
Out[11]:
```

	Borough	Neighborhood	latitude	longitude
0	North West Delhi	Adarsh Nagar	28.614192	77.071541
1	North West Delhi	Ashok Vihar	28.699453	77.184826
2	North West Delhi	Azadpur	28.707657	77.175547
3	North West Delhi	Bawana	28.799660	77.032885
5	North West Delhi	Dhaka	39.031714	-90.261223
6	North West Delhi	Jahangirpuri	28.725972	77.162658
7	North West Delhi	Karala	28.735140	77.032511
8	North West Delhi	Keshav Puram	28.688926	77.161683
9	North West Delhi	Kingsway Camp	28.614262	77.201555
10	North West Delhi	Kohat Enclave	28.698041	77.140539

```
In [12]: print('The dataframe has {} boroughs and {} neighborhoods.'.format(
          len(neighborhoods['Borough'].unique()),
          neighborhoods.shape[0]
        ))
```

The dataframe has 9 boroughs and 163 neighborhoods.

Here, we see that Delhi has 9 boroughs, also known as districts and 163 neighbourhoods. The districts in Delhi are very generic geographically, North Delhi, North West Delhi, South Delhi, South West Delhi and so on.

After this step we get the location of Delhi, to get a central geolocation which will help us orient our maps accordingly. We use nominatim from geocoder for this process, just a few simple lines of code and done.

```
In [43]: delhi_address = 'New Delhi, India'

geolocator = Nominatim(timeout=3, user_agent="foursquare_agent")
delhi_location = geolocator.geocode(address)
delhi_latitude = delhi_location.latitude
delhi_longitude = delhi_location.longitude
print('The georapical coordinate of New Delhi are {}, {}'.format(delhi_latitude, delhi_longitude))

The georapical coordinate of New Delhi are 28.6141793, 77.2022662.
```

Now the next file we are going to explore is a cornerstone to this analysis. The restaurants.csv file contains neighbourhood, neighbourhood latitude and longitude, venue, venue latitude and longitude and venue category. We have a record of 849 restaurants in 116 neighbourhoods.

Now we import available restaurant data as per locations in Delhi to get a rough idea as to which places have a higher density of food joints.

```
In [24]: restaurants=pd.read_csv('restaurant_dataSet.csv')
restaurants=restaurants[pd.notnull(restaurants['Venue Longitude'])]
restaurants=restaurants[pd.notnull(restaurants['Venue Latitude'])]
restaurants
```

Out[24]:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Adarsh Nagar	28.614192	77.071541	Eagle Boys Pizza	28.615595	77.070784	Pizza Place
1	Adarsh Nagar	28.614192	77.071541	Bikanerwala	28.613391	77.076084	Indian Restaurant
2	Adarsh Nagar	28.614192	77.071541	Bikano East Patel Nagar	28.616190	77.066978	Fast Food Restaurant
3	Adarsh Nagar	28.614192	77.071541	McDonald's	28.616330	77.067034	Fast Food Restaurant
4	Ashok Vihar	28.699453	77.184826	Nat Khat Caterers	28.699630	77.187832	Indian Restaurant
...
844	Tilak Nagar	28.639650	77.094039	CCD tilak nagar	28.636264	77.097048	Café
845	Vikaspuri	28.638419	77.070836	Domino's Pizza	28.638000	77.075000	Pizza Place
846	Vikaspuri	28.638419	77.070836	McDonald's	28.639752	77.075190	Fast Food Restaurant
847	Vikaspuri	28.638419	77.070836	Asian Garden	28.639661	77.074482	Indian Restaurant
848	Vikaspuri	28.638419	77.070836	Mela	28.635672	77.067237	Indian Restaurant

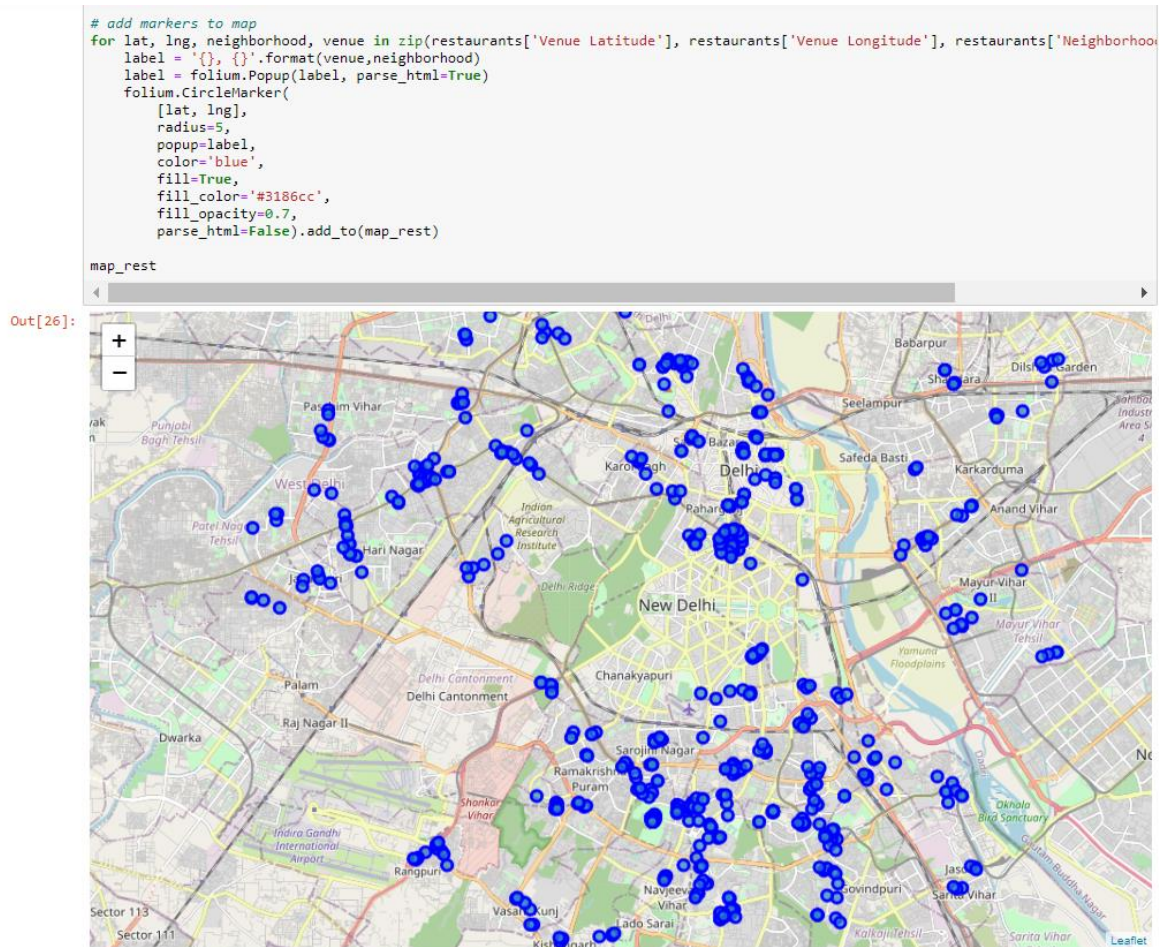
849 rows × 7 columns

```
In [25]: print('The dataframe has {} restaurants in {} neighbourhoods'.format(
    restaurants['Venue'].shape[0],
    len(restaurants['Neighborhood'].unique())
)
```

The dataframe has 849 restaurants in 116 neighbourhoods

Now in order to understand this data it is very essential for us to visualize and see for ourselves what is the density of the network of restaurants, is it according to the boroughs or are there any other central hubs, or are these restaurants evenly distributed throughout Delhi. (very unlikely, given Delhi is an old city that has grown

to accommodate all the exceeding requirements and expectations of people.) Let's take a look:



As we can see it is definitely not a uniform distribution, moreover some places seem to have a cluster of eating joints. Upon inspection of this very map, we found 5 neighbourhoods that had a huge number of restaurants. This was to make sure our analysis is productive and we don't end up exploring a sparse neighbourhood.

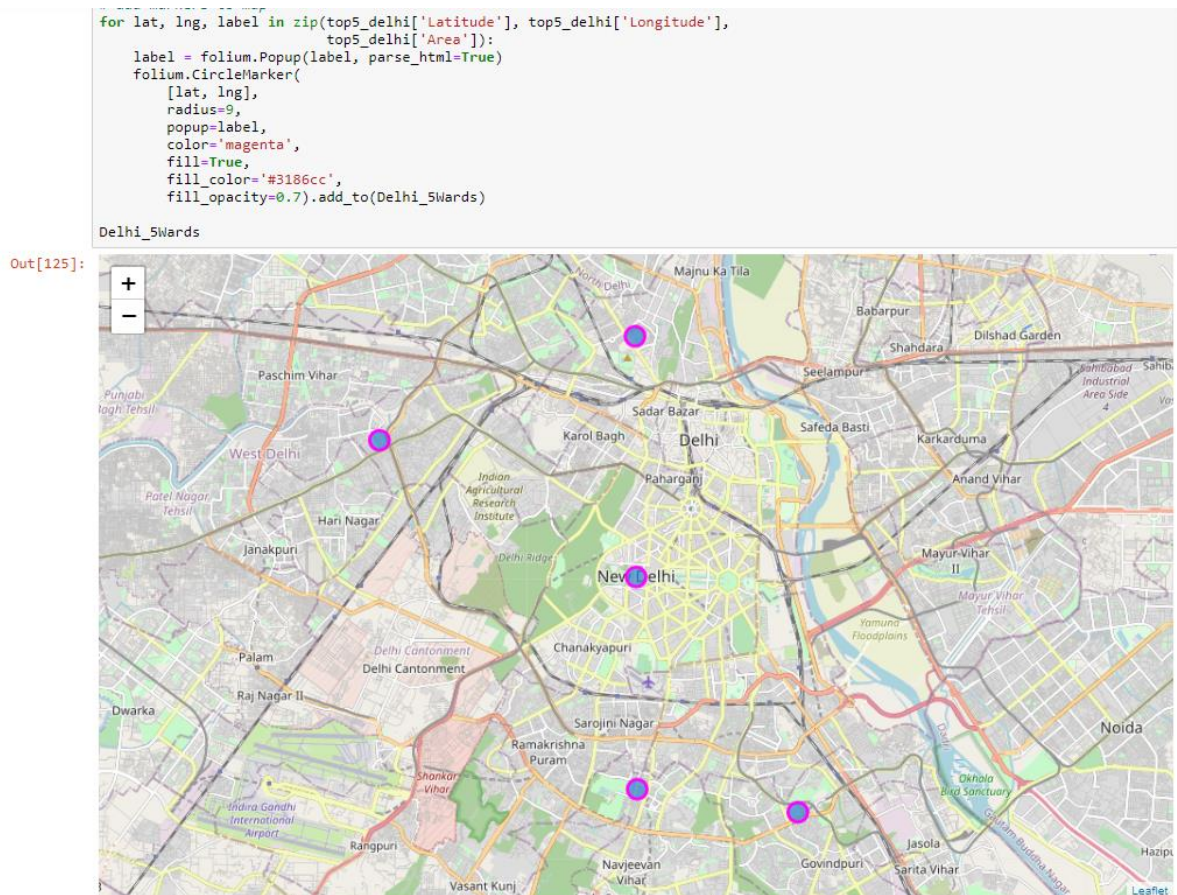
```
In [123]: lst=[address1,latitude1, longitude1],
            [address2,latitude2, longitude2],
            [address3,latitude3, longitude3],
            [address4,latitude4, longitude4],
            [address5,latitude5, longitude5]]
cols=cols + ['Area', 'Latitude', 'Longitude']
top5_delhi = pd.DataFrame(lst, columns=cols)
top5_delhi
```

Out[123]:

	Area	Latitude	Longitude
0	Connaught Place, Delhi, India	28.614179	77.202266
1	Green Park, Delhi, India	28.555537	77.202497
2	Nehru Place, Delhi, India	28.549257	77.252952
3	Shivaji Place, Delhi, India	28.651657	77.121703
4	Kamla Nagar, Delhi, India	28.680344	77.202129

For anyone who has been to Delhi even once, names like Connaught Place and Green Park would sound very familiar. A very important point to note here is that even though we are analysing restaurants and that led us to these 5 locations, these are

actually located in the most posh and populated boroughs of Delhi. These 5 neighbourhoods are prosperous in terms of food business, but are not restricted to just that, Connaught Place is also a shopping hub, Green Park has stadiums and sports parks likewise there are other important venues in these neighbourhoods as well.



Now that we have a more focussed approach towards Delhi neighbourhoods, let's move on to use the Foursquare API to explore these neighbourhoods, we obtain a list of all the venues within 5 km range of the locations we have using a loop to make calls to the API. We obtained 494 venues in these 5 neighbourhoods. Now, as a next step we filter out restaurants from the category column and form a data frame of restaurants alone. There are 179 restaurants out of the 494 total venues.

```
In [127]: Delhi_5_Dist_Venues = getNearbyVenues(names=top5_delhi['Area'],
        latitudes=top5_delhi['Latitude'],
        longitudes=top5_delhi['Longitude']
        )
```

Connaught Place, Delhi, India
 Green Park, Delhi, India
 Nehru Place, Delhi, India
 Shivaji Place, Delhi, India
 Kamla Nagar, Delhi, India

We now fetch the venues within 5km range of these centres and subsequently filter the restaurants from the results returned by foursquare API

```
In [128]: print ("Shape of the Venues Dataframe: ", Delhi_5_Dist_Venues.shape)
Delhi_5_Dist_Venues.tail(3)
```

Shape of the Venues Dataframe: (494, 7)

Out[128]:

	District	Dist_Latitude	Dist_Longitude	Venue	Venue_Lat	Venue_Long	Venue_Category
491	Kamla Nagar, Delhi, India	28.680344	77.202129	Glenze Cafe & Bakers	28.647162	77.188480	Café
492	Kamla Nagar, Delhi, India	28.680344	77.202129	Alfa Spice	28.644484	77.178748	Multicuisine Indian Restaurant
493	Kamla Nagar, Delhi, India	28.680344	77.202129	Dukes Pastry Shop	28.641820	77.186111	Bakery

```
In [129]: # Create a Data-Frame out of it to Concentrate Only on Restaurants
```

```
Delhi_5_Dist_Venues_only_restaurant = Delhi_5_Dist_Venues[Delhi_5_Dist_Venues['Venue_Category']\
        .str.contains('Restaurant')].reset_index(drop=True)
Delhi_5_Dist_Venues_only_restaurant.index = np.arange(1, len(Delhi_5_Dist_Venues_only_restaurant)+1)
print ("Shape of the Data-Frame with Venue Category only Restaurant: ", Delhi_5_Dist_Venues_only_restaurant.shape)
Delhi_5_Dist_Venues_only_restaurant.head(3)
```

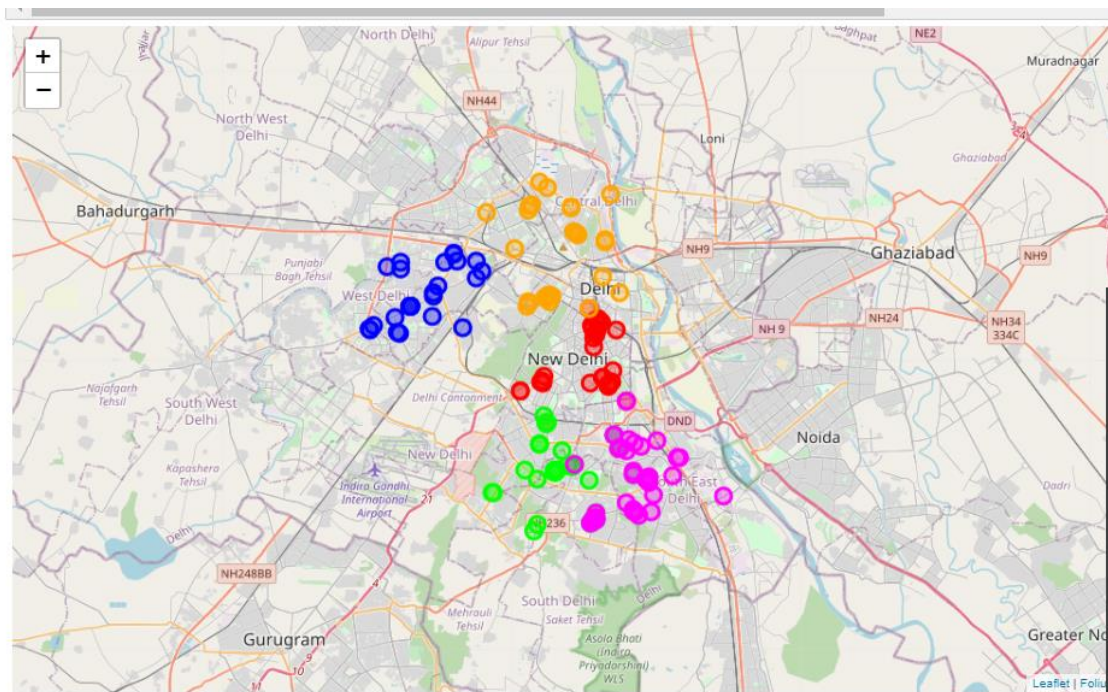
Shape of the Data-Frame with Venue Category only Restaurant: (179, 7)

Out[129]:

	District	Dist_Latitude	Dist_Longitude	Venue	Venue_Lat	Venue_Long	Venue_Category
1	Connaught Place, Delhi, India	28.614179	77.202266	Tamra	28.620543	77.218174	Restaurant
2	Connaught Place, Delhi, India	28.614179	77.202266	Varq वरूँ	28.604547	77.223781	Indian Restaurant
3	Connaught Place, Delhi, India	28.614179	77.202266	Sanadige	28.601969	77.187020	Karnataka Restaurant

Now we plot all the restaurants as clusters according to the 5 neighbourhoods.

Out[131]:



The next step gets very interesting, each of these restaurants has a category, It could be a bakery, a café, an Italian restaurant, a Pizza place or what not. Let's have a look at the diversity of cuisine you can get in Delhi.

We also see the distinct categories of restaurants as it will be beneficial in recommending which type of restaurants would be popular in which area

```
In [132]: ## Number of Unique Categories in the Dataframe
print('There are {} unique categories.'.format(len(Delhi_5_Dist_Venues['Venue_Category'].unique())))
## We can check some of the categories randomly
print (Delhi_5_Dist_Venues[['Venue_Category']][70:75])
```

```
There are 102 unique categories.
Venue_Category
70    Chinese Restaurant
71    Mediterranean Restaurant
72         Hotel
73         Restaurant
74    Indian Restaurant
```

There are 102 unique categories of restaurants in these 5 neighbourhoods!

We can also see which category of restaurants are most frequent, in all of these 5 neighbourhoods.

```
In [134]: # create a dataframe of top 10 categories
Delhi_5_Dist_Venues_Top10 = Delhi_5_Dist_Venues['Venue_Category'].value_counts()[0:10].to_frame(name='frequency')
Delhi_5_Dist_Venues_Top10=Delhi_5_Dist_Venues_Top10.reset_index()
#Tokyo_5_Dist_Venues_Top10

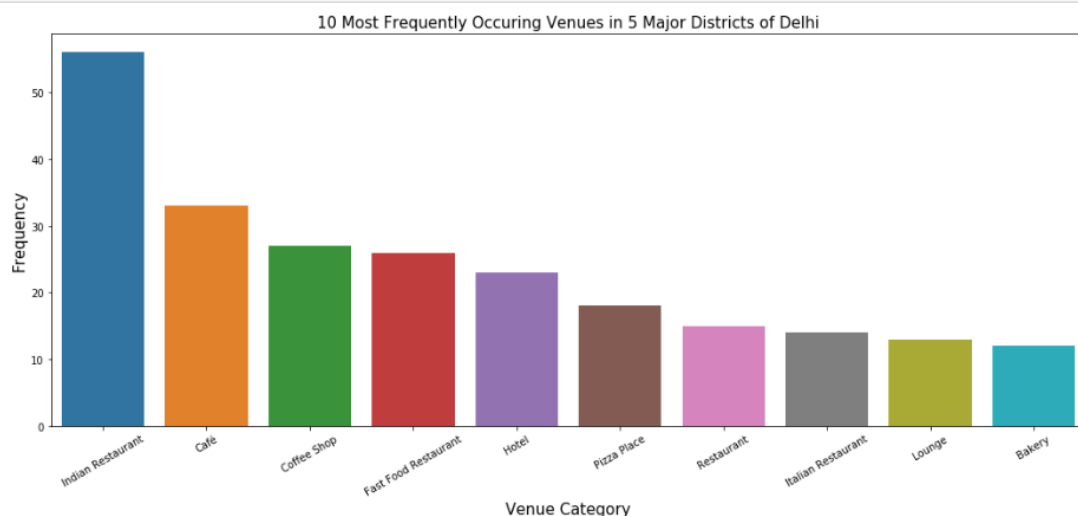
Delhi_5_Dist_Venues_Top10.rename(index=str, columns={"index": "Venue_Category", "frequency": "Frequency"}, inplace=True)
Delhi_5_Dist_Venues_Top10
```

```
Out[134]:
```

	Venue_Category	Frequency
0	Indian Restaurant	56
1	Café	33
2	Coffee Shop	27
3	Fast Food Restaurant	26
4	Hotel	23
5	Pizza Place	18
6	Restaurant	15
7	Italian Restaurant	14
8	Lounge	13
9	Bakery	12

This can be visualized as a bar chart, and Indians tend to like their cuisine the best, in the heart of Delhi at least.

```
In [135]: import seaborn as sns
fig = plt.figure(figsize=(18,7))
s=sns.barplot(x="Venue_Category", y="Frequency", data=Delhi_5_Dist_Venues_Top10)
s.set_xticklabels(s.get_xticklabels(), rotation=30)
plt.title('10 Most Frequently Occuring Venues in 5 Major Districts of Delhi', fontsize=15)
plt.xlabel("Venue Category", fontsize=15)
plt.ylabel("Frequency", fontsize=15)
plt.savefig("Most_Freq_Venues.png", dpi=300)
plt.show()
```



```
In [136]: Delhi_5_Dist_Venues_onehot = pd.get_dummies(Delhi_5_Dist_Venues[['Venue_Category']], prefix="", prefix_sep="")

### add district column back to dataframe
Delhi_5_Dist_Venues_onehot['District'] = Delhi_5_Dist_Venues['District']
### move district column to the first column
fixed_columns = [Delhi_5_Dist_Venues_onehot.columns[-1]] + list(Delhi_5_Dist_Venues_onehot.columns[:-1])
Delhi_5_Dist_Venues_onehot = Delhi_5_Dist_Venues_onehot[fixed_columns]
Delhi_5_Dist_Venues_onehot.head(3)
```

Out[136]:

	District	American Restaurant	Arcade	Art Gallery	Art Museum	Asian Restaurant	Athletics & Sports	BBQ Joint	Bakery	Bar	...	Sports Bar	Stadium	Tea Room
0	Connaught Place, Delhi, India	0	0	0	0	0	0	0	0	0	...	0	0	0
1	Connaught Place, Delhi, India	0	0	0	0	0	0	0	0	0	...	0	0	0
2	Connaught Place, Delhi, India	0	0	0	0	0	0	0	0	0	...	0	0	0

3 rows × 103 columns

```
In [137]: Delhi_5_Dist_Venues_Grouped = Delhi_5_Dist_Venues_onehot.groupby('District').mean().reset_index()
Delhi_5_Dist_Venues_Grouped.index = np.arange(1, len(Delhi_5_Dist_Venues_Grouped)+1)
Delhi_5_Dist_Venues_Grouped
```

Out[137]:

	District	American Restaurant	Arcade	Art Gallery	Art Museum	Asian Restaurant	Athletics & Sports	BBQ Joint	Bakery	Bar	...	Sports Bar	Stadium	Tea Room
1	Connaught Place, Delhi, India	0.000000	0.010000	0.02	0.01	0.02	0.00	0.020000	0.010000	0.030000	...	0.000000	0.00	0.00
2	Green Park	0.000000	0.000000	0.00	0.00	0.00	0.00	0.000000	0.000000	0.000000	...	0.000000	0.00	0.00

Result

Now to proceed towards the results of our analysis, we can observe what kind of restaurants are located in which neighbourhood.

The following result is the most important takeaway from our data analysis project:

We see here that out of the 5 neighbourhoods we choose, Connaught place has a greater frequency of restaurants in general and even for certain specific kinds of restaurants, cafes and bars.

Let us note our observations:

1. Bar and Lounge only come up in the list for Connaught Place
2. Chinese restaurants are available in Kamla Nagar, Connaught Place and Nehru Place.
3. Cafes are pretty popular in all of these 5 neighbourhoods (Everyone likes coffee) and occupy median position in the frequencies table as well.
4. Indian Restaurants seem to flourish in all the neighbourhoods.

%%%%%%%%%% Connaught Place, Delhi, India %%%%%%%%%%		
	Venue	Freq
0	Indian Restaurant	0.14
1	Café	0.11
2	Hotel	0.09
3	Lounge	0.04
4	Chinese Restaurant	0.03
5	Bar	0.03
6	South Indian Restaurant	0.02
7	Restaurant	0.02
8	Deli / Bodega	0.02
9	Mediterranean Restaurant	0.02

%%%%%%%%%% Green Park, Delhi, India %%%%%%%%%%		
	Venue	Freq
0	Indian Restaurant	0.10
1	Italian Restaurant	0.06
2	Lounge	0.05
3	Coffee Shop	0.05
4	Asian Restaurant	0.04
5	Bakery	0.04
6	Café	0.04
7	Shopping Mall	0.03
8	Market	0.03
9	Dessert Shop	0.03

%%%%%%%%%% Kamla Nagar, Delhi, India %%%%%%%%%%		
	Venue	Freq
0	Fast Food Restaurant	0.12
1	Pizza Place	0.10
2	Indian Restaurant	0.10
3	Coffee Shop	0.09
4	Snack Place	0.06
5	Café	0.06
6	Hotel	0.05
7	Bakery	0.04
8	Sandwich Place	0.04
9	Chinese Restaurant	0.04

%%%%%%%%%% Nehru Place, Delhi, India %%%%%%%%%%		
	Venue	Freq
0	Indian Restaurant	0.13
1	Restaurant	0.07
2	Market	0.06
3	Café	0.05
4	Italian Restaurant	0.05
5	Dessert Shop	0.04
6	Lounge	0.04
7	Coffee Shop	0.04
8	Hotel	0.04
9	Chinese Restaurant	0.03

%%%%%%%%%% Shivaji Place, Delhi, India %%%%%%%%%%		
	Venue	Freq
0	Fast Food Restaurant	0.11
1	Indian Restaurant	0.10
2	Pizza Place	0.09
3	Café	0.07
4	Coffee Shop	0.07
5	Donut Shop	0.05
6	Sandwich Place	0.04
7	Shopping Mall	0.03
8	BBQ Joint	0.03
9	Restaurant	0.03

Discussion

An important limitation of this project is that we haven't taken into account the revenues of these restaurants into account, that could have given us an insight into the monetary aspect of them as businesses. However, if we take the liberty of making the assumption, that a significant % of the restaurants are doing well, and given the growing population of the city of Delhi, the demand is not flattening. How would we go about recommending restaurant owners which location to pick?

I'd say the safest bet would be to open a franchise in Connaught Place, even though there will be a lot of competition, there is clearly a good-sized audience. A look at the categories gives us Deli, Lounges, Bar, Mediterranean restaurants reflects the cosmopolitan culture of CP.

If say someone wants to open a Chinese restaurant, which we can see are quite popular, there can be two possible recommendations in terms of location,

1. To go for a place that already has consumers, Kamla Nagar, Connaught Place or Nehru Place.
2. To experiment at a new location, like Shivaji Place, this will require establishing a presence and creating an audience for the restaurant.

Conclusion

After all the analysis we did some of the point which were worth noting that there are lesser foreign food chains than Indian restaurants in Delhi. Cafes have outlets in most of the neighbourhoods of Delhi.

The Indian food service market has come a long way from the early Nineties when it was dominated by unorganised players and few brands. The revolution began in 1996 with McDonalds, Pizza Hut, Dominos Pizza, Subway and Yo! China, among others, setting up shop in the country.

Since then, the food services market has been continuously growing. Quick service restaurants are a mainstay of the Indian food service market, and are growing fast. Fine dining is gaining prominence too. Both multi-cuisine and single-cuisine establishments have shown tremendous growth. Delhi is an inspiring mish-mash of old and new, and has plenty of options on offer, regardless of whether you are visiting only, or setting up shop in the restaurant business.