```
In [1]: import numpy as np
        import pandas as pd
        pd.set_option('display.max_columns', None)
        pd.set_option('display.max_rows', None)

        #!conda install -c conda-forge geopy --yes
        from geopy.geocoders import Nominatim
        import urllib.request
        import json
        from bs4 import BeautifulSoup
        from urllib.request import urlopen
        import requests
        from pandas.io.json import json_normalize

        import matplotlib.cm as cm
        import matplotlib.colors as colors
        # Matplotlib and associated plotting modules
        import matplotlib.cm as cm
        import matplotlib.pyplot as plt
        import matplotlib.colors as colors
        %matplotlib inline
        from sklearn.cluster import KMeans

        #!conda install -c conda-forge folium=0.5.0 --yes
        import folium

        print('Libraries imported.')
```

Libraries imported.

```
In [2]: # libraries for displaying images
        from IPython.display import Image
        from IPython.core.display import HTML

        # tranforming json file into a pandas dataframe library
        from pandas.io.json import json_normalize
```
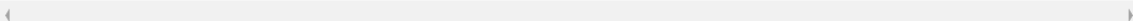
```
In [3]: address = 'New Delhi, India'

        geolocator = Nominatim(user_agent="delhi_exp")
        location = geolocator.geocode(address)
        latitude = location.latitude
        longitude = location.longitude
        print('The geograpical coordinates of Delhi, India are {}, {}.'.format(latitude, longitude))
```

The geograpical coordinates of Delhi, India are 28.6141793, 77.2022662.

```
In [4]: map_delhi = folium.Map(location=[latitude, longitude], zoom_start=10)
        map_delhi
```

Out[4]:

Defining Foursquare credentials and version

In [5]:
```python
CLIENT_ID = 'PBFP5HJT00LXB0QRPXGDTZQI0PO2MFMG4IFQ3K1OKH0WPKPX' # your Foursquare ID
CLIENT_SECRET = '5VVO420NBKH1FDD2AWTYUUQIWUF5IM3LQBHKNRRDHQFRCNY4' # your Foursquare Secret
VERSION = '20180605' # Foursquare API version

print('Your credentails:')
print('CLIENT_ID: ' + CLIENT_ID)
print('CLIENT_SECRET:' + CLIENT_SECRET)
```

```
Your credentails:
CLIENT_ID: PBFP5HJT00LXB0QRPXGDTZQI0PO2MFMG4IFQ3K1OKH0WPKPX
CLIENT_SECRET:5VVO420NBKH1FDD2AWTYUUQIWUF5IM3LQBHKNRRDHQFRCNY4
```

Now, let's get the top 100 venues that are within a radius of 50 kilometers.

In [6]:
```python
# function that extracts the category of the venue
def get_category_type(row):
    try:
        categories_list = row['categories']
    except:
        categories_list = row['venue.categories']

    if len(categories_list) == 0:
        return None
    else:
        return categories_list[0]['name']
```

In [7]:
```python
radius = 50000
LIMIT = 100




offset = 0
total_venues = 0
foursquare_venues = pd.DataFrame(columns = ['name', 'categories', 'lat', 'lng','city'])

while (True):


    url = "https://api.foursquare.com/v2/venues/explore?client_id={}&client_secret={}&v={}&ll={},{}&radius={}&limit={}&offset={}".format
(CLIENT_ID,CLIENT_SECRET,VERSION,

latitude,longitude,radius,LIMIT,offset)

    result = requests.get(url).json()
    venues_fetched = len(result['response']['groups'][0]['items'])
    total_venues = total_venues + venues_fetched
    print("Total {} venues fetched within a total radius of {} Km".format(venues_fetched, radius/1000))

    venues = result['response']['groups'][0]['items']
    venues = json_normalize(venues)

    # Filter the columns
    filtered_columns = ['venue.name', 'venue.categories', 'venue.location.lat', 'venue.location.lng','venue.location.city']
    venues = venues.loc[:, filtered_columns]
```

```
        # Filter the category for each row
        venues['venue.categories'] = venues.apply(get_category_type, axis = 1)

        # Clean all column names
        venues.columns = [col.split(".")[-1] for col in venues.columns]
        foursquare_venues = pd.concat([foursquare_venues, venues], axis = 0, sort = False)

        if (venues_fetched < 50):
            break
        else:
            offset = offset + 20


foursquare_venues = foursquare_venues.reset_index(drop = True)
print("\nTotal {} venues fetched".format(total_venues))
```

Total 100 venues fetched within a total radius of 50.0 Km

/home/jupyterlab/conda/envs/python/lib/python3.6/site-packages/ipykernel_launcher.py:23: FutureWarning: pandas.io.json.json_normalize is deprecated, use pandas.json_normalize instead

Total 100 venues fetched within a total radius of 50.0 Km
Total 100 venues fetched within a total radius of 50.0 Km
Total 100 venues fetched within a total radius of 50.0 Km
Total 98 venues fetched within a total radius of 50.0 Km
Total 78 venues fetched within a total radius of 50.0 Km
Total 58 venues fetched within a total radius of 50.0 Km
Total 38 venues fetched within a total radius of 50.0 Km

Total 672 venues fetched

In [8]:
```
#venues = result['response']['groups'][0]['items']
#venues = json_normalize(venues)
#venues
```

In [9]:
```
foursquare_venues.shape
```

Out[9]: (672, 5)

In [10]:
```
four_df=foursquare_venues
#foursquare_venues=four_df
```

Checking for the duplicate values in data as the fetching limit was only 100 and I tried to fetch it multiple times by changing offset

In [11]:
```
foursquare_venues.drop_duplicates(subset=None, keep='first', inplace=True)
foursquare_venues.shape
```

Out[11]: (178, 5)

In [12]:
```
foursquare_venues.head(20)
```

Out[12]:

	name	categories	lat	lng	city
0	The Imperial	Hotel	28.625548	77.218664	New Delhi
1	Tamra	Restaurant	28.620543	77.218174	New Delhi
2	Pandey Paan	Smoke Shop	28.622249	77.201075	New Delhi
3	The Big Chill Cafe	Italian Restaurant	28.600686	77.227636	New Delhi
4	Amour Bistro	Café	28.601569	77.185923	NaN
5	The Leela Palace	Hotel	28.579923	77.189291	New Delhi
6	Varq	वर्क	Indian Restaurant	28.604547	77.223781	New Delhi
7	Naturals Ice Cream	Ice Cream Shop	28.634455	77.222139	New Delhi
8	The Lodhi Hotel	Hotel	28.591669	77.238131	New Delhi
9	Khan Market	खान मार्केट (Khan Market)	Market	28.600342	77.226923	New Delhi
10	Lodhi Gardens (लोधी बाग़) (Lodhi Gardens)	Park	28.591424	77.220899	New Delhi
11	Jamavar Kashmiri Restaurant @ The Leela	Indian Restaurant	28.580143	77.189267	Delhi
12	Lantern's	Bar	28.643183	77.177746	New Delhi
13	Indian Accent	Indian Restaurant	28.592026	77.238006	New Delhi
14	Bukhara	North Indian Restaurant	28.596914	77.173358	New Delhi
15	National Crafts Museum	राष्ट्रीय शिल्प संग्र...	Art Museum	28.613598	77.242039	New Delhi
16	Humayun's Tomb	हुमायूँ का मकबरा (Humayun's T...	Monument / Landmark	28.593197	77.246047	New Delhi
17	SODABOTTLEOPENERWALA	Irani Cafe	28.600141	77.226273	New Delhi
18	Bengali Market	बंगाली मार्केट	বাংলা বাজার	Indian Restaurant	28.629498	77.232020	New Delhi
19	Delhi Golf Club	Golf Course	28.599968	77.233288	New Delhi

In [13]:
```python
foursquare_venues.dropna(subset=["city"],axis=0,inplace=True)
foursquare_venues.head()
```

Out[13]:

	name	categories	lat	lng	city
0	The Imperial	Hotel	28.625548	77.218664	New Delhi
1	Tamra	Restaurant	28.620543	77.218174	New Delhi
2	Pandey Paan	Smoke Shop	28.622249	77.201075	New Delhi
3	The Big Chill Cafe	Italian Restaurant	28.600686	77.227636	New Delhi
5	The Leela Palace	Hotel	28.579923	77.189291	New Delhi

In [14]:
```python
# Map of Delhi already createdusing latitude and longitude values


# add markers to map
for lat, lng, name, categories in zip(foursquare_venues['lat'],foursquare_venues['lng'],foursquare_venues['name'],foursquare_venues['categories']):
    label = '{}, {}'.format(name, categories)
    label = folium.Popup(label, parse_html=True)
    folium.CircleMarker([lat, lng],radius=2,popup=label,color='blue',fill=True,fill_color='#3186cc',fill_opacity=0.7).add_to(map_delhi)

map_delhi
```

Out[14]:

In [15]:
```python
df_one=foursquare_venues
#foursquare_venues=df_one
```

In [16]:
```python
foursquare_venues
```

	name	categories	lat	lng	city
0	The Imperial	Hotel	28.625548	77.218664	New Delhi
1	Tamra	Restaurant	28.620543	77.218174	New Delhi
2	Pandey Paan	Smoke Shop	28.622249	77.201075	New Delhi
3	The Big Chill Cafe	Italian Restaurant	28.600686	77.227636	New Delhi
5	The Leela Palace	Hotel	28.579923	77.189291	New Delhi
6	Varq	वर्क	Indian Restaurant	28.604547	77.223781	New Delhi
7	Naturals Ice Cream	Ice Cream Shop	28.634455	77.222139	New Delhi
8	The Lodhi Hotel	Hotel	28.591669	77.238131	New Delhi
9	Khan Market	खान मार्केट (Khan Market)	Market	28.600342	77.226923	New Delhi
10	Lodhi Gardens (लोधी बाग़) (Lodhi Gardens)	Park	28.591424	77.220899	New Delhi
11	Jamavar Kashmiri Restaurant @ The Leela	Indian Restaurant	28.580143	77.189267	Delhi
12	Lantern's	Bar	28.643183	77.177746	New Delhi
13	Indian Accent	Indian Restaurant	28.592026	77.238006	New Delhi
14	Bukhara	North Indian Restaurant	28.596914	77.173358	New Delhi
15	National Crafts Museum	राष्ट्रीय शिल्प संग्र...	Art Museum	28.613598	77.242039	New Delhi
16	Humayun's Tomb	हुमायूँ का मकबरा (Humayun's T...	Monument / Landmark	28.593197	77.246047	New Delhi
17	SODABOTTLEOPENERWALA	Irani Cafe	28.600141	77.226273	New Delhi
18	Bengali Market	बंगाली मार्केट	বাংলা বাজার	Indian Restaurant	28.629498	77.232020	New Delhi
19	Delhi Golf Club	Golf Course	28.599968	77.233288	New Delhi
20	The Oberoi	Hotel	28.596564	77.239634	New Delhi

22	Hauz Khas Village	Market	28.554311	77.195140	New Delhi
23	The Claridges	Hotel	28.600705	77.216601	Delhi
24	ITC Maurya	Hotel	28.597130	77.173643	New Delhi
25	Gulati Restaurant	Indian Restaurant	28.608010	77.229989	New Delhi
26	PCO	Speakeasy	28.561635	77.155590	Delhi
27	Blue Bar	Bar	28.595403	77.170529	New Delhi
28	Connaught Place	कनॉट प्लेस (Connaught Place)	Plaza	28.632731	77.220018	New Delhi
29	L'Opera	Café	28.599784	77.226144	New Delhi
30	Taj Palace Hotel	Hotel	28.595098	77.170913	New Delhi
31	PVR Director's Cut	Movie Theater	28.540773	77.154499	New Delhi
32	Fitness First Platinum, Select City Walk	Gym / Fitness Center	28.528615	77.218361	New Delhi
33	JW Marriott Hotel New Delhi Aerocity	Hotel	28.552782	77.121491	New Delhi
34	Big Yellow Door	Café	28.693245	77.204948	Delhi
35	The Big Chill Cafe	Italian Restaurant	28.552728	77.241923	New Delhi
36	Olive Bar & Kitchen	Italian Restaurant	28.526103	77.183976	New Delhi
37	The Big Chill Cafe	Restaurant	28.528201	77.217748	New Delhi
38	Maxims	Bakery	28.552849	77.242065	New Delhi
39	Khan Chacha	Indian Restaurant	28.600618	77.227237	New Delhi
40	Select Citywalk	Shopping Mall	28.528678	77.219136	New Delhi
41	Kunzum Travel Cafe	Art Gallery	28.553684	77.194368	New Delhi
42	Indian Accent	Indian Restaurant	28.570423	77.256753	Delhi

44	Sagar South Indian Restaurant	South Indian Restaurant	28.573393	77.230293	New Delhi
45	Kainoosh	Asian Restaurant	28.542631	77.156886	New Delhi
46	Thai High	Thai Restaurant	28.522127	77.181831	Mahrauli
47	Jolly Creations Designer Boutique	Boutique	28.662689	77.226300	New Delhi
48	Qutub Minar	क़ुतुब मीनार (Qutub Minar)	Monument / Landmark	28.525265	77.186593	New Delhi
49	Imperfecto	Mediterranean Restaurant	28.554657	77.195092	New Delhi
50	ISA KHAN'S GARDEN TOMB	Historic Site	28.592797	77.246147	New Delhi
51	Dum Pukht	Indian Restaurant	28.597194	77.173288	New Delhi
52	Big Chill	Italian Restaurant	28.542758	77.156446	Delhi
55	b-bar	Lounge	28.528593	77.218341	Delhi
56	Aloft New Delhi Aerocity	Hotel	28.552446	77.123437	New Delhi
57	Khan Chacha	Indian Restaurant	28.528170	77.219207	New Delhi
58	DLF Emporio	Shopping Mall	28.543215	77.156855	New Delhi
59	Ambience Mall	Shopping Mall	28.541012	77.155128	New Delhi
60	Mahabelly	Restaurant	28.527303	77.216800	Delhi
61	Trident	Hotel	28.501032	77.089236	Gurgaon
62	Blue Tokai Coffee Roasters	Coffee Shop	28.517214	77.200021	New Delhi
63	Woodbox Cafe	Café	28.694334	77.204811	New Delhi
65	Cafe Wink	Italian Restaurant	28.657311	77.317098	New Delhi
66	Monkey Bar (Monkey Bar New Delhi)	Gastropub	28.536578	77.147669	New Delhi
68	DLF CyberHub	Plaza	28.495342	77.088741	Gurgaon

69	Q'La	Bar	28.527281	77.185795	New Delhi
70	Lighthouse 13	Lounge	28.529997	77.219804	New Delhi
71	Defence Colony Market	Market	28.573653	77.230134	New Delhi
73	The Oberoi	Hotel	28.501941	77.088162	Gurgaon
74	Cafe Dori @ Nappa Dori Warehouse	Café	28.503955	77.184991	New Delhi
76	au bon pain	Café	28.494842	77.088421	Gurgaon
77	PVR bluO	Bowling Alley	28.502957	77.097185	Gurgaon
78	Pita Pit	Falafel Restaurant	28.495092	77.088534	Gurgaon
79	Starbucks	Coffee Shop	28.495766	77.088946	Gurgaon
80	Saffron	Indian Restaurant	28.500975	77.089319	Gurgaon
81	The Leela Ambience	Hotel	28.505570	77.096530	Gurgaon
83	Jawaharlal Nehru Stadium	Stadium	28.583104	77.232587	New Delhi
84	Galleria Market	Market	28.467333	77.081923	Gurgaon
85	Vivanta by Taj	Hotel	28.455952	77.070524	NCR
86	Pvr Ecx Chanakya	Multiplex	28.584352	77.190767	New Delhi
87	Cafe Di Ghent	Café	28.468248	77.082918	Gurgaon
88	Leisure Valley	Park	28.469142	77.064690	Gurgaon
89	Cafe Delhi Heights	Café	28.468201	77.083786	Gurgaon
90	Di Ghent Boulangerie	Bagel Shop	28.468401	77.083420	Gurgaon
92	Sakley's Mountain Cafe	Café	28.467584	77.081695	Gurgaon
93	Book India Holiday	Pub	28.460506	77.095925	Noida

94	Big Bazaar, Ambience Mall	Department Store	28.540983	77.155005	New Delhi
96	Biryani Blues	Indian Restaurant	28.462280	77.087233	Gurgaon
98	Kingdom Of Dreams	General Entertainment	28.468011	77.068125	Gurgaon
99	Cilantro	Restaurant	28.500921	77.089363	Gurgaon
180	Café Parmesan	Café	28.395362	77.323522	Farīdābād
182	Windsor	Food Court	28.638558	77.365699	Ghāziābād
183	Open Tap	Brewery	28.405130	77.044332	Gurgaon
185	Starbucks	Coffee Shop	28.406085	77.044741	Gurgaon
186	DoubleTree by Hilton	Hotel	28.422142	77.104506	Gurgaon
187	Hyatt Regency Gurgaon	Hotel	28.392552	76.976269	Gurgaon
188	Madison & Pike	Café	28.414426	77.065792	Gurgaon
189	JAYPEE GREENS GOLF & SPA RESORT	Hotel	28.467824	77.510819	Greater Noida
190	Chickenette	Fried Chicken Joint	28.395114	77.324468	Farīdābād
191	Sagar Ratna	Indian Restaurant	28.421042	77.038509	Gurgaon
193	Buddh International Circuit	Racetrack	28.350988	77.535791	Greater Noida
194	Park Plaza	Hotel	28.429159	77.296713	Farīdābād
195	Vapour Bar Exchange	Brewery	28.406734	77.041820	Gurgaon
196	Shopper's Stop	Clothing Store	28.413425	77.042197	Gurgaon
197	Domino's Pizza	Pizza Place	28.384309	77.324458	Farīdābād
198	Hunger Cure	Deli / Bodega	28.394460	77.329733	Farīdābād
199	Pizza Hut	Pizza Place	28.397976	77.313145	Farīdābād

280	Coriander Leaf	Indian Restaurant	28.405901	77.045079	Gurgaon
281	More than Paranthas	Restaurant	28.397856	77.312765	Farīdābād
282	Radisson BLU	Hotel	28.449954	77.527910	Greater Noida
283	Subway	Sandwich Place	28.421112	77.090562	Gurgaon
284	Haldiram's	Indian Sweet Shop	28.458301	77.527828	Greater Noida
286	Taj, The Gateway Resort, Damdama Lake	Resort	28.301366	77.123282	Gurgaon
287	Subway	Sandwich Place	28.413677	77.072352	Gurgaon
288	Domino's Pizza	Pizza Place	28.422000	77.087000	Gurgaon
289	Casablanca - DoubleTree by Hilton	Moroccan Restaurant	28.422498	77.104746	Gurgaon
290	Jain Shikanji	Juice Bar	28.830293	77.571528	Modinagar
291	Harish Bakery	Bakery	28.428250	77.099828	Gurgaon
292	Subway	Sandwich Place	28.394843	77.322007	Farīdābād
293	Cafe Coffee Day	Café	28.398256	77.299131	Farīdābād
294	Fill Up	Fast Food Restaurant	28.468363	77.535987	Greater Noida
297	Pahalwan Da Dhaba	Indian Restaurant	29.031215	77.071786	Rāi
298	Karma Lake Land Golf Course	Golf Course	28.356872	76.938551	Gurgaon
299	Big Bazaar	Supermarket	28.419886	77.040358	Gurgaon
380	Subway	Sandwich Place	28.418435	77.038667	Gurgaon
381	Baskin-Robbins	Ice Cream Shop	28.413002	77.043069	Gurgaon
383	Cafe Coffee Day	Café	28.395455	77.323707	Farīdābād
384	Domino's Pizza	Pizza Place	28.410000	77.048000	Gurgaon

```python
In [17]: foursquare_venues.groupby(['city','categories']).count()
```

Out[17]:

city	categories	name	lat	lng
Delhi	Café	1	1	1
	Hotel	1	1	1
	Indian Restaurant	2	2	2
	Italian Restaurant	1	1	1
	Lounge	1	1	1
	Restaurant	1	1	1
	Speakeasy	1	1	1
Farīdābād	Café	3	3	3
	Deli / Bodega	1	1	1
	Fried Chicken Joint	1	1	1
	Hotel	1	1	1
	Indian Restaurant	1	1	1
	Multiplex	1	1	1
	Pizza Place	3	3	3
	Restaurant	1	1	1
	Sandwich Place	1	1	1

Ghāziābād	Food Court	1	1	1
Greater Noida	Fast Food Restaurant	3	3	3
	Hotel	2	2	2
	Indian Sweet Shop	1	1	1
	Monument / Landmark	1	1	1
	Multiplex	1	1	1
	Racetrack	1	1	1
	Shopping Mall	2	2	2
Gurgaon	Bagel Shop	1	1	1
	Bakery	1	1	1
	Bowling Alley	1	1	1
	Brewery	2	2	2
	Café	5	5	5
	Clothing Store	1	1	1
	Coffee Shop	2	2	2
	Falafel Restaurant	1	1	1
	Fast Food Restaurant	2	2	2
	General Entertainment	1	1	1
	Golf Course	1	1	1
	Hotel	5	5	5
	Ice Cream Shop	1	1	1

Mahrauli	Thai Restaurant	1	1	1
Meerut	Juice Bar	1	1	1
Modinagar	Juice Bar	1	1	1
Murthal	Indian Restaurant	1	1	1
NCR	Hotel	1	1	1
New Delhi	Art Gallery	1	1	1
	Art Museum	1	1	1
	Asian Restaurant	1	1	1
	Bakery	1	1	1
	Bar	3	3	3
	Boutique	1	1	1
	Café	3	3	3
	Coffee Shop	1	1	1
	Department Store	1	1	1
	Gastropub	1	1	1
	Golf Course	1	1	1
	Gym / Fitness Center	1	1	1
	Historic Site	1	1	1
	Hotel	8	8	8
	Ice Cream Shop	1	1	1
	Indian Restaurant	7	7	7
Noida	Pub	1	1	1
Rāi	Indian Restaurant	1	1	1
Sohna	Hotel	1	1	1
Sonīpat	Fast Food Restaurant	1	1	1
	Indian Restaurant	1	1	1
	Park	1	1	1
	Pizza Place	1	1	1

In [18]: `foursquare_venues.replace('Mahrauli','Gurgaon',inplace=True)`

Let's group the data by city and categories

In [19]: `foursquare_venues.groupby(['city','categories']).count()`

city	categories	name	lat	lng
Delhi	Café	1	1	1
	Hotel	1	1	1
	Indian Restaurant	2	2	2
	Italian Restaurant	1	1	1
	Lounge	1	1	1
	Restaurant	1	1	1
	Speakeasy	1	1	1
Farīdābād	Café	3	3	3
	Deli / Bodega	1	1	1
	Fried Chicken Joint	1	1	1
	Hotel	1	1	1
	Indian Restaurant	1	1	1
	Multiplex	1	1	1
	Pizza Place	3	3	3
	Restaurant	1	1	1
	Sandwich Place	1	1	1
Ghāziābād	Food Court	1	1	1
Noida	Pub	1	1	1
Rāi	Indian Restaurant	1	1	1
Sohna	Hotel	1	1	1
Sonīpat	Fast Food Restaurant	1	1	1
	Indian Restaurant	1	1	1
	Park	1	1	1
	Pizza Place	1	1	1

	Fast Food Restaurant	3	3	3
	Hotel	2	2	2
	Indian Sweet Shop	1	1	1
Greater Noida	Monument / Landmark	1	1	1
	Multiplex	1	1	1
	Racetrack	1	1	1
	Shopping Mall	2	2	2
	Bagel Shop	1	1	1
	Bakery	1	1	1
	Bowling Alley	1	1	1
	Brewery	2	2	2
	Café	5	5	5
	Clothing Store	1	1	1
	Coffee Shop	2	2	2
	Falafel Restaurant	1	1	1
	Fast Food Restaurant	2	2	2
	General Entertainment	1	1	1
	Golf Course	1	1	1
	Hotel	5	5	5
Gurgaon	Ice Cream Shop	1	1	1
	Indian Restaurant	7	7	7
	Market	2	2	2
	Moroccan Restaurant	1	1	1
	Park	1	1	1
	Pizza Place	3	3	3
	Plaza	1	1	1
	Resort	1	1	1
	Restaurant	1	1	1
	Sandwich Place	3	3	3
	Shopping Mall	2	2	2
	Supermarket	1	1	1
	Thai Restaurant	1	1	1

Meerut	Juice Bar	1	1	1
Modinagar	Juice Bar	1	1	1
Murthal	Indian Restaurant	1	1	1
NCR	Hotel	1	1	1
New Delhi	Art Gallery	1	1	1
	Art Museum	1	1	1
	Asian Restaurant	1	1	1
	Bakery	1	1	1
	Bar	3	3	3
	Boutique	1	1	1
	Café	3	3	3
	Coffee Shop	1	1	1
	Department Store	1	1	1
	Gastropub	1	1	1
	Golf Course	1	1	1
	Gym / Fitness Center	1	1	1
	Historic Site	1	1	1
	Hotel	8	8	8
	Ice Cream Shop	1	1	1
	Indian Restaurant	7	7	7
	Irani Cafe	1	1	1
	Italian Restaurant	4	4	4
	Lounge	1	1	1
	Market	3	3	3
	Mediterranean Restaurant	1	1	1
	Monument / Landmark	2	2	2
	Movie Theater	1	1	1
	Multiplex	1	1	1
	North Indian Restaurant	1	1	1
	Park	1	1	1
	Plaza	1	1	1
	Restaurant	2	2	2
	Shopping Mall	3	3	3
	Smoke Shop	1	1	1
	South Indian Restaurant	1	1	1

Analysing the venues

In [20]:
```python
# one hot encoding
df_onehot = pd.get_dummies(foursquare_venues[['categories']], prefix="", prefix_sep="")

# add postal, borough and neighborhood column back to dataframe

df_onehot['city'] = foursquare_venues[['city']]

# move postal, borough and neighborhood column to the first column
fixed_columns = list(df_onehot.columns[-1:]) + list(df_onehot.columns[:-1])
df_onehot = df_onehot[fixed_columns]

print(df_onehot.shape)
df_onehot.head(10)
```

(150, 54)

Out[20]:

	city	Art Gallery	Art Museum	Asian Restaurant	Bagel Shop	Bakery	Bar	Boutique	Bowling Alley	Brewery	Café	Clothing Store	Coffee Shop	Deli / Bodega	Department Store	Falafel Restaurant	Fast Food Restaurant
0	New Delhi	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	New Delhi	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	New Delhi	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	New Delhi	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5	New Delhi	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6	New Delhi	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7	New Delhi	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
8	New Delhi	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
9	New Delhi	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
10	New Delhi	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Next, let's group rows by city and by taking the mean of the frequency of occurrence of each category

In [21]:
```python
df_grouped = df_onehot.groupby(["city"]).mean().reset_index()

print(df_grouped.shape)
df_grouped
```

(14, 54)

Out[21]:

	city	Art Gallery	Art Museum	Asian Restaurant	Bagel Shop	Bakery	Bar	Boutique	Bowling Alley	Brewery	Café	Clothing Store	Coffee Shop	Deli / Bodega	Departm Store
0	Delhi	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.125000	0.000000	0.000000	0.000000	0.000000
1	Farīdābād	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.230769	0.000000	0.000000	0.076923	0.000000
2	Ghāziābād	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
3	Greater Noida	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
4	Gurgaon	0.000000	0.000000	0.000000	0.020833	0.020833	0.000000	0.000000	0.020833	0.041667	0.104167	0.020833	0.041667	0.000000	0.000000
5	Meerut	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
6	Modinagar	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
7	Murthal	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
8	NCR	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
9	New Delhi	0.017241	0.017241	0.017241	0.000000	0.017241	0.051724	0.017241	0.000000	0.000000	0.051724	0.000000	0.017241	0.000000	0.017241
10	Noida	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
11	Rāi	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
12	Sohna	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
13	Sonīpat	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000

Let's write a function to sort the venues in descending order.

```
In [22]: def return_most_common_venues(row, num_top_venues):
             row_categories = row.iloc[1:]
             row_categories_sorted = row_categories.sort_values(ascending=False)

             return row_categories_sorted.index.values[0:num_top_venues]
```

Now let's create the new dataframe and display the top 10 venues for each City.

```
In [48]: num_top_venues = 10

         indicators = ['st', 'nd', 'rd']

         # create columns according to number of top venues
         columns = ['city']
         for ind in np.arange(num_top_venues):
             try:
                 columns.append('{}{} Most Common Venue'.format(ind+1, indicators[ind]))
             except:
                 columns.append('{}th Most Common Venue'.format(ind+1))

         # create a new dataframe
         df_sorted = pd.DataFrame(columns=columns)
         df_sorted['city'] = df_grouped['city']

         for ind in np.arange(df_grouped.shape[0]):
             df_sorted.iloc[ind, 1:] = return_most_common_venues(df_grouped.iloc[ind, :], num_top_venues)

         df_sorted.head()
```

	city	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Delhi	Indian Restaurant	Speakeasy	Italian Restaurant	Restaurant	Lounge	Café	Hotel	Thai Restaurant	Falafel Restaurant	Gym / Fitness Center
1	Farīdābād	Pizza Place	Café	Fried Chicken Joint	Sandwich Place	Multiplex	Hotel	Restaurant	Deli / Bodega	Indian Restaurant	Gastropub
2	Ghāziābād	Food Court	Thai Restaurant	Indian Restaurant	Hotel	Historic Site	Gym / Fitness Center	Golf Course	General Entertainment	Gastropub	Fried Chicken Joint
3	Greater Noida	Fast Food Restaurant	Shopping Mall	Hotel	Indian Sweet Shop	Multiplex	Racetrack	Monument / Landmark	Gym / Fitness Center	Golf Course	General Entertainment
4	Gurgaon	Indian Restaurant	Café	Hotel	Sandwich Place	Pizza Place	Brewery	Shopping Mall	Fast Food Restaurant	Market	Coffee Shop

Cluster Neighborhoods

Run _k_-means to cluster the neighborhood into 5 clusters. ¶

```
In [49]:  # set number of clusters
          kclusters = 9

          df_grouped_clustering = df_grouped.drop('city', 1)

          # run k-means clustering
          kmeans = KMeans(n_clusters=kclusters).fit(df_grouped_clustering)

          # check cluster labels generated for each row in the dataframe
          kmeans.labels_[0:100]
```

```
Out[49]: array([1, 8, 0, 7, 1, 2, 2, 3, 4, 1, 5, 3, 4, 6], dtype=int32)
```

Let's create a new dataframe that includes the cluster as well as the top 10 venues for each city

```
In [50]:  # add clustering labels
          df_sorted.insert(0, 'Cluster Labels', kmeans.labels_)
```

```
In [ ]:
```

```
In [51]:  df_merged = foursquare_venues

          # merge df_grouped with df_data to add latitude/longitude for each neighborhood
          df_merged = df_merged.join(df_sorted.set_index('city'), on='city')

          df_merged.head() # check the last columns!
```

	name	categories	lat	lng	city	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue
0	The Imperial	Hotel	28.625548	77.218664	New Delhi	1	Hotel	Indian Restaurant	Italian Restaurant	Shopping Mall	Café	Bar	Market	Restaurant	Monument / Landmark
1	Tamra	Restaurant	28.620543	77.218174	New Delhi	1	Hotel	Indian Restaurant	Italian Restaurant	Shopping Mall	Café	Bar	Market	Restaurant	Monument / Landmark
2	Pandey Paan	Smoke Shop	28.622249	77.201075	New Delhi	1	Hotel	Indian Restaurant	Italian Restaurant	Shopping Mall	Café	Bar	Market	Restaurant	Monument / Landmark
3	The Big Chill Cafe	Italian Restaurant	28.600686	77.227636	New Delhi	1	Hotel	Indian Restaurant	Italian Restaurant	Shopping Mall	Café	Bar	Market	Restaurant	Monument / Landmark
5	The Leela Palace	Hotel	28.579923	77.189291	New Delhi	1	Hotel	Indian Restaurant	Italian Restaurant	Shopping Mall	Café	Bar	Market	Restaurant	Monument / Landmark

Now let's visualise the resulting clusters

In [52]:
```python
# create map
map_clusters = folium.Map(location=[latitude, longitude], zoom_start=11)

# set color scheme for the clusters
x = np.arange(kclusters)
ys = [i + x + (i*x)**2 for i in range(kclusters)]
colors_array = cm.rainbow(np.linspace(0, 1, len(ys)))
rainbow = [colors.rgb2hex(i) for i in colors_array]

# add markers to the map
markers_colors = []
for lat, lon, poi, cluster in zip(df_merged['lat'], df_merged['lng'],df_merged['name'],df_merged['Cluster Labels']):
    label = folium.Popup(str(poi) + ' Cluster ' + str(cluster), parse_html=True)
    folium.CircleMarker(
        [lat, lon],
        radius=5,
        popup=label,
        color=rainbow[cluster-1],
        fill=True,
        fill_color=rainbow[cluster-1],
        fill_opacity=0.7).add_to(map_clusters)

map_clusters
```

Out[52]:

Conclusion

According to the results obtained, it is more profitable to open malls in Delhi or Gurgaon Region.