Name: Sagar Tarasariya

Capstone Project - the Battle of Neighborhoods

Introduction

Ahmedabad is located near the banks of the Sabarmati River, 25 km (16 mi), from the capital of Gujarat, Gandhinagar, also known as its twin city. Ahmedabad has emerged as an important economic and industrial hub in India.

And also Ahmedabad is very much famous for Food all over Gujarat.

There are many restaurants in Ahmedabad City, each belonging to different categories like Chinese, Italian, North Indian, French, etc.

So as part of this project, we will list and visualize all major parts of Ahmedabad, Gujarat.

Dataset:

For this project we need the following data:

Ahmedabad restaurants data that contains list Locality, Restaurant name, rating along with their latitude and longitude.

Data source: Zomato Indian Restaurant kaggel dataset->Dataset

Description: This data set contains the required information. And we will use this data set to explore various locality of Ahmedabad city.

Nearby places in each locality of Ahmedabad city.

Data source: Foursquare API->Foursquare

Description: By using this API we will get all the venues in each neighborhood.

Questions that can be asked using the above mentioned datasets:

Q: which places are have best restaurant in Ahmedabad?

Q: Which Restaurant have most outlets in Ahmedabad city?

Q: Top Quick Bites restaurant in Ahmedabad?

Q: Top Casual Dining Restaurants in Ahmedabad?

Name: Sagar Tarasariya

Q: Which place are suitable for edible person in Ahmedabad city?

Q: What are the best Chinese restaurant in Bodakdev?

Q: How many Restaurant Accepting Digital Payment in Ahmedabad?

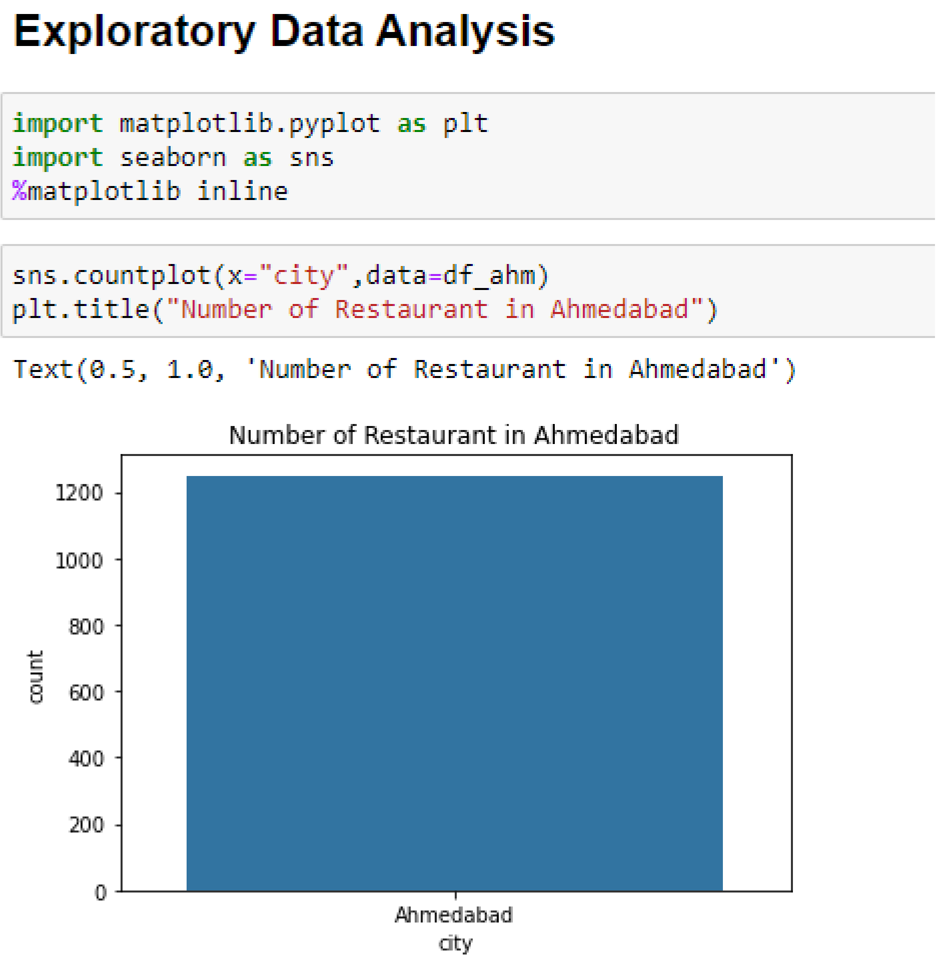
Approach Followed:

- Collect the Ahmedabad city data from Zomato kaggel dataset
- Using Foursquare API we will find all venues for each neighborhood.
- Filter out all venues that are nearby by locality.
- Using aggregative rating for each restaurant to find the best places.
- Visualize the Ranking of neighborhoods using folium library(python)

Required Libraries:

- **Numpy**
- **Pandas**
- **Matplotlib**
- **Sklearn**
- **Folium**
- **Geocoder**

Name: Sagar Tarasariya



Exploratory Data Analysis

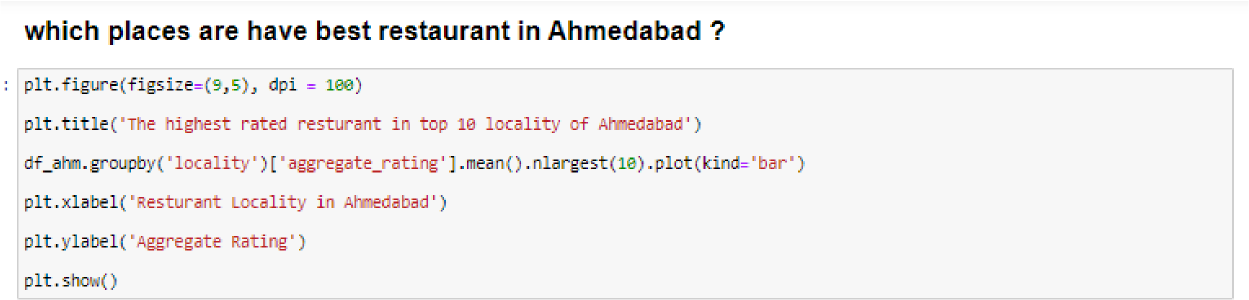
```
import matplotlib.pyplot as plt
import seaborn as sns
%matplotlib inline
```

```
sns.countplot(x="city",data=df_ahm)
plt.title("Number of Restaurant in Ahmedabad")
```

```
Text(0.5, 1.0, 'Number of Restaurant in Ahmedabad')
```

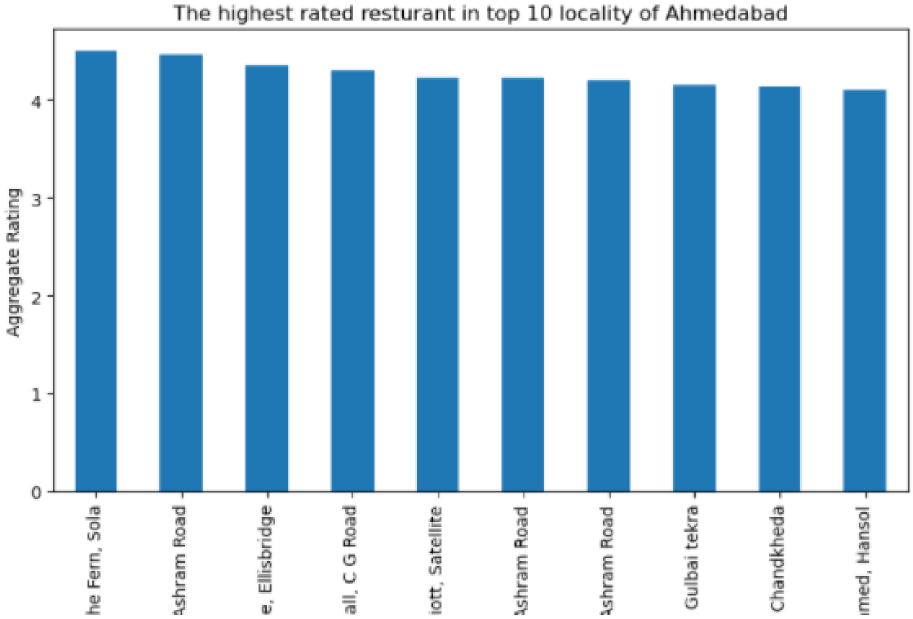
By using Count Plot of seaborn library I have plotted Ahmedabad city's total Restaurant

Name: Sagar Tarasariya



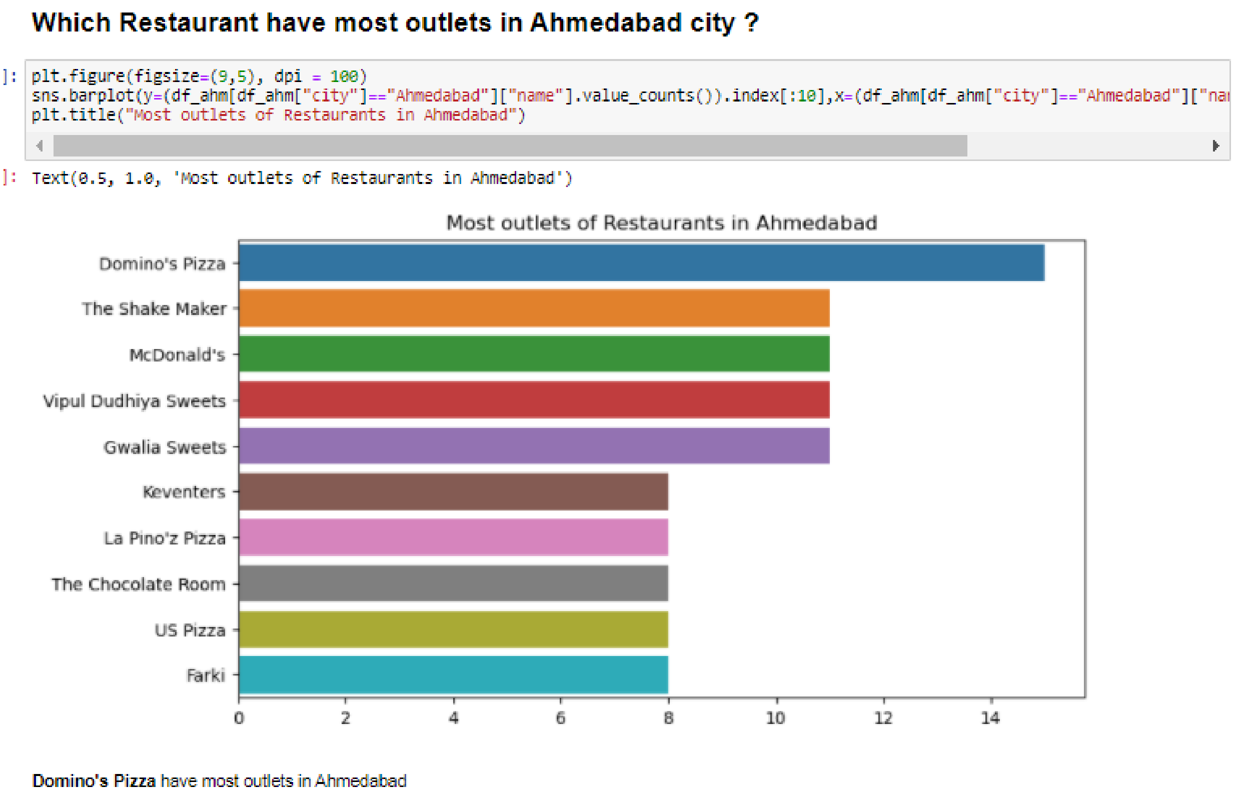
which places are have best restaurant in Ahmedabad ?

```
plt.figure(figsize=(9,5), dpi = 100)
plt.title('The highest rated resturant in top 10 locality of Ahmedabad')
df_ahm.groupby('locality')['aggregate_rating'].mean().nlargest(10).plot(kind='bar')
plt.xlabel('Resturant Locality in Ahmedabad')
plt.ylabel('Aggregate Rating')
plt.show()
```



Ploted by calculating mean of aggregation rating of restaurant

Name: Sagar Tarasariya

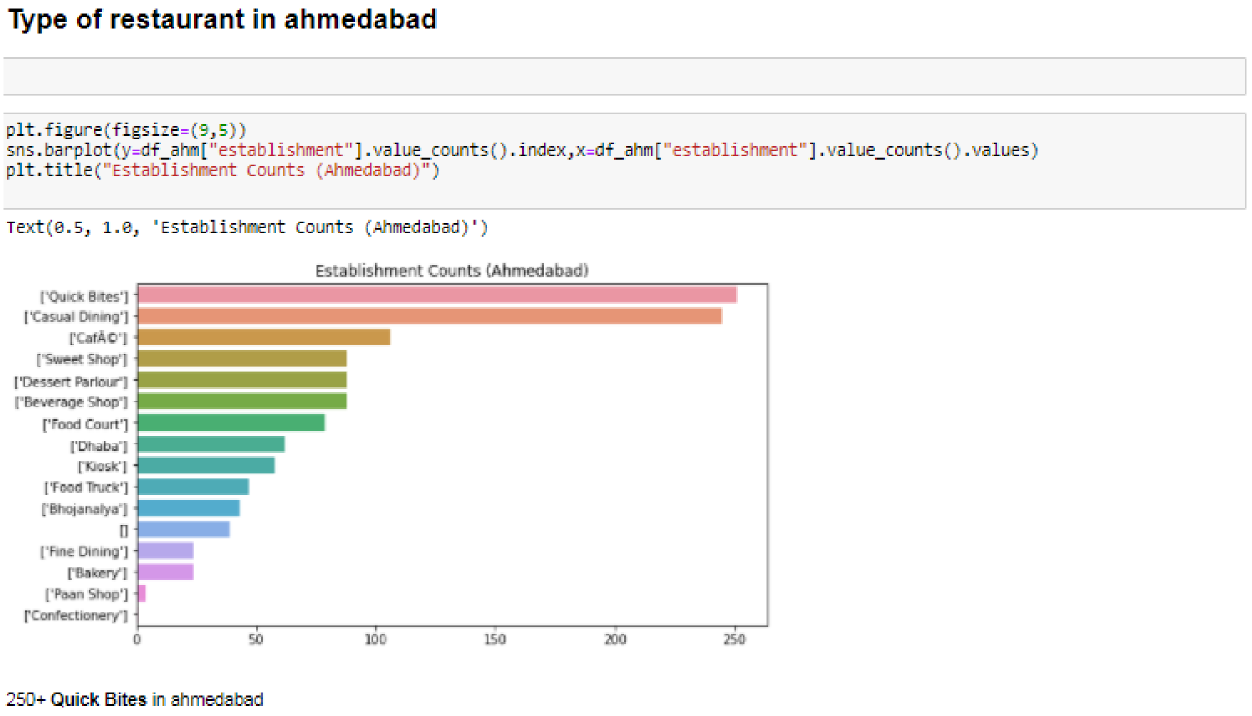


Which Restaurant have most outlets in Ahmedabad city ?

```
]: plt.figure(figsize=(9,5), dpi = 100)
   sns.barplot(y=(df_ahm[df_ahm["city"]=="Ahmedabad"]["name"].value_counts()).index[:10],x=(df_ahm[df_ahm["city"]=="Ahmedabad"]["na
   plt.title("Most outlets of Restaurants in Ahmedabad")

]: Text(0.5, 1.0, 'Most outlets of Restaurants in Ahmedabad')
```

Domino's Pizza have most outlets in Ahmedabad

In this I have just Counted Number Of restaurant in Ahmedabad city.



Type of restaurant in ahmedabad

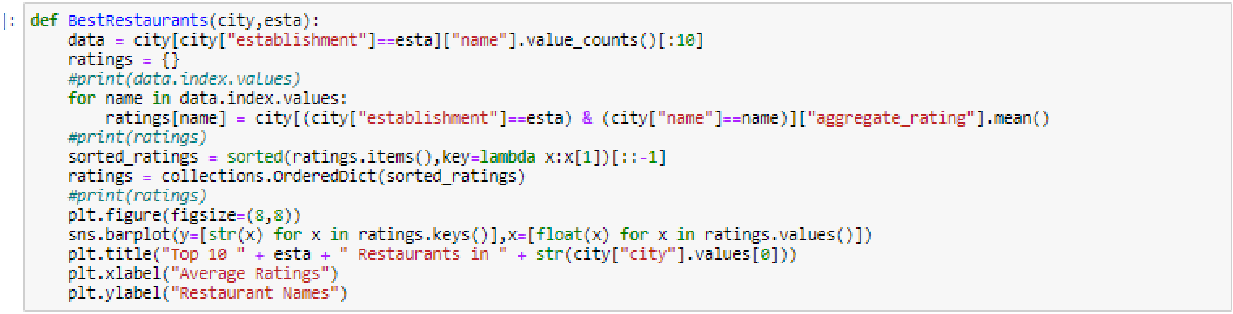
```
plt.figure(figsize=(9,5))
sns.barplot(y=df_ahm["establishment"].value_counts().index,x=df_ahm["establishment"].value_counts().values)
plt.title("Establishment Counts (Ahmedabad)")

Text(0.5, 1.0, 'Establishment Counts (Ahmedabad)')
```

250+ **Quick Bites** in ahmedabad

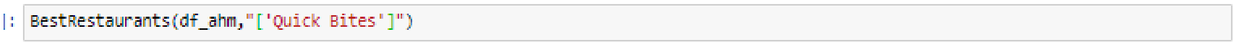
Type of Restaurant by counting Establishment.

Name: Sagar Tarasariya

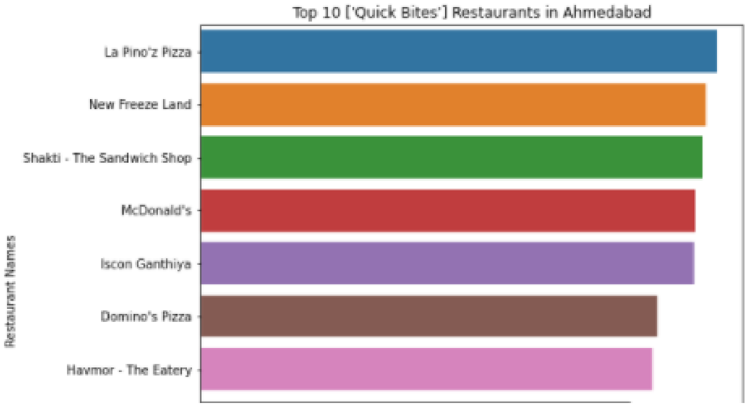
Top Quick Bites restaurant in Ahmedabad



```
def BestRestaurants(city,esta):
    data = city[city["establishment"]==esta]["name"].value_counts()[:10]
    ratings = {}
    #print(data.index.values)
    for name in data.index.values:
        ratings[name] = city[(city["establishment"]==esta) & (city["name"]==name)]["aggregate_rating"].mean()
    #print(ratings)
    sorted_ratings = sorted(ratings.items(),key=lambda x:x[1])[::-1]
    ratings = collections.OrderedDict(sorted_ratings)
    #print(ratings)
    plt.figure(figsize=(8,8))
    sns.barplot(y=[str(x) for x in ratings.keys()],x=[float(x) for x in ratings.values()])
    plt.title("Top 10 " + esta + " Restaurants in " + str(city["city"].values[0]))
    plt.xlabel("Average Ratings")
    plt.ylabel("Restaurant Names")
```

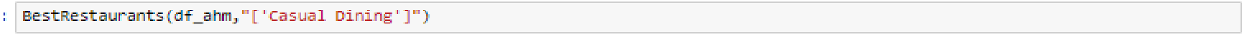


```
BestRestaurants(df_ahm,"['Quick Bites']")
```

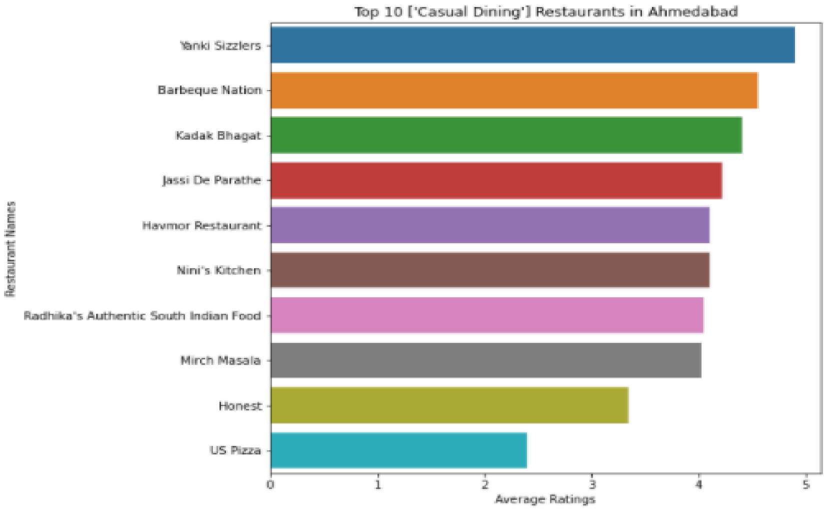


By putting condition that establishment must be Quick Bites and then on that Establishments aggregation rating's mean counted and printed top rated Quick bites restaurant.

Top Casual Dining Restaurants in Ahmedabad



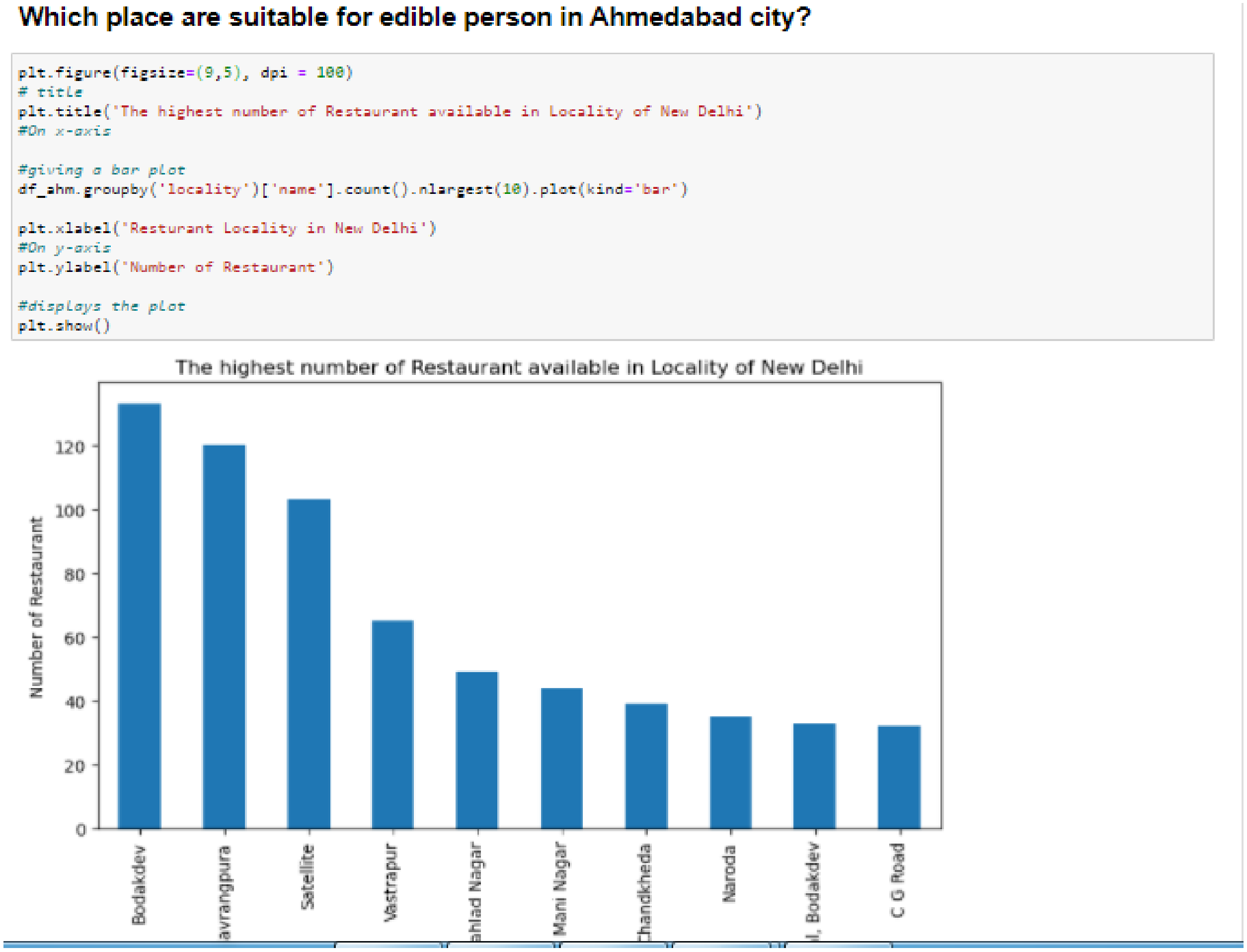
```
BestRestaurants(df_ahm,"['Casual Dining']")
```



Yanki Sizzlers is No.1 Casual Dining Restaurant

Name: Sagar Tarasariya

Same function with just passing Casual Dining Restaurant



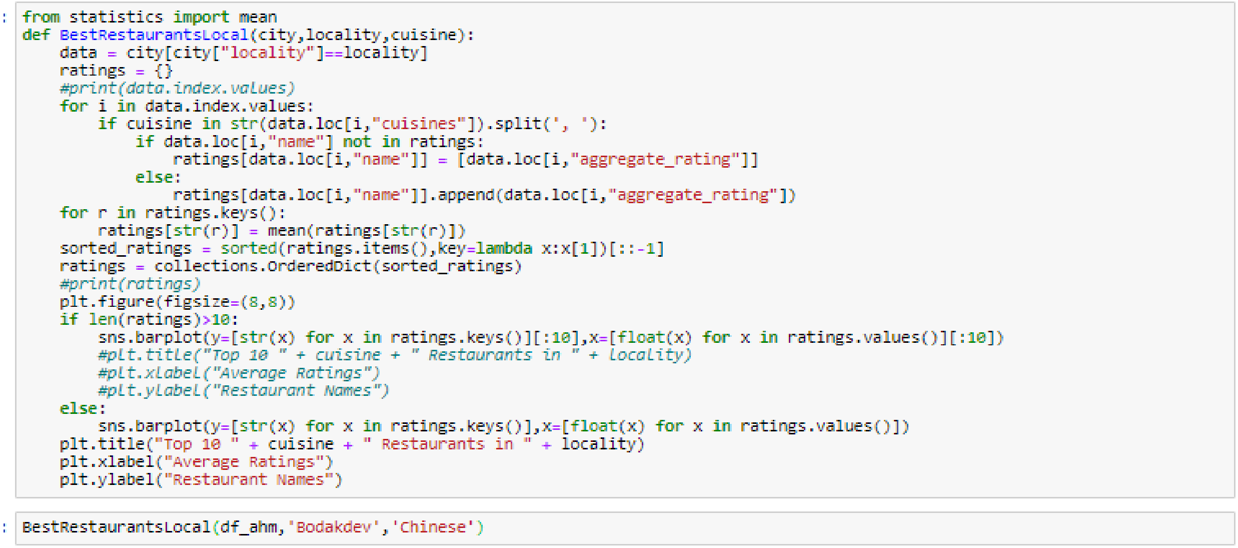
Which place are suitable for edible person in Ahmedabad city?

```
plt.figure(figsize=(9,5), dpi = 100)
# title
plt.title('The highest number of Restaurant available in Locality of New Delhi')
#On x-axis

#giving a bar plot
df_ahm.groupby('locality')['name'].count().nlargest(10).plot(kind='bar')

plt.xlabel('Resturant Locality in New Delhi')
#On y-axis
plt.ylabel('Number of Restaurant')

#displays the plot
plt.show()
```

Counted number of Restaurant in each area->most restaurant are there in area which is best for edible person.

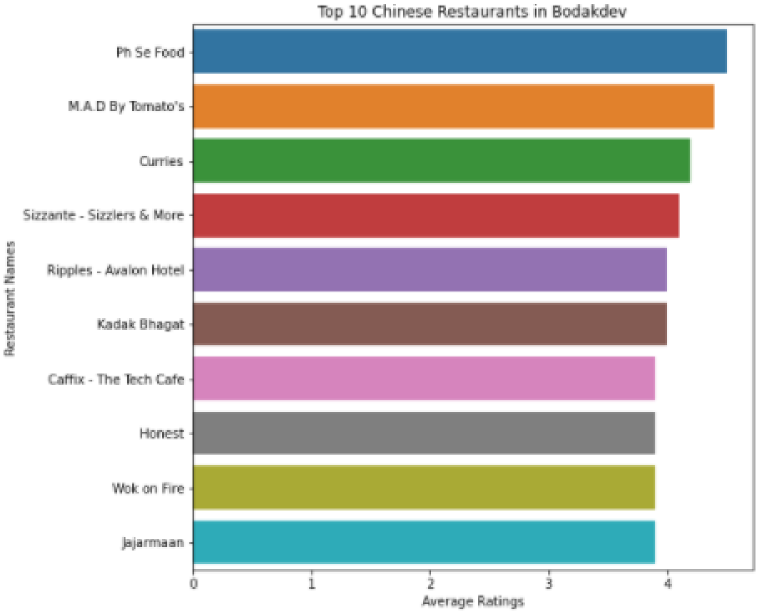
Name: Sagar Tarasariya

What are the best chinese restaurant in Bodakdev?



```
from statistics import mean
def BestRestaurantsLocal(city,locality,cuisine):
    data = city[city["locality"]==locality]
    ratings = {}
    #print(data.index.values)
    for i in data.index.values:
        if cuisine in str(data.loc[i,"cuisines"]).split(', '):
            if data.loc[i,"name"] not in ratings:
                ratings[data.loc[i,"name"]] = [data.loc[i,"aggregate_rating"]]
            else:
                ratings[data.loc[i,"name"]].append(data.loc[i,"aggregate_rating"])
    for r in ratings.keys():
        ratings[str(r)] = mean(ratings[str(r)])
    sorted_ratings = sorted(ratings.items(),key=lambda x:x[1])[::-1]
    ratings = collections.OrderedDict(sorted_ratings)
    #print(ratings)
    plt.figure(figsize=(8,8))
    if len(ratings)>10:
        sns.barplot(y=[str(x) for x in ratings.keys()][:10],x=[float(x) for x in ratings.values()][:10])
        #plt.title("Top 10 " + cuisine + " Restaurants in " + locality)
        #plt.xlabel("Average Ratings")
        #plt.ylabel("Restaurant Names")
    else:
        sns.barplot(y=[str(x) for x in ratings.keys()],x=[float(x) for x in ratings.values()])
    plt.title("Top 10 " + cuisine + " Restaurants in " + locality)
    plt.xlabel("Average Ratings")
    plt.ylabel("Restaurant Names")
```

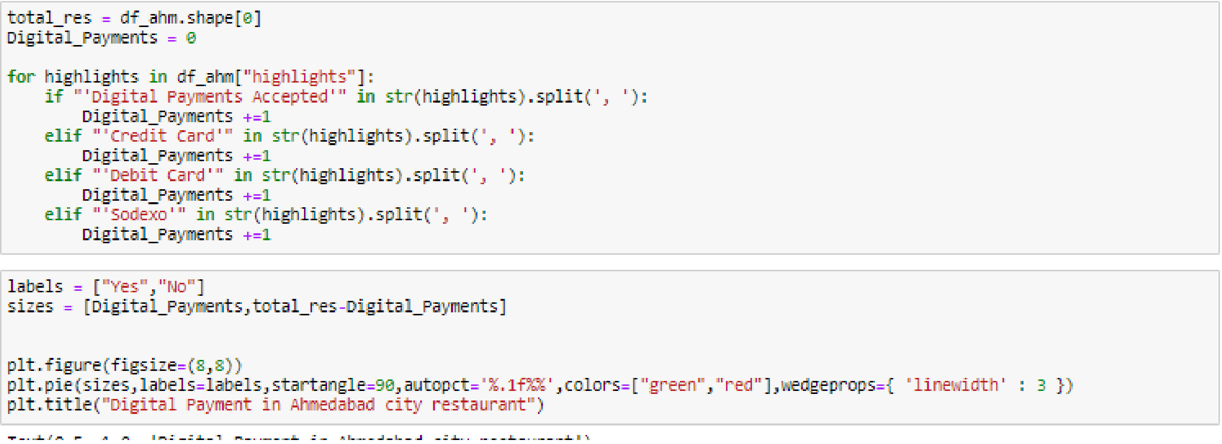
```
BestRestaurantsLocal(df_ahm,'Bodakdev','Chinese')
```



Ph Se Food is top rated Chinese restaurant in **bodakdev** area.

Name: Sagar Tarasariya

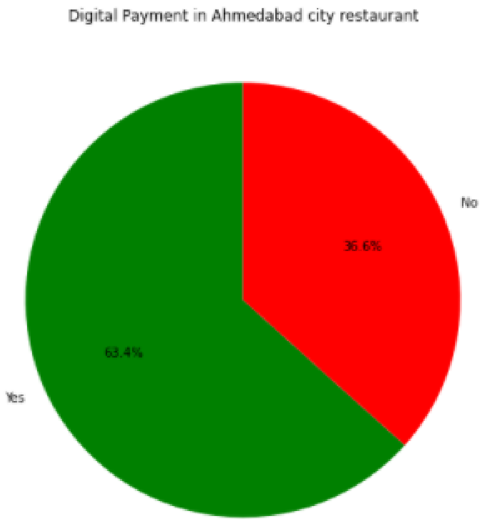
How many Restaurant Accepting Digital Payment in Ahmedabad?



```
total_res = df_ahm.shape[0]
Digital_Payments = 0

for highlights in df_ahm["highlights"]:
    if "'Digital Payments Accepted'" in str(highlights).split(', '):
        Digital_Payments +=1
    elif "'Credit Card'" in str(highlights).split(', '):
        Digital_Payments +=1
    elif "'Debit Card'" in str(highlights).split(', '):
        Digital_Payments +=1
    elif "'Sodexo'" in str(highlights).split(', '):
        Digital_Payments +=1
```

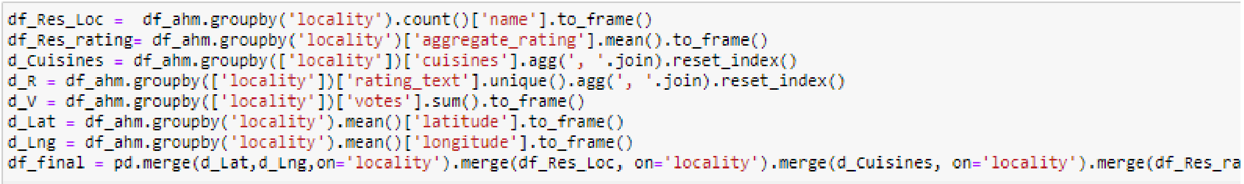
```
labels = ["Yes","No"]
sizes = [Digital_Payments,total_res-Digital_Payments]


plt.figure(figsize=(8,8))
plt.pie(sizes,labels=labels,startangle=90,autopct='%.1f%%',colors=["green","red"],wedgeprops={ 'linewidth' : 3 })
plt.title("Digital Payment in Ahmedabad city restaurant")
```

```
Text(0.5, 1.0, 'Digital Payment in Ahmedabad city restaurant')
```



More than **60%** restaurant accepting **Digital Payment** in Ahmedabad city

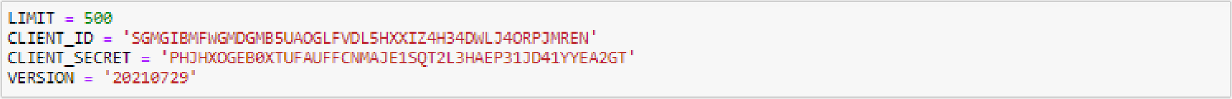
In highlight column of dataset payment method data available so what I did is I have searched for particular payment method in that column and counted how many restaurant are accepting digital payment and many are not.



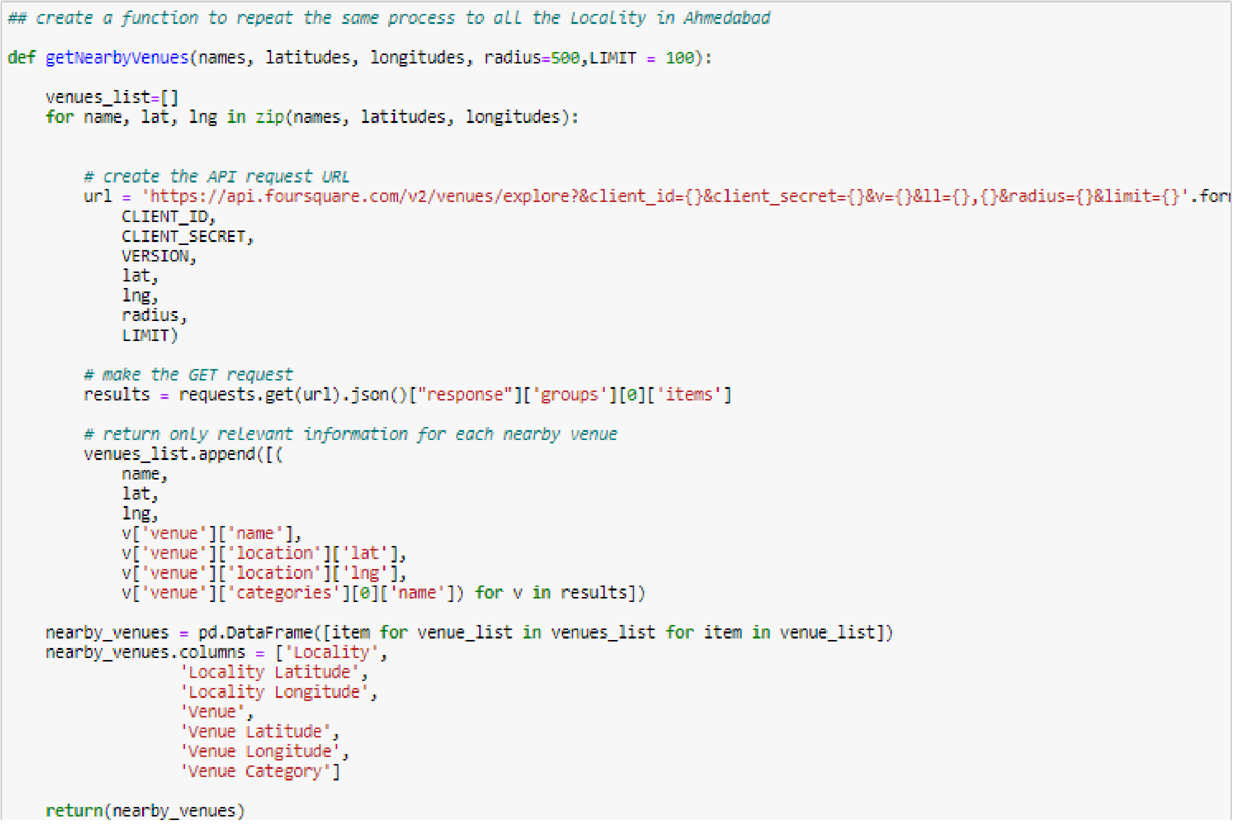
```
df_Res_Loc =  df_ahm.groupby('locality').count()['name'].to_frame()
df_Res_rating= df_ahm.groupby('locality')['aggregate_rating'].mean().to_frame()
d_Cuisines = df_ahm.groupby(['locality'])['cuisines'].agg(', '.join).reset_index()
d_R = df_ahm.groupby(['locality'])['rating_text'].unique().agg(', '.join).reset_index()
d_V = df_ahm.groupby(['locality'])['votes'].sum().to_frame()
d_Lat = df_ahm.groupby('locality').mean()['latitude'].to_frame()
d_Lng = df_ahm.groupby('locality').mean()['longitude'].to_frame()
df_final = pd.merge(d_Lat,d_Lng,on='locality').merge(df_Res_Loc, on='locality').merge(d_Cuisines, on='locality').merge(df_Res_ra
```

Prepared data frame for getting neighborhood venue using foursquare API.

Name: Sagar Tarasariya

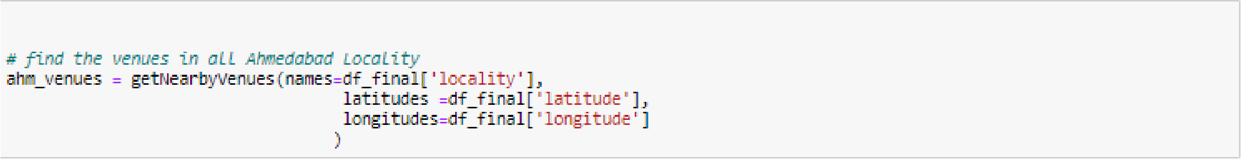


```
LIMIT = 500
CLIENT_ID = 'SGMGIBMFWGMDGMB5UAOGLFVDL5HXXIZ4H34DWLJ4ORPJMREN'
CLIENT_SECRET = 'PHJHXOGEB0XTUFAUFFCNMAJE1SQT2L3HAEP31JD41YYEA2GT'
VERSION = '20210729'
```

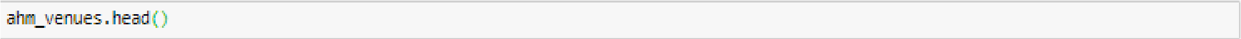


```
## create a function to repeat the same process to all the Locality in Ahmedabad

def getNearbyVenues(names, latitudes, longitudes, radius=500,LIMIT = 100):

    venues_list=[]
    for name, lat, lng in zip(names, latitudes, longitudes):


        # create the API request URL
        url = 'https://api.foursquare.com/v2/venues/explore?&client_id={}&client_secret={}&v={}&ll={},{}&radius={}&limit={}'.for
            CLIENT_ID,
            CLIENT_SECRET,
            VERSION,
            lat,
            lng,
            radius,
            LIMIT)

        # make the GET request
        results = requests.get(url).json()["response"]['groups'][0]['items']

        # return only relevant information for each nearby venue
        venues_list.append([(
            name,
            lat,
            lng,
            v['venue']['name'],
            v['venue']['location']['lat'],
            v['venue']['location']['lng'],
            v['venue']['categories'][0]['name']) for v in results])

    nearby_venues = pd.DataFrame([item for venue_list in venues_list for item in venue_list])
    nearby_venues.columns = ['Locality',
                  'Locality Latitude',
                  'Locality Longitude',
                  'Venue',
                  'Venue Latitude',
                  'Venue Longitude',
                  'Venue Category']

    return(nearby_venues)
```

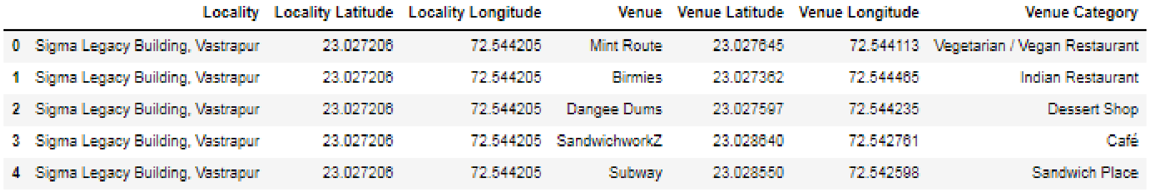
Passing Foursquare API credential like secret ID, Client ID, Server etc. and creating API request, in which data are coming in json format so converting it to list and then creating new data frame.



```
# find the venues in all Ahmedabad Locality
ahm_venues = getNearbyVenues(names=df_final['locality'],
                                   latitudes =df_final['latitude'],
                                   longitudes=df_final['longitude']
                                  )
```



```
ahm_venues.head()
```

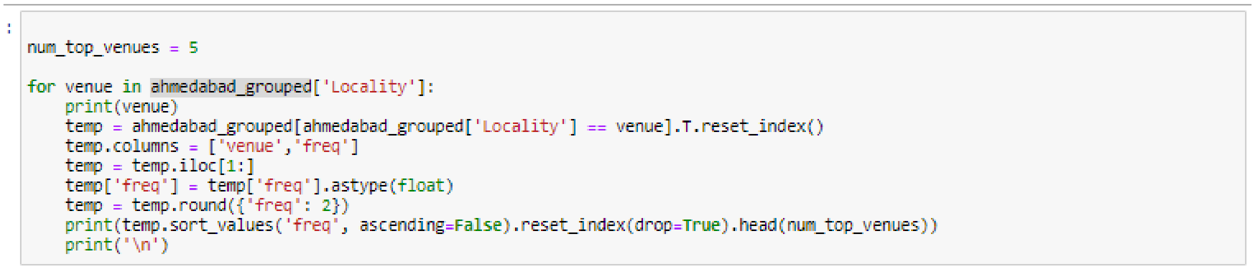


	Locality	Locality Latitude	Locality Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Sigma Legacy Building, Vastrapur	23.027206	72.544205	Mint Route	23.027645	72.544113	Vegetarian / Vegan Restaurant
1	Sigma Legacy Building, Vastrapur	23.027206	72.544205	Birmies	23.027362	72.544465	Indian Restaurant
2	Sigma Legacy Building, Vastrapur	23.027206	72.544205	Dangee Dums	23.027597	72.544235	Dessert Shop
3	Sigma Legacy Building, Vastrapur	23.027206	72.544205	SandwichworkZ	23.028640	72.542761	Café
4	Sigma Legacy Building, Vastrapur	23.027206	72.544205	Subway	23.028550	72.542598	Sandwich Place

Here is the output of the venues.

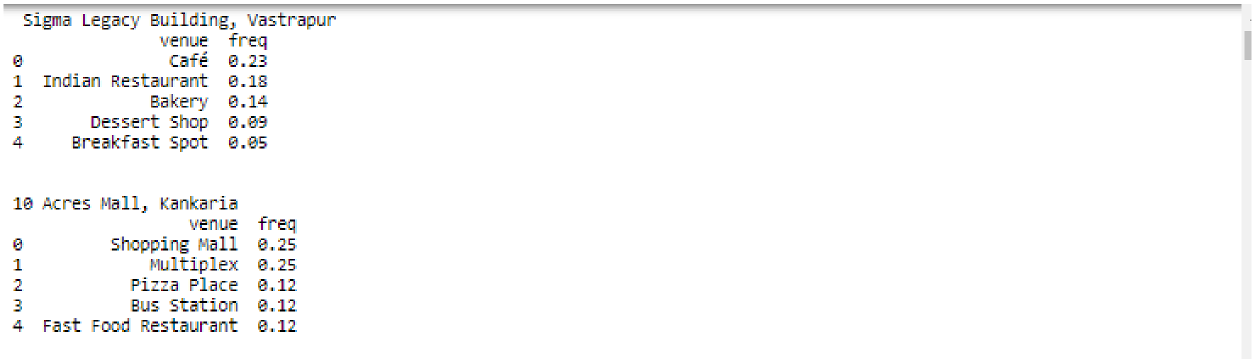
Name: Sagar Tarasariya

Displaying Top 5 most comman venues



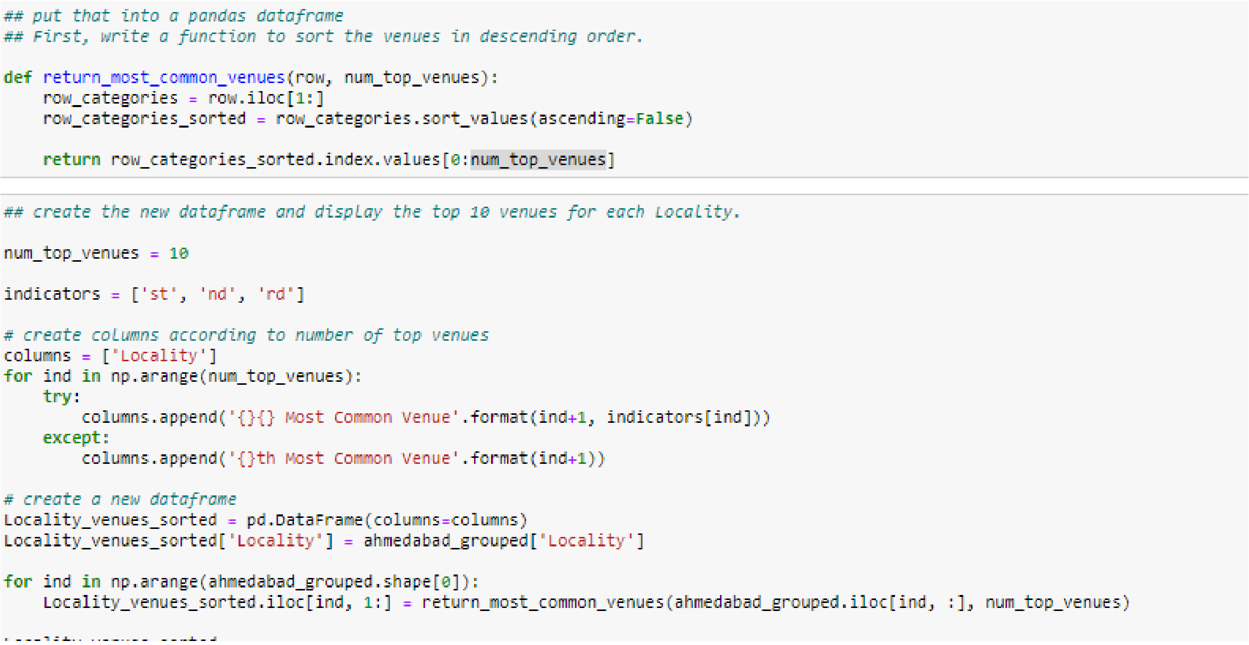
```
:
num_top_venues = 5

for venue in ahmedabad_grouped['Locality']:
    print(venue)
    temp = ahmedabad_grouped[ahmedabad_grouped['Locality'] == venue].T.reset_index()
    temp.columns = ['venue','freq']
    temp = temp.iloc[1:]
    temp['freq'] = temp['freq'].astype(float)
    temp = temp.round({'freq': 2})
    print(temp.sort_values('freq', ascending=False).reset_index(drop=True).head(num_top_venues))
    print('\n')
```

By sorting displaying top 5 most common venues.



```
 Sigma Legacy Building, Vastrapur
               venue  freq
0               Café  0.23
1  Indian Restaurant  0.18
2             Bakery  0.14
3       Dessert Shop  0.09
4     Breakfast Spot  0.05


10 Acres Mall, Kankaria
                  venue  freq
0         Shopping Mall  0.25
1             Multiplex  0.25
2           Pizza Place  0.12
3           Bus Station  0.12
4  Fast Food Restaurant  0.12
```

This is output sample.

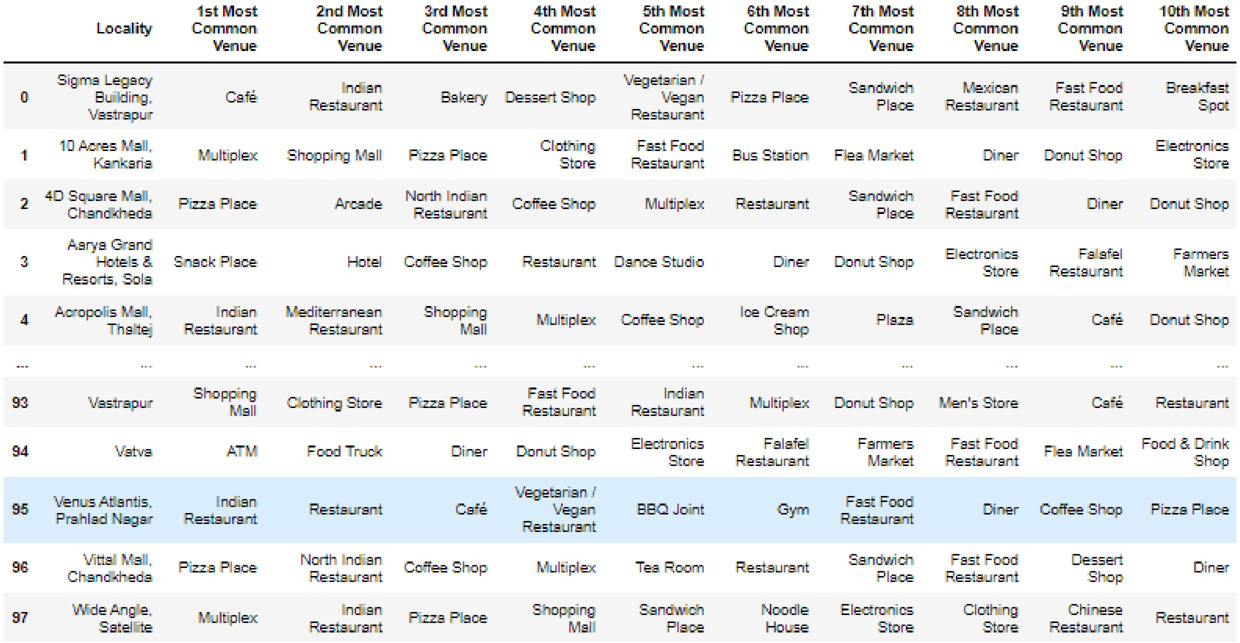


```
## put that into a pandas dataframe
## First, write a function to sort the venues in descending order.

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]
```

```
## create the new dataframe and display the top 10 venues for each Locality.

num_top_venues = 10

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

# create columns according to number of top venues
columns = ['Locality']
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
Locality_venues_sorted = pd.DataFrame(columns=columns)
Locality_venues_sorted['Locality'] = ahmedabad_grouped['Locality']

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

Locality_venues_sorted
```

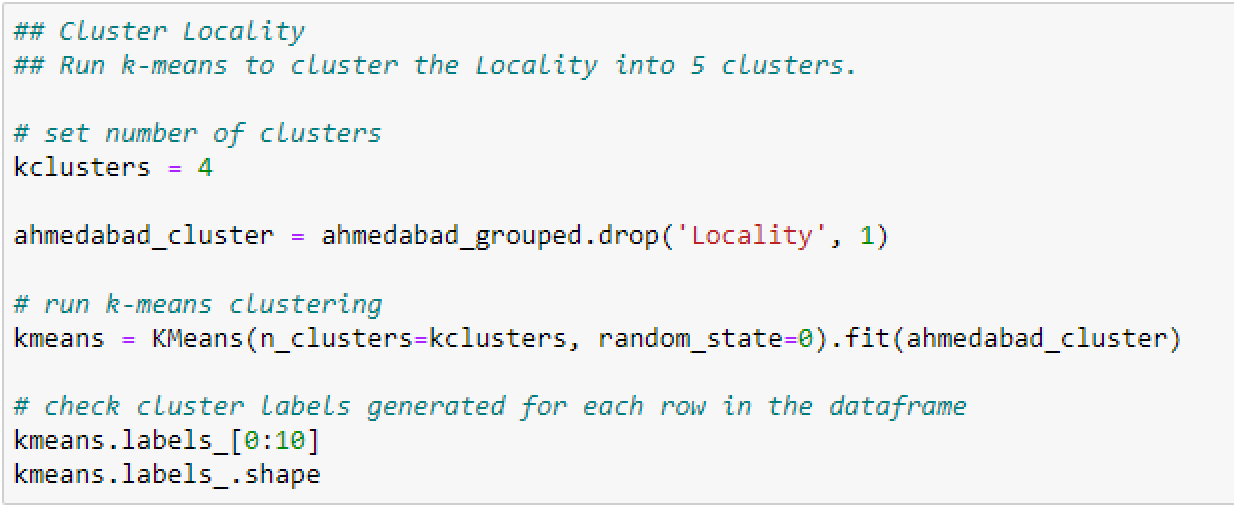
Putting it to the data frame and displaying top 10 most common venues.

Name: Sagar Tarasariya



	Locality	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	Sigma Legacy Building, Vastrapur	Café	Indian Restaurant	Bakery	Dessert Shop	Vegetarian / Vegan Restaurant	Pizza Place	Sandwich Place	Mexican Restaurant	Fast Food Restaurant	Breakfast Spot
1	10 Acres Mall, Kankaria	Multiplex	Shopping Mall	Pizza Place	Clothing Store	Fast Food Restaurant	Bus Station	Flea Market	Diner	Donut Shop	Electronics Store
2	4D Square Mall, Chandkheda	Pizza Place	Arcade	North Indian Restaurant	Coffee Shop	Multiplex	Restaurant	Sandwich Place	Fast Food Restaurant	Diner	Donut Shop
3	Aarya Grand Hotels & Resorts, Sola	Snack Place	Hotel	Coffee Shop	Restaurant	Dance Studio	Diner	Donut Shop	Electronics Store	Falafel Restaurant	Farmers Market
4	Acropolis Mall, Thaltej	Indian Restaurant	Mediterranean Restaurant	Shopping Mall	Multiplex	Coffee Shop	Ice Cream Shop	Plaza	Sandwich Place	Café	Donut Shop
...	...	...	...	...	...	...	...	...	...	...	...
93	Vastrapur	Shopping Mall	Clothing Store	Pizza Place	Fast Food Restaurant	Indian Restaurant	Multiplex	Donut Shop	Men's Store	Café	Restaurant
94	Vatva	ATM	Food Truck	Diner	Donut Shop	Electronics Store	Falafel Restaurant	Farmers Market	Fast Food Restaurant	Flea Market	Food & Drink Shop
95	Venus Atlantis, Prahlad Nagar	Indian Restaurant	Restaurant	Café	Vegetarian / Vegan Restaurant	BBQ Joint	Gym	Fast Food Restaurant	Diner	Coffee Shop	Pizza Place
96	Vittal Mall, Chandkheda	Pizza Place	North Indian Restaurant	Coffee Shop	Multiplex	Tea Room	Restaurant	Sandwich Place	Fast Food Restaurant	Dessert Shop	Diner
97	Wide Angle, Satellite	Multiplex	Indian Restaurant	Pizza Place	Shopping Mall	Sandwich Place	Noodle House	Electronics Store	Clothing Store	Chinese Restaurant	Restaurant

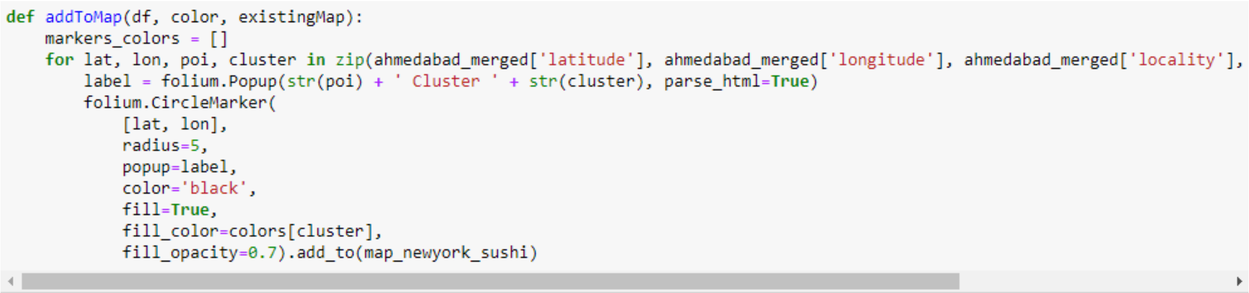
Output.



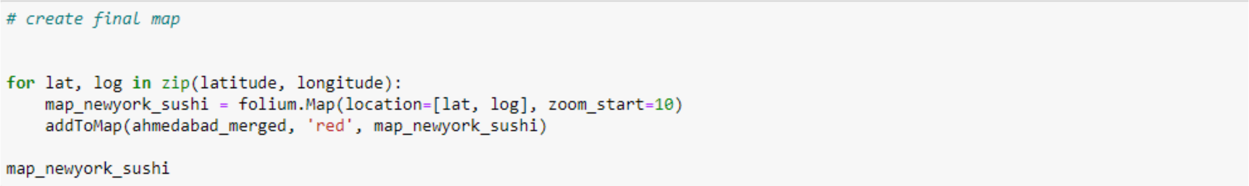
```
## Cluster Locality
## Run k-means to cluster the Locality into 5 clusters.

# set number of clusters
kclusters = 4

ahmedabad_cluster = ahmedabad_grouped.drop('Locality', 1)

# run k-means clustering
kmeans = KMeans(n_clusters=kclusters, random_state=0).fit(ahmedabad_cluster)

# check cluster labels generated for each row in the dataframe
kmeans.labels_[0:10]
kmeans.labels_.shape
```

Performing K-Means-Clustering (4 Cluster)

Name: Sagar Tarasariya

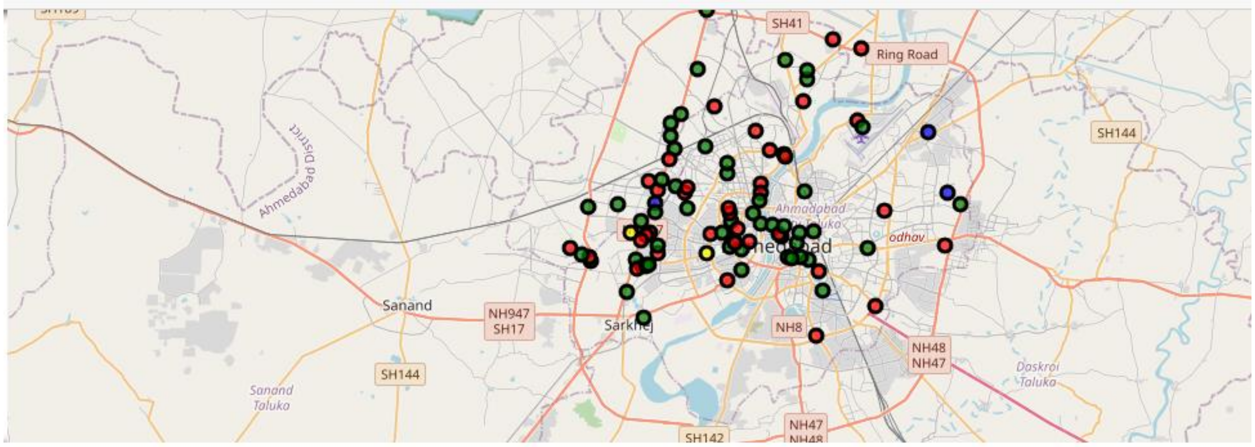


```
def addToMap(df, color, existingMap):
    markers_colors = []
    for lat, lon, poi, cluster in zip(ahmedabad_merged['latitude'], ahmedabad_merged['longitude'], ahmedabad_merged['locality'],
        label = folium.Popup(str(poi) + ' Cluster ' + str(cluster), parse_html=True)
        folium.CircleMarker(
            [lat, lon],
            radius=5,
            popup=label,
            color='black',
            fill=True,
            fill_color=colors[cluster],
            fill_opacity=0.7).add_to(map_newyork_sushi)
```

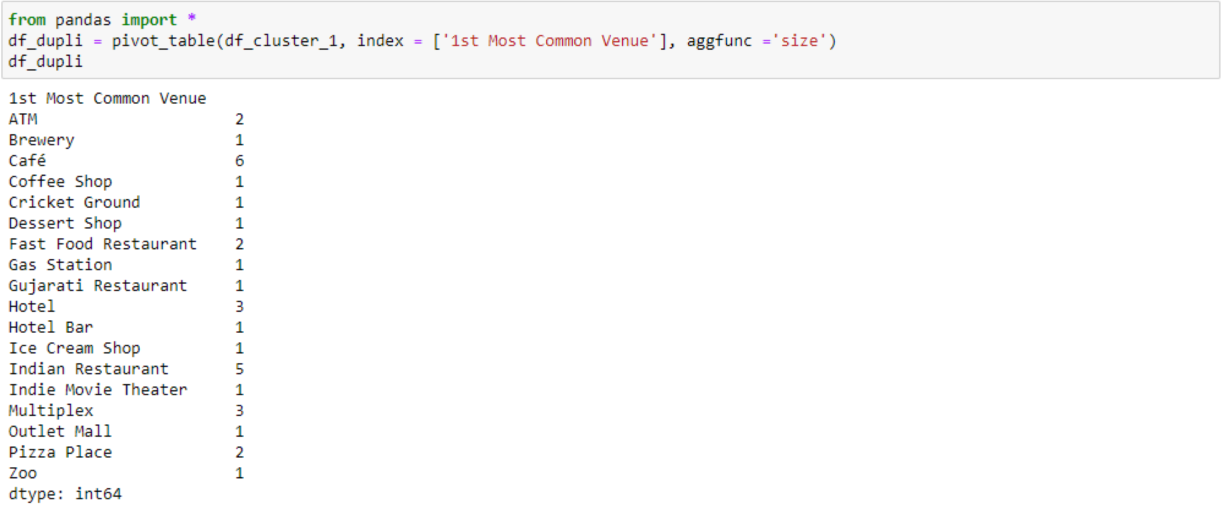


```
# create final map


for lat, log in zip(latitude, longitude):
    map_newyork_sushi = folium.Map(location=[lat, log], zoom_start=10)
    addToMap(ahmedabad_merged, 'red', map_newyork_sushi)

map_newyork_sushi
```

Displaying Clusters in Folium Map.



Output



```
from pandas import *
df_dupli = pivot_table(df_cluster_1, index = ['1st Most Common Venue'], aggfunc ='size')
df_dupli
```

```
1st Most Common Venue
ATM                        2
Brewery                    1
Café                       6
Coffee Shop                1
Cricket Ground             1
Dessert Shop               1
Fast Food Restaurant       2
Gas Station                1
Gujarati Restaurant        1
Hotel                      3
Hotel Bar                  1
Ice Cream Shop             1
Indian Restaurant          5
Indie Movie Theater        1
Multiplex                  3
Outlet Mall                1
Pizza Place                2
Zoo                        1
dtype: int64
```

Checking in each cluster which venue is best in which cluster.

Name: Sagar Tarasariya

Conclusion:

Q: which places are have best restaurant in Ahmedabad?
A: Sola, Ashram road, Ellisbridge, C G Road, stellite, Gulbai tekra and Hansoi are the best places which have **4+** ratings restaurant.

Q: Which Restaurant have most outlets in Ahmedabad city?
A: Domino's Pizza have most outlets in Ahmedabad

Q: Top Quick Bites restaurant in Ahmedabad?
A: La Pinoz Pizza is no. 1 Quick bite restaurant in Ahmedabad

Q: Top Casual Dining Restaurants in Ahmedabad?
A: Yanki Sizzlers,Barbeque Nation etc. are the best Dining Restaurant in Ahmedabad

Q:Which place are suitable for edible person in Ahmedabad city?
A: Bodakdev and Navrangpura are the best place for edible person

Q: What are the best chinese restaurant in Bodakdev?
A: Ph Se Food is the best Chinese Restaurant in Bodakdev

Q: How many Restaurant Accepting Digital Payment in Ahmedabad?
A: 63.4% restaurant are accepting digital payment but 36.6% are not accepting Digital payment

Cluster 1:

It is most recommended for Cafe.

Cluster 2:

It is most recommended for Indian Restaurant.

Cluster 3:

It is most recommended for Coffee Shope and Pizza place.

Cluster 4:

It is most recommended for Food Truck and hotel.

In this Project I mostly learned about Foursquare API and Data Analytics Skills.