# Bangalore Food Preference by Neighborhood

**By: Manas Sinha**

**In this Jupyter Notebook we'll try to find which areas in Bangalore prefer which type of food and wh**

## 1. Introduction

Bangalore is the third most populous city in India, and is home to many technology companies. Acco
Bangalore's population consists of migrants which makes the city's population quite diverse.
Undoubtedly when there is such a large quantity of migrants, there will be a lot of diversity in the fooc
project is to categorically segment the neighborhoods of Bangalore into major clusters and examine
examine the neighborhood cluster's food habits and taste. Further examination might reveal if food h
neighborhood.
This project will help to understand the diversity of a neighborhood by leveraging venue data from Fo
clustering' machine learning algorithm. Exploratory Data Analysis (EDA) will help to discover further a
neighborhood. Stakeholders would be the one who are interested to use this quantifiable analysis to
cultures and cuisines over one of the the most diverse cities in India - Bangalore". This project can al
willing to open his or her restaurant. Or by a government authority to examine and study their city's cu

## 2. Data

The following data sources are used to implement our project:

1. Bangalore Neighborhoods : Which provides us with geocoded data for the city of Bangalore. Th
2. Foursquare API:

Link: https://developer.foursquare.com/docs
Description: Foursquare API, a location data provider, will be used to make RESTful API calls to retrie
neighborhoods. This is the link to Foursquare Venue Category Hierarchy. Venues retrieved from all th
into "Arts & Entertainment", "College & University", "Event", "Food", "Nightlife Spot", "Outdoors & Recrea
follows:
'categories': [{'id': '4bf58dd8d48988d110941735',
'name': 'Italian Restaurant',
'pluralName': 'Italian Restaurants',
'shortName': 'Italian',
'icon': {'prefix': 'https://ss3.4sqi.net/img/categories_v2/food/italian_',
'suffix': '.png'},
'primary': True}],

'verified': False,
'stats': {'tipCount': 17},
'url': 'http://eccorestaurantny.com',
'price': {'tier': 4, 'message': 'Very Expensive', 'currency'

## 3. Methodology

### Importing Bangalore neighborhood dataset found on Kaggle

For simplicity reasons we have downloaded the .csv file beforehand instead of using the Kaggle API

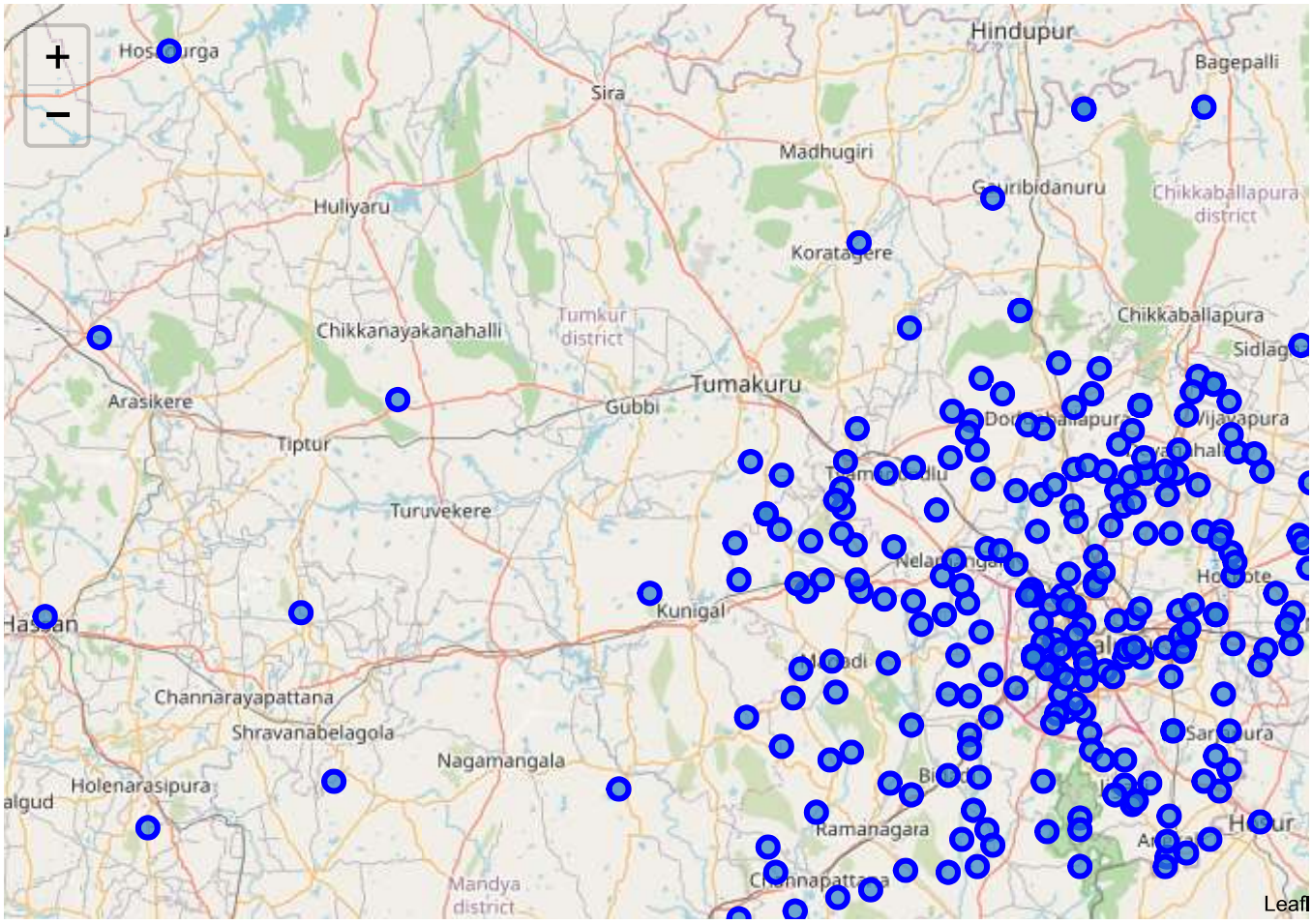
```
#keeping only the necessary columns
blr_df = blr_df[['Neighborhood','Latitude','Longitude']]
blr_df.head()
```

| | Neighborhood | Latitude | Longitude |
|---|---|---|---|
| 0 | Agram | 45.813177 | 15.977048 |
| 1 | Amruthahalli | 13.066513 | 77.596624 |
| 2 | Attur | 11.663711 | 78.533551 |
| 3 | Banaswadi | 13.014162 | 77.651854 |
| 4 | Bellandur | 58.235358 | 26.683116 |

### Map of Neighborhoodds

```
blr_lat = 12.97
blr_lon = 77.58
map_blr = folium.Map(location=[blr_lat,blr_lon], zoom_start=9)

# add markers to map
for lat, lng, neighborhood in zip(blr_df['Latitude'], blr_df['Longitude'], blr_df['Neighborhc
    label = '{}'.format(neighborhood)
    label = folium.Popup(label, parse_html=True)
    folium.CircleMarker(
        [lat, lng],
        radius=5,
        popup=label,
        color='blue',
        fill=True,
        fill_color='#3186cc',
        fill_opacity=0.7,
        parse_html=False).add_to(map_blr)
```

```
map_blr
```




## Understanding The Foursquare API better

```
results['response']['venues']
```

```
[{'categories': [{'icon': {'prefix': 'https://ss3.4sqi.net/img/categories_v2/food/defau
    'suffix': '.png'},
   'id': '54135bf5e4b08f3d2429dfde',
   'name': 'South Indian Restaurant',
   'pluralName': 'South Indian Restaurants',
   'primary': True,
   'shortName': 'South Indian'}],
  'hasPerk': False,
  'id': '4e1ace8d183880768f580e21',
  'location': {'address': 'Near Civil Court',
   'cc': 'IN',
   'city': 'Bangalore',
   'country': 'India',
   'distance': 485,
   'formattedAddress': ['Near Civil Court', 'Bangalore', 'Karnātaka', 'India'],
   'labeledLatLngs': [{'label': 'display',
     'lat': 12.972336965593202,
     'lng': 77.58377805364935}],
   'lat': 12.972336965593202,
   'lng': 77.58377805364935,
   'state': 'Karnātaka'},
  'name': 'Kaveri Bhavan',
  'referralId': 'v-1587056319'}]
```


The name of the restaurant is **Kaveri Bhavan** and the category is **South Indian Restaurant**

## 4. Analysis and Machine Learning

```
# one hot encoding
blr_onehot = pd.get_dummies(blr_venues[['Venue Category']], prefix="", prefix_sep="")
blr_onehot.head()
```

| | Afghan Restaurant | American Restaurant | Andhra Restaurant | Asian Restaurant | Austrian Restaurant | BBQ Joint | Belgian Restaurant | Brewer |
|---|---|---|---|---|---|---|---|---|
| **0** | 0 | 1 | 0 | 0 | 0 | 0 | 0 | |
| **1** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| **2** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| **3** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |
| **4** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | |

```
# move neighborhood column to the first column
Neighborhood = blr_onehot['Neighborhood']

blr_onehot.drop(labels=['Neighborhood'], axis=1,inplace = True)
blr_onehot.insert(0, 'Neighborhood', Neighborhood)
```

```
blr_onehot.head()
```

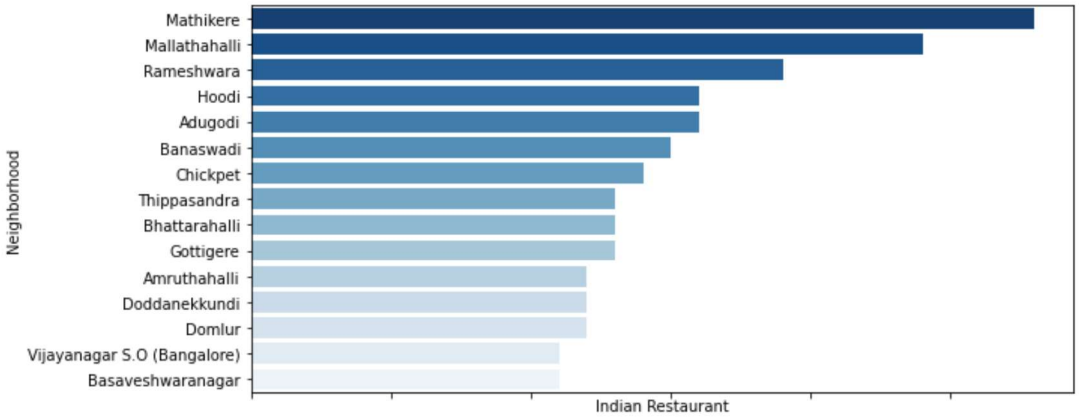
| | Neighborhood | Afghan Restaurant | American Restaurant | Andhra Restaurant | Asian Restaurant | Austrian Restaurant | BBQ Joint | B Rest |
|---|---|---|---|---|---|---|---|---|
| **0** | Agram | 0 | 1 | 0 | 0 | 0 | 0 | |
| **1** | Agram | 0 | 0 | 0 | 0 | 0 | 0 | |
| **2** | Agram | 0 | 0 | 0 | 0 | 0 | 0 | |
| **3** | Agram | 0 | 0 | 0 | 0 | 0 | 0 | |
| **4** | Agram | 0 | 0 | 0 | 0 | 0 | 0 | |

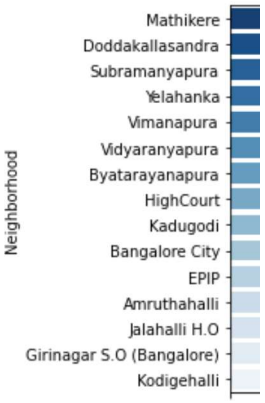
### Finding the top 15 food categories in Bangalore

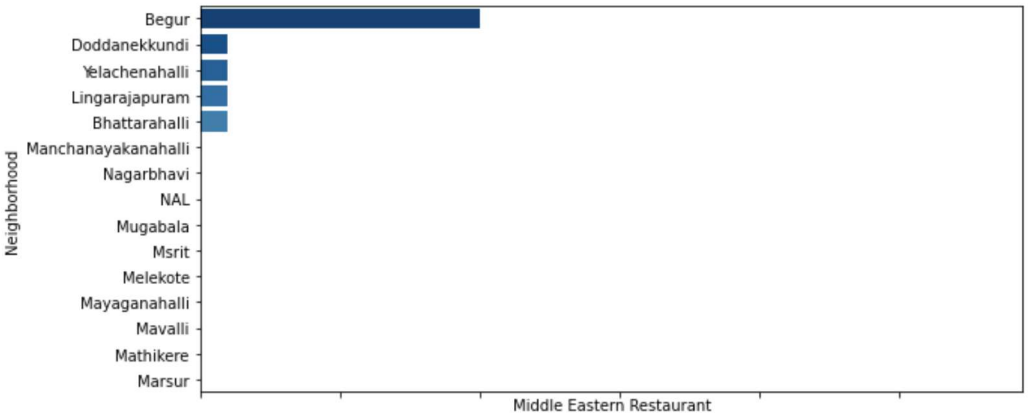
```
venue_counts_described = venue_counts.describe().transpose()

venue_top15 = venue_counts_described.sort_values('max', ascending=False)[0:15]
venue_top15
```

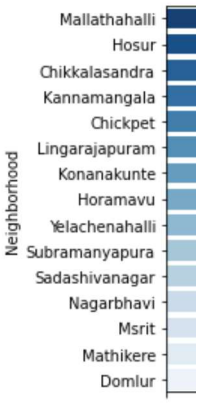
| | count | mean | std | min | 25% | 50% | 75% | max |
|---|---|---|---|---|---|---|---|---|
| **Indian Restaurant** | 100.0 | 5.54 | 5.618188 | 0.0 | 1.0 | 4.0 | 9.00 | 28.0 |
| **Fast Food Restaurant** | 100.0 | 1.53 | 1.839274 | 0.0 | 0.0 | 1.0 | 2.00 | 10.0 |
| **Middle Eastern Restaurant** | 100.0 | 0.14 | 1.015237 | 0.0 | 0.0 | 0.0 | 0.00 | 10.0 |
| **South Indian Restaurant** | 100.0 | 0.69 | 1.125418 | 0.0 | 0.0 | 0.0 | 1.00 | 6.0 |
| **Pizza Place** | 100.0 | 0.72 | 1.073840 | 0.0 | 0.0 | 0.0 | 1.00 | 6.0 |
| **Chinese Restaurant** | 100.0 | 0.57 | 0.890522 | 0.0 | 0.0 | 0.0 | 1.00 | 5.0 |
| **Italian Restaurant** | 100.0 | 0.26 | 0.733333 | 0.0 | 0.0 | 0.0 | 0.00 | 4.0 |
| **Vegetarian / Vegan Restaurant** | 100.0 | 0.66 | 1.036895 | 0.0 | 0.0 | 0.0 | 1.00 | 4.0 |
| **North Indian Restaurant** | 100.0 | 0.20 | 0.635642 | 0.0 | 0.0 | 0.0 | 0.00 | 4.0 |
| **Kerala Restaurant** | 100.0 | 0.13 | 0.505625 | 0.0 | 0.0 | 0.0 | 0.00 | 4.0 |
| **Doner Restaurant** | 100.0 | 0.04 | 0.400000 | 0.0 | 0.0 | 0.0 | 0.00 | 4.0 |
| **Japanese Restaurant** | 100.0 | 0.07 | 0.355477 | 0.0 | 0.0 | 0.0 | 0.00 | 3.0 |
| **Karnataka Restaurant** | 100.0 | 0.32 | 0.617587 | 0.0 | 0.0 | 0.0 | 0.25 | 3.0 |
| **Multicuisine Indian Restaurant** | 100.0 | 0.14 | 0.449916 | 0.0 | 0.0 | 0.0 | 0.00 | 3.0 |
| **Mediterranean Restaurant** | 100.0 | 0.06 | 0.342893 | 0.0 | 0.0 | 0.0 | 0.00 | 3.0 |

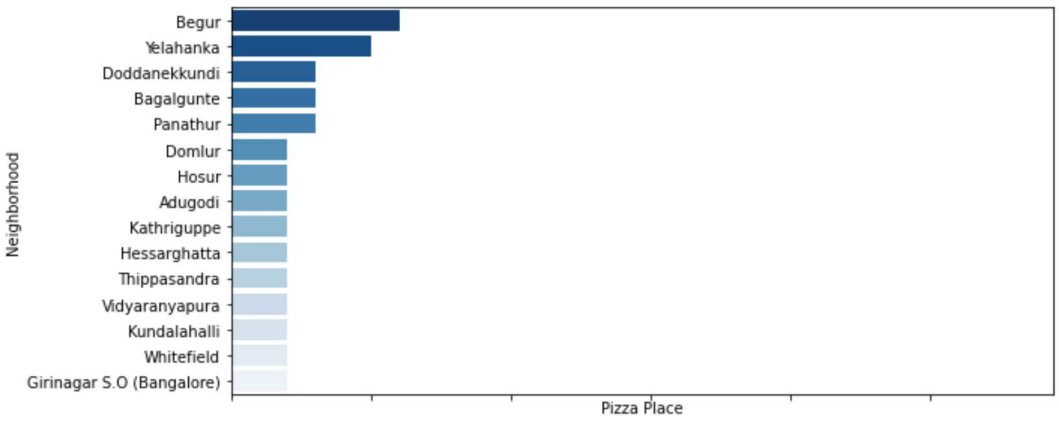
```
import seaborn as sns
import matplotlib.pyplot as plt

fig, axes =plt.subplots(5, 2, figsize=(20,20), sharex=True)
axes = axes.flatten()

for ax, category in zip(axes, venue_top15_list):
    data = venue_counts[[category]].sort_values([category], ascending=False)[0:15]
    pal = sns.color_palette("Blues", len(data))
    sns.barplot(x=category, y=data.index, data=data, ax=ax, palette=np.array(pal[::-1]))

plt.tight_layout()
plt.show();
```

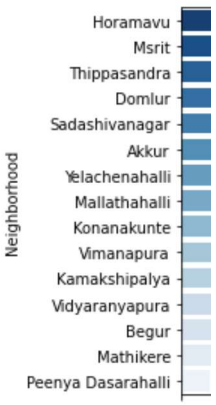

Mathikere
Mallathahalli
Rameshwara
Hoodi
Adugodi
Banaswadi
Chickpet
Thippasandra
Bhattarahalli
Gottigere
Amruthahalli
Doddanekkundi
Domlur
Vijayanagar S.O (Bangalore)
Basaveshwaranagar
Neighborhood
Indian Restaurant

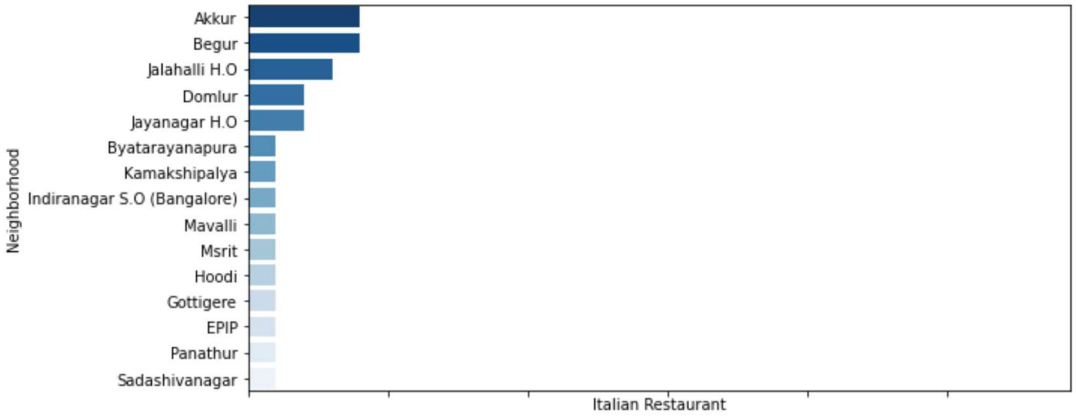

Mathikere
Doddakallasandra
Subramanyapura
Yelahanka
Vimanapura
Vidyaranyapura
Byatarayanapura
HighCourt
Kadugodi
Bangalore City
EPIP
Amruthahalli
Jalahalli H.O
Girinagar S.O (Bangalore)
Kodigehalli
Neighborhood

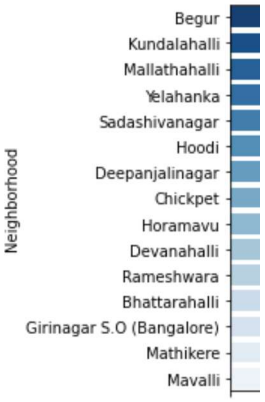

Begur
Doddanekkundi
Yelachenahalli
Lingarajapuram
Bhattarahalli
Manchanayakanahalli
Nagarbhavi
NAL
Mugabala
Msrit
Melekote
Mayaganahalli
Mavalli
Mathikere
Marsur
Neighborhood
Middle Eastern Restaurant


Mallathahalli
Hosur
Chikkalasandra
Kannamangala
Chickpet
Lingarajapuram
Konanakunte
Horamavu
Yelachenahalli
Subramanyapura
Sadashivanagar
Nagarbhavi
Msrit
Mathikere
Domlur
Neighborhood

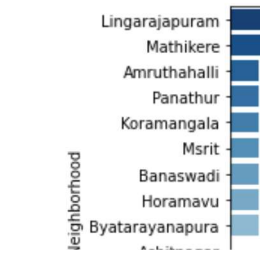

Begur
Yelahanka
Doddanekkundi
Bagalgunte
Panathur
Domlur
Hosur
Adugodi
Kathriguppe
Hessarghatta
Thippasandra
Vidyaranyapura
Kundalahalli
Whitefield
Girinagar S.O (Bangalore)
Neighborhood
Pizza Place


Horamavu
Msrit
Thippasandra
Domlur
Sadashivanagar
Akkur
Yelachenahalli
Mallathahalli
Konanakunte
Vimanapura
Kamakshipalya
Vidyaranyapura
Begur
Mathikere
Peenya Dasarahalli
Neighborhood


Akkur
Begur
Jalahalli H.O
Domlur
Jayanagar H.O
Byatarayanapura
Kamakshipalya
Indiranagar S.O (Bangalore)
Mavalli
Msrit
Hoodi
Gottigere
EPIP
Panathur
Sadashivanagar
Neighborhood
Italian Restaurant


Begur
Kundalahalli
Mallathahalli
Yelahanka
Sadashivanagar
Hoodi
Deepanjalinagar
Chickpet
Horamavu
Devanahalli
Rameshwara
Bhattarahalli
Girinagar S.O (Bangalore)
Mathikere
Mavalli
Neighborhood


Mallathahalli
Kannamangala
Mathikere
Hoodi
Banaswadi
Devanahalli
Yelachenahalli
Nagarbhavi
Mayaganahalli
leighborhood


Lingarajapuram
Mathikere
Amruthahalli
Panathur
Koramangala
Msrit
Banaswadi
Horamavu
Byatarayanapura
leighborhood


## Grouping rows by neighborhood and by taking the mean of the frequency of occurrence of each cat

```
blr_grouped = blr_onehot.groupby('Neighborhood').mean().reset_index()
blr_grouped.head()
```

| | Neighborhood | Afghan Restaurant | American Restaurant | Andhra Restaurant | Asian Restaurant | Austrian Restaurant | BBQ Joint | B Rest |
|---|---|---|---|---|---|---|---|---|
| 0 | Achitnagar | 0.0 | 0.000000 | 0.000000 | 0.200000 | 0.0 | 0.000 | |
| 1 | Adugodi | 0.0 | 0.000000 | 0.000000 | 0.000000 | 0.0 | 0.000 | |
| 2 | Agram | 0.0 | 0.153846 | 0.000000 | 0.076923 | 0.0 | 0.000 | |
| 3 | Akkur | 0.0 | 0.000000 | 0.000000 | 0.125000 | 0.0 | 0.125 | |
| 4 | Amruthahalli | 0.0 | 0.000000 | 0.068966 | 0.034483 | 0.0 | 0.000 | |

```
#new size of dataframe blr_grouped
blr_grouped.shape
```

```
(100, 70)
```

## Creating new dataframe to display the top 5 venues for each neighborhood

```
for ind in np.arange(blr_grouped.shape[0]):
    neighborhoods_venues_sorted.iloc[ind, 1:] = return_most_common_venues(blr_grouped.iloc[ir

neighborhoods_venues_sorted.head()
```

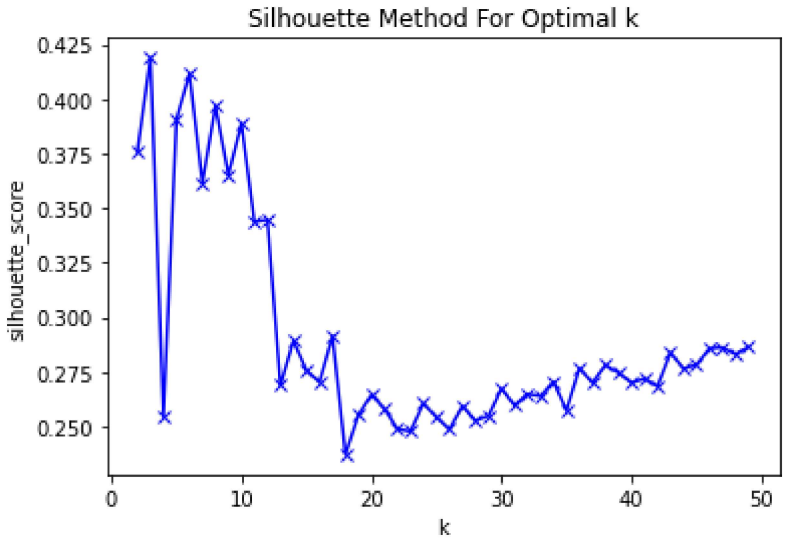
| | Neighborhood | 1st Most Common Venue | 2nd Most Common Venue | 3rd Most Common Venue | 4 |
|---|---|---|---|---|---|
| 0 | Achitnagar | Fast Food Restaurant | Indian Restaurant | Asian Restaurant | |
| 1 | Adugodi | Indian Restaurant | Pizza Place | Lebanese Restaurant | |
| 2 | Agram | Mediterranean Restaurant | Pizza Place | American Restaurant | |
| 3 | Akkur | Italian Restaurant | Turkish Restaurant | Asian Restaurant | |
| 4 | Amruthahalli | Indian Restaurant | Fast Food Restaurant | Andhra Restaurant | |

## Clustering Neighborhoods

```
blr_grouped_clustering = blr_grouped.drop('Neighborhood', 1)
```

**Determine the optimal number of clusters for k-means clustering using Silhouette Method**

```
plt.plot(K_sil, sil, 'bx-')
plt.xlabel('k')
plt.ylabel('silhouette_score')
plt.title('Silhouette Method For Optimal k')
plt.show()
```




**At k=2 and k=6 there is peak. Let's use k=6**

Let's create a new dataframe that includes the cluster as well as the top 5 venues for each neighborh

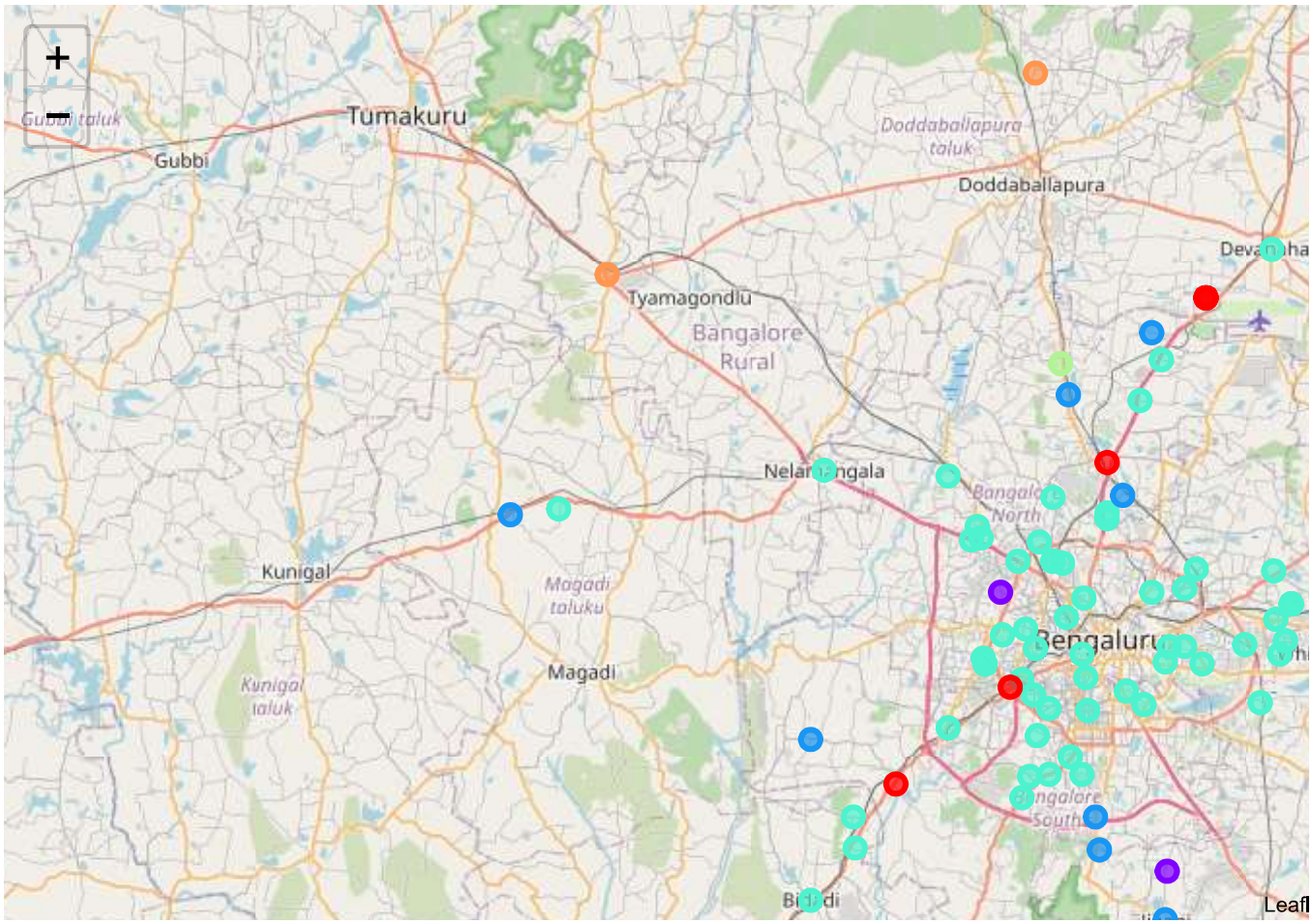
```
# merge neighborhoods_venues_sorted with blr_data to add latitude/longitude for each neighbor
blr_merged = neighborhoods_venues_sorted.join(blr_df.set_index('Neighborhood'), on='Neighborh
blr_merged.head()
```

| | Cluster Labels | Neighborhood | 1st Most Common Venue | 2nd Most Common Venue | 3rd Most Common Venue | 4th |
|---|---|---|---|---|---|---|

## Creating a Map for Visualization

Restaurant

```
map_clusters = folium.Map(location=[12.97, 77.58], zoom_start=10)

# set color scheme for the clusters
colors_array = cm.rainbow(np.linspace(0, 1, kclusters))
rainbow = [colors.rgb2hex(i) for i in colors_array]

# add markers to the map
markers_colors = []
for lat, lon, poi, cluster in zip(blr_merged['Latitude'], blr_merged['Longitude'], blr_merged
    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
```




# 5. Results

### Cluster 0

```
cluster_0 = blr_merged.loc[blr_merged['Cluster Labels'] == 0, blr_merged.columns[1:12]]
cluster_0.head(5)
```

| | Neighborhood | 1st Most Common Venue | 2nd Most Common Venue | 3rd Most Common Venue | 4th Most Common Venue | 5th Most Common Venue | Latit |
|---|---|---|---|---|---|---|---|
| 2 | Agram | Mediterranean Restaurant | Pizza Place | American Restaurant | Fast Food Restaurant | Japanese Restaurant | 45.813 |
| 3 | Akkur | Italian Restaurant | Turkish Restaurant | Asian Restaurant | BBQ Joint | Belgian Restaurant | 50.733 |
| 3 | Akkur | Italian Restaurant | Turkish Restaurant | Asian Restaurant | BBQ Joint | Belgian Restaurant | 50.733 |

## Cluster 1

```
cluster_1 = blr_merged.loc[blr_merged['Cluster Labels'] == 1, blr_merged.columns[1:12]]
cluster_1.head(5)
```

| | Neighborhood | 1st Most Common Venue | 2nd Most Common Venue | 3rd Most Common Venue | 4th M |
|---|---|---|---|---|---|
| 16 | Byagadadenahalli | Pizza Place | Halal Restaurant | Falafel Restaurant | Fast Fooc |
| 39 | Hulimangala | Pizza Place | Halal Restaurant | Falafel Restaurant | Fast Fooc |
| 60 | Laggere | Pizza Place | Halal Restaurant | Falafel Restaurant | Fast Fooc |
| 73 | NAL | Pizza Place | Halal Restaurant | Falafel Restaurant | Fast Fooc |
| 96 | Whitefield | Pizza Place | Chinese Restaurant | Dumpling Restaurant | Falafe |

## Cluster 2

```
cluster_2 = blr_merged.loc[blr_merged['Cluster Labels'] == 2, blr_merged.columns[1:12]]
cluster_2.head(5)
```

| | 1st Most | 2nd Most | 3rd Most | 4th Most | 5th Most |
|---|---|---|---|---|---|

## Cluster 3

```
cluster_3 = blr_merged.loc[blr_merged['Cluster Labels'] == 3, blr_merged.columns[1:12]]
cluster_3.head(5)
```

| | Neighborhood | 1st Most Common Venue | 2nd Most Common Venue | 3rd Most Common Venue | 4th Most Common Venue | 5th Most Common Venue | Latit |
|---|---|---|---|---|---|---|---|
| **0** | Achitnagar | Fast Food Restaurant | Indian Restaurant | Asian Restaurant | Yemeni Restaurant | Halal Restaurant | 13.091 |
| **1** | Adugodi | Indian Restaurant | Pizza Place | Lebanese Restaurant | Rajasthani Restaurant | Fast Food Restaurant | 12.942 |
| **4** | Amruthahalli | Indian Restaurant | Fast Food Restaurant | Andhra Restaurant | Pizza Place | Karnataka Restaurant | 13.066 |

## Cluster 4

```
cluster_4 = blr_merged.loc[blr_merged['Cluster Labels'] == 4, blr_merged.columns[1:12]]
cluster_4.head(5)
```

| | Neighborhood | 1st Most Common Venue | 2nd Most Common Venue | 3rd Most Common Venue | 4th Most Common Venue | 5th Most Common Venue | Latit |
|---|---|---|---|---|---|---|---|
| **13** | Bellandur | Fast Food Restaurant | Yemeni Restaurant | Halal Restaurant | Falafel Restaurant | French Restaurant | 58.235 |
| **34** | HighCourt | Fast Food Restaurant | Pizza Place | Halal Restaurant | Falafel Restaurant | French Restaurant | 53.783 |
| **82** | Rajanakunte | Fast Food Restaurant | Yemeni Restaurant | Halal Restaurant | Falafel Restaurant | French Restaurant | 13.169 |

## Cluster 5

```
cluster_5 = blr_merged.loc[blr_merged['Cluster Labels'] == 5, blr_merged.columns[1:12]]
cluster_5.head(5)
```

| | | 1st Most | 2nd Most | 3rd Most | 4th Most | 5th Most | |
|---|---|---|---|---|---|---|---|
| **30** | [illegible] | [illegible] | Karnataka Restaurant | Yemeni Restaurant | Halal Restaurant | Falafel Restaurant | [illegible] |
| **31** | Hadonahalli | Vegetarian / Vegan Restaurant | Yemeni Restaurant | Halal Restaurant | Falafel Restaurant | Fast Food Restaurant | 13.369 |

## Thank You