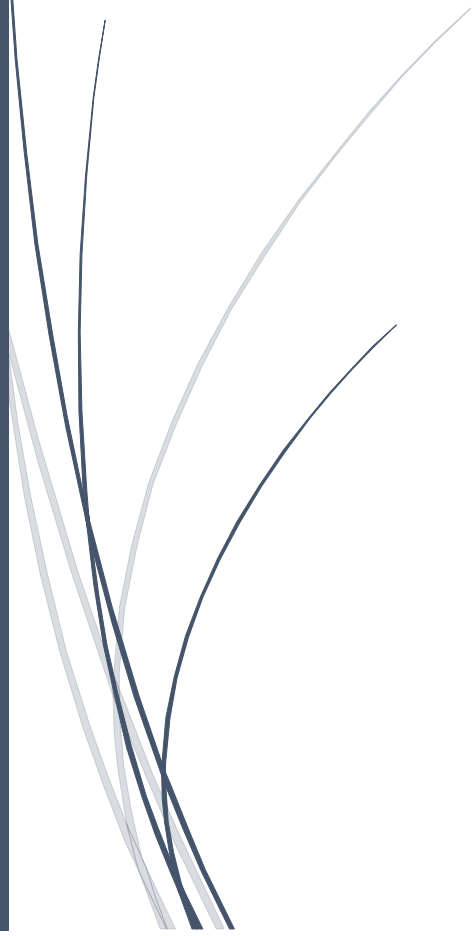
7/12/2020

Indian Cuisine Recommendation System - For Massachusetts, USA

Capstone Project - The Battle of Neighbourhoods



Thilak Rethinekumar

Introduction:

Massachusetts with its diverse culture comes with diverse cuisines like Indian, Chinese, Mediterranean, Mexican etc.
We wanted to do some detailed exploratory data analysis of all the neighbourhoods in Massachusetts and find answers to below questions.

Problem statement:

Find answers to following questions:
1. What are best location in Massachusetts for Indian Cuisine?
2. Which areas have most Indian Restaurants ?
3. Which all areas lack Indian Restaurants but potential to improve?
4. Which is the biggest chain of Indian restaurants in Massachusetts?

Target Audience:

Anyone who wish to find an Indian Restaurant for a weekend lunch or buffet

Why it is important?:

Massachusetts is a metropoliton city with many immigrants from various countries who have different food tastes and styles. This system can also be used to find not only Indian Restaurant but also other types of restaurants.

Data Sourcing:

For this project I would need list of Neighborhoods in Massachusetts along with their latitude and longitude.

I would take above mentioned data from:

https://public.opendatasoft.com/explore/dataset/us-zip-code-latitude-and-longitude/export/?refine.state=MA

It would consist of Zip code, Neighborhood name, Latitude and Longitude

Example:

Zip City Latitude Longitude 1720 Acton 42.477142 -71.4422

I would need to find the Indian restaurants in each neighbourhood of Massachusetts based on above data. For that I would be using Foursquare API.

By using this API I will get all the venues in each neighbourhood. I can then filter these venues to get only Indian restaurants.

Methodology:

1. We begin by collecting the Massachusetts data from the following link https://public.opendatasoft.com/explore/dataset/us-zip-code-latitude-and-longitude/export/?refine.state=MA

```
Geospatial_Coordinates = pd.read_csv('us-zip-code-latitude-and-longitude.csv')
Geospatial_Coordinates.columns=["Zip","City","Latitude","Longitude"]
Geospatial_Coordinates['Zip']=Geospatial_Coordinates['Zip'].apply(lambda x: '{0:0>5}'.format(x))
Geospatial_Coordinates
```

2. Get Four square API credentials

3. Define a function to get venues using Four square API

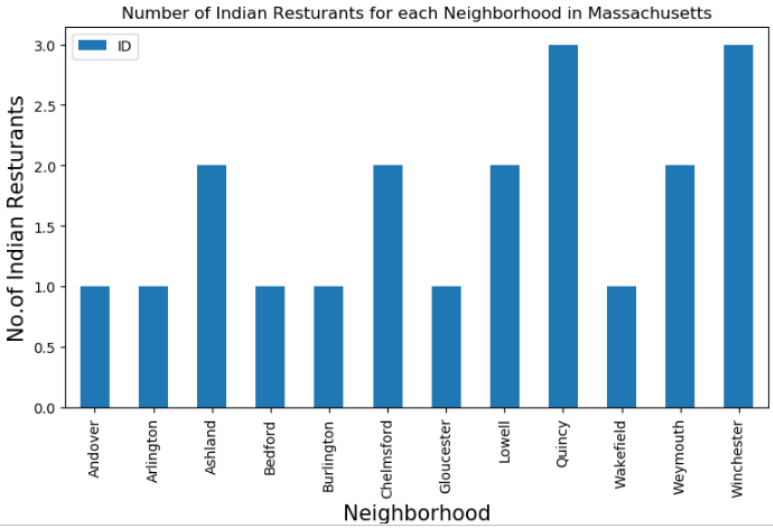
4. For each and every Neighborhood, need to call the above defined function to get the venues and then filter only Indian Restaurants and assign them all to a Pandas data frame

```
6]: # prepare neighborhood list that contains indian resturants
column_names=['Neighborhood', 'ID','Name']
indian_rest_bo=pd.DataFrame(columns=column_names)
count=1
for row in Geospatial_Coordinates.values.tolist():
    Zip, City, Latitude, Longitude=row
    venues = get_venues(Latitude,Longitude)
    indian_resturants=venues[venues['Category']=='Indian Restaurant']
    print('(',count,'/',len(Geospatial_Coordinates),')','Indian Resturants in '+City+':'+str(len(indian_resturants)))
    for resturant_detail in indian_resturants.values.tolist():
        id, name , category=resturant_detail
        indian_rest_bo = indian_rest_ny.append({'Neighborhood': City,
                                                'ID': id,
                                                'Name' : name
                                               }, ignore_index=True)
    count+=1
    indian_rest_bo
```

5. Create a Bar diagram to show which Neighborhood has got most number of Indian Restaurants in Massachusetts

Bar diagram to show which Neighborhood has got most number of Indian Restaurants in Massachusetts

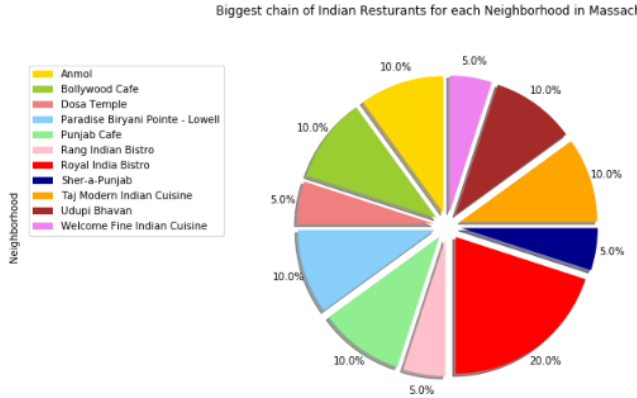
```
plt.figure(figsize=(9,5), dpi = 100)
# title
plt.title('Number of Indian Resturants for each Neighborhood in Massachusetts')
#On x-axis
plt.xlabel('Neighborhood', fontsize = 15)
#On y-axis
plt.ylabel('No.of Indian Resturants', fontsize=15)
#giving a bar plot
indian_rest_bo.groupby('Neighborhood')['ID'].count().plot(kind='bar')
#Legend
plt.legend()
#displays the plot
plt.show()
```



6. Create a Pie chart to show the biggest chain of Indian Restaurants for each Neighborhood in Massachusetts

Pie chart to show the biggest chain of Indian Restaurants for each Neighborhood in Massachusetts

```
# group countries by continents and apply sum() function
indian_rest_bos = indian_rest_bo.groupby('Name', axis=0).count()
indian_rest_bos

colors_list = ['gold', 'yellowgreen', 'lightcoral', 'lightskyblue', 'lightgreen', 'pink','red','darkblue','orange','brown','violet']
explode_list = [0.1, 0.1, 0.1, 0.1, 0.1, 0.1,0.1,0.1,0.1,0.1,0.1] # ratio for each continent with which to offset each wedge.

indian_rest_bos['Neighborhood'].plot(kind='pie',
                            figsize=(15, 6),
                            autopct='%1.1f%%',
                            startangle=90,
                            shadow=True,
                            labels=None,         # turn off labels on pie chart
                            pctdistance=1.12,    # the ratio between the center of each pie slice and the start of the text generated by autopct
                            colors=colors_list,  # add custom colors
                            explode=explode_list # 'explode' lowest 3 continents
                            )

plt.axis('equal')

plt.legend(labels=indian_rest_bos.index, loc='upper left')

plt.title('Biggest chain of Indian Resturants for each Neighborhood in Massachusetts', y=1.12)

plt.show()
```



7. Create a function to get more details like Ratings, Likes and Tips for each and every restaurant using Foursquare API
8. Create a pandas data frame with Likes, Ratings and Tips for each and every Indian Restaurant in Massachusetts

Creating a pandas data frame with Likes, Ratings and Tips for each and every Indian Restaurant in Massachusetts

```
# prepare neighborhood list that contains indian resturants
column_names=['Neighborhood', 'ID','Name','Likes','Rating','Tips']
indian_rest_stats_bo=pd.DataFrame(columns=column_names)
count=1


for row in indian_rest_bo.values.tolist():
    Neighborhood,ID,Name=row
    try:
        venue_details=get_venue_details(ID)
        print(venue_details)
        id,name,likes,rating,tips=venue_details.values.tolist()[0]
    except IndexError:
        print('No data available for id=',ID)
        # we will assign 0 value for these resturants as they may have been
        #recently opened or details does not exist in FourSquare Database
        id,name,likes,rating,tips=[0]*5
    print('(',count,'/',len(indian_rest_bo),')','processed')
    indian_rest_stats_bo = indian_rest_stats_bo.append({
                                            'Neighborhood': Neighborhood,
                                            'ID': id,
                                            'Name' : name,
                                            'Likes' : likes,
                                            'Rating' : rating,
                                            'Tips' : tips
                                           }, ignore_index=True)
    count+=1
```

9. Find Most liked restaurant in Massachusetts

Most liked restaurant in Massachusetts

```
[17]: indian_rest_stats_bo.iloc[indian_rest_stats_bo['Likes'].idxmax()]
```

```
[17]: Neighborhood                  Arlington
      ID              4d274b6f55a8b60cc384c2c0
      Name                  Royal India Bistro
      Likes                                 64
      Rating                               8.7
      Tips                                  22
      Name: 1, dtype: object
```

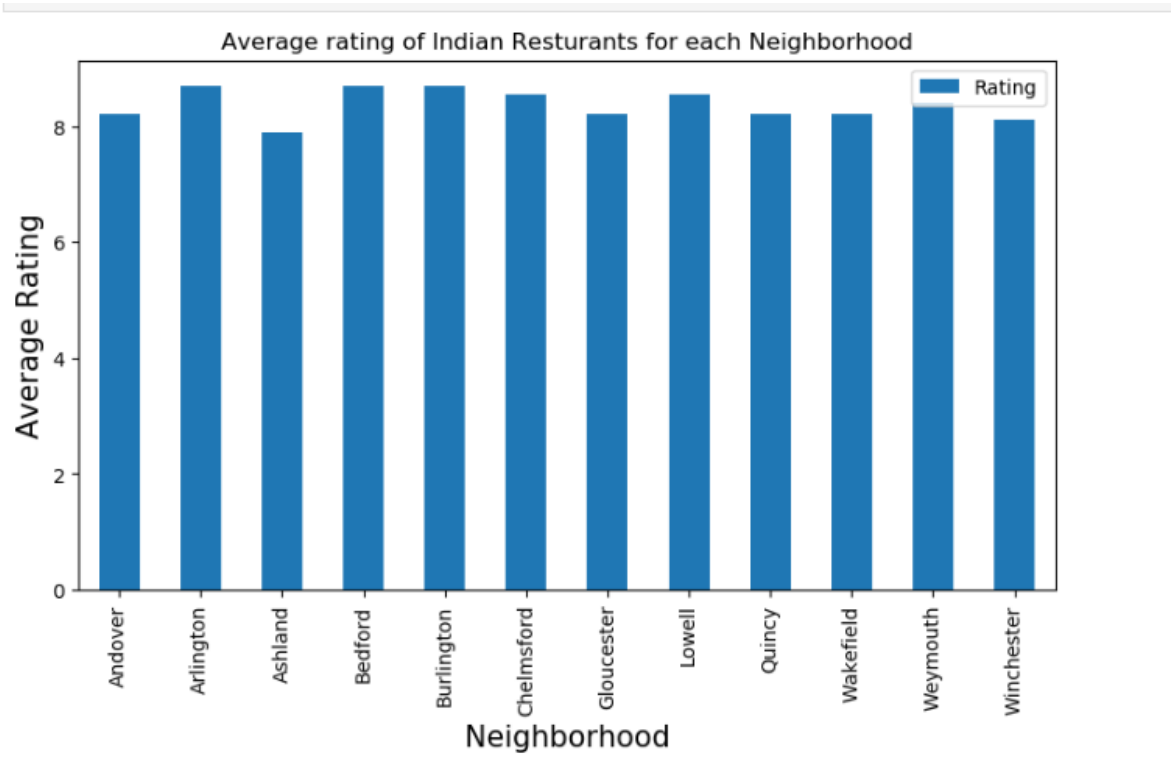
10. Find Highly rated restaurant in Massachusetts

Highly rated restaurant in Massachusetts

```
[19]: indian_rest_stats_bo.iloc[indian_rest_stats_bo['Rating'].idxmax()]
```

```
[19]: Neighborhood                  Chelmsford
      ID              4a82fbdef964a520bcf91fe3
      Name                        Udupi Bhavan
      Likes                                 33
      Rating                               8.9
      Tips                                  21
      Name: 6, dtype: object
```

11. Find top locations in Massachusetts for Indian Restaurants

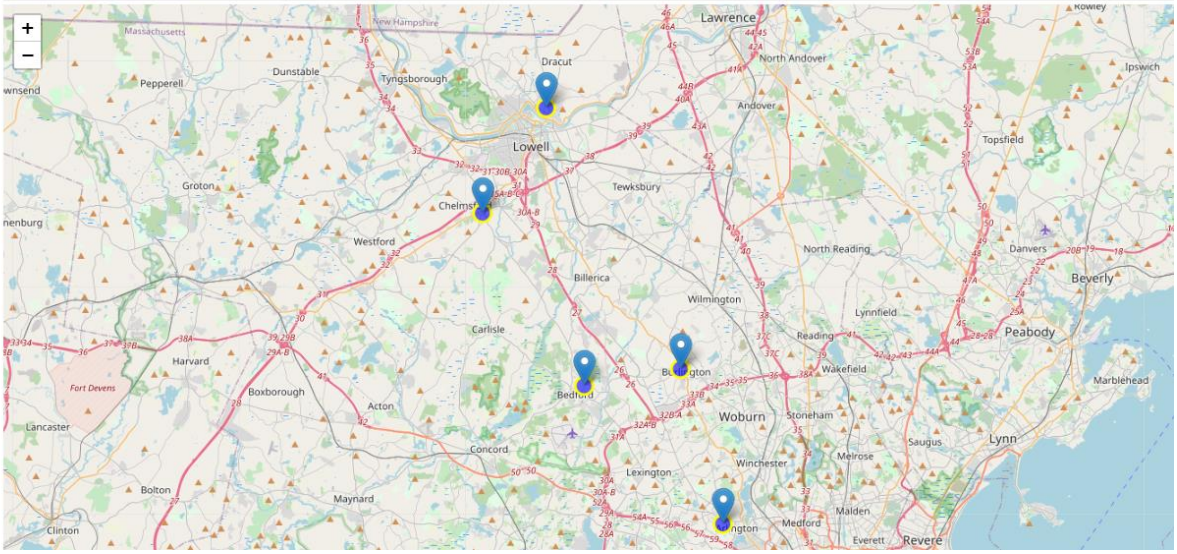


12. Map showing top 5 areas in Massachusetts for good Indian Restaurants

Maps showing top 5 areas in Massachusetts for good Indian Restaurants

```
[71]: # add pop-up text to each marker on the map
for lat, lng, label in bos_neighborhood_stats[['Latitude','Longitude','Label']].values:
    folium.Marker([lat, lng], popup=label).add_to(bos_map)
# add incidents to map
bos_map.add_child(incidents)
```

[71]:



Discussion & Conclusion:

1) What are best location in Massachusetts for Indian Cuisine?

Arlington, Bedford, Burlington, Chelmsford, Lowell. Refer above map as well as the bar diagram.

2) Which areas have most Indian Restaurants ?

Quincy and Winchester has most Indian Restaurants. Refer the bar diagram.

3) Which all areas lack Indian Restaurants but potential to improve?

Andover, Arlington, Bedford, Burlington, Gloucestar, Wakefield are the places which lack Indian restaurants.

However, Arlington, Bedford, Burlington already has some best Indian Restaurants.

So, Andover, Gloucestar and Wakefield has got the potential to improve. If one want to start an Indian Restaurant, then it would be the best place.

4) Which is the biggest chain of Indian restaurants in Massachusetts?

Royal Indian Bistro is the biggest chain of Indian restaurants in Massachusetts. Refer the pie chart as well.

Note:

There is always room for improvement and hence the above solution provided can also be improved for best results depending upon the data we have.

We can also develop prediction model which can accept a town as input parameter and it can recommend best Indian Restaurant in that town.