IBM Data Science Peer Graded Assignment

The Battle of Neighborhoods

Week 4 : Part a)

Title : Restaurant recommender system in Gurgaon

Gurgaon, officially named Gurugram, is a city located in the northern Indian state of Haryana. It is situated near the Delhi-Haryana border, about 30 kilometres southwest of the national capital New Delhi and 268 km (167 mi) south of Chandigarh, the state capital. It is one of the major satellite cities of Delhi and is part of the National Capital Region of India. Gurgaon had a population of 876,900. Gurgaon has become a leading financial and industrial hub with the third-highest per capita income in India.The diversity of the cuisine available is reflective of the social and economic diversity of Gurgaon. Roadside vendors, tea stalls, South Indian, North Indian, Muslim food, Chinese and Western fast food are all very popular in the city. The Chinese food and the Thai food served in most of the restaurants are can be customised to cater to the tastes of the Indian population.Also,ithas become a big center for serving freshly brewed Beer and other liquors.

In [ ]:

Problem description :

I travel and keep changing places very frequently. This is very hectic and plus I get to experience very different types of environment, of which i donot have much knowledge about. In such situation, food can be an important factor for decided how you rate your trips and plus also recommmending it to the people. Food can also attract people around to world to try it out if it were to be the best. In such scenarios, we need to find the right place, at reasonable cost, to serve us the best possible way. So there are few questions that must be addresed, such as :

1.How many types of foods are available in the restaurant ? 2.which is the most nearest to me with good rating ? 3.How many "similar" restaurants are available near by me ? 4.Do the "similar" restaurants cost more ? if so, what speciality do that have ?

To address such question, XXYZ company's manager decides to allocate this project to me not just to find out solutions to the questions but also build a system that can help in recommending new places based on their rankings compared to the previously visited by me.

Expectations from this recommender system is to get answer for the questions, and in such a way that it uncovers all the perspective of managing recommendations. It is sighted to show :

1.What types of restaurants are present in a particular area ?

2.where are the similar restaurant present based on a preference to particular food ?

3.How do different restaurants rank with respect to my preferences ?

Target audience :

Target audiences for this project does not limit to a person who keeps travelling but everyone. People could simply decide to look for a similar restaurant all the time because they are addicted to a specific category of food. People who rarely use restaurants would prefer to have the most rated restaurants nearby them and all this could be easily handed by our recommender system. So target for this project is basically everyone who is exploring different places or similar places.

Week 4 : Part b) -

Title : Restaurant recommender system in Gurgaon

Data requirements :

To find a solution to the questions and build a recommender model, we need lots of data. Data can answer question which are unimaginable and non answerable by humans because humans do not have the tendency to analyse such large dataset and produce analtics to find a solutions.

Let's consider the base scenario :

Suppose I want to find a restaurant, then logically, I need 3 things :

1. Its geographical coordinates(latitide and longitude) to find our where exactly it is located.
2. Population of the neighborhood where the restaurant is located.
3. Average income of neighborhood to know how much is the restaurant worth.

Lets take a closer look at each of these :

To access location of a restaurant, its Latitude and Longitude is to be known so that we can point at its coordinates and create a map displaying all the restaurants with its labels respectively. Population of a neighborhood is very important factor in determining a restaurant's growth and amount of customers who turn up to eat. Logically, the more the population of a neighborhood, the more people will be interested to walk openly into a restaurant and less the population, less number of people frequently visit a restaurant. Also if more people visit, better the restaurant is rated because it is accessed by different people with different taste. Hence it is a very important factor. Income of a neighborhood is also very important factor as population was. Income is directly proportional to the rich status of a neighborhood. If people in a neighborhood earn more than an average income, then it is very much possible that they will spend more ,however that is not always true . Therefore,a restaurant assessment is proportional to the income of a neighborhood.

Data collection : Collecting geographical coordinates was a bit tricky, it was not available on open source data websites such as wikipedia, india gov website, census report websites etc. So I decided to use Google maps API to fetch latitude and longitude but google API has limited number of calls that I could make with my free account. Initially I scrapped list of neighbor's using beautifulSoup4 from wikipedia. The table headings becoming the boroughs and data becoming the neighborhoods. Gurgaon has 8 boroghs and 64 neighborhoods. So I manually googled each neighborhood to find its corresponding latitude and longitude. After doing so, I produced the following dataframe.

import requests url ="https://raw.githubusercontent.com/DeepentiA/Capstone_Project/master/Gurgaon_dataset.csv (https://raw.githubusercontent.com/DeepentiA/Capstone_Project/master/Gurgaon_dataset.csv)" df = pd.read_csv(url) df.head(20)

Population by neighborhood is again easy to find out given that its readily available. But incase of Gurgaon, it is again not the case. I was able to find population data for few cities. Here is the https://www.census2011.co.in/census/district/225-gurgaon.html (https://www.census2011.co.in/census/district/225-gurgaon.html) . Rest other neighborhood population is assumed and may be inaccurate but since this is a demonstrating project, the main idea to get the working model. Income by neighborhood is again easy to find out given that its readily available. But incase of Gurgaon, it is again not the case. I was able to find Income data for main city.https://www.payscale.com/research/IN/Location=Gurgaon-Haryana/SalaryNeighborhood (https://www.payscale.com/research/IN/Location=Gurgaon-Haryana/SalaryNeighborhood) Income is assumed and may be inaccurate but since this is a demonstrating project, the main idea to get the working model.

Gurgaon Income

In [30]:

```
url ="https://raw.githubusercontent.com/DeepentiA/Capstone_Project/master/Gurgaon_income.csv"
df = pd.read_csv(url)
df.head(20)
```

Out[30]:

	Borough	Neighborhoods	AverageIncome	Unnamed: 3
0	0	Central	Cantonment area	18944.099790
1	1	Central	Domlur	56837.022200
2	2	Central	Indiranagar	41991.817440
3	3	Central	Jeevanbheemanagar	6667.447632
4	4	Central	Malleswaram	53270.063890
5	5	Central	Pete area	50712.430220
6	6	Central	Rajajinagar	60967.535870
7	7	Central	Sadashivanagar	59943.541560
8	8	Central	Seshadripuram	58407.090340
9	9	Central	Shivajinagar	55850.962100
10	10	Central	Ulsoor	41007.219540
11	11	Central	Vasanth Nagar	26168.448090
12	12	Eastern	Bellandur	7227.731930
13	13	Eastern	CV Raman Nagar	54335.368710
14	14	Eastern	Hoodi	22591.063480
15	15	Eastern	Krishnarajapuram	36934.737730
16	16	Eastern	Mahadevapura	35915.973330
17	17	Eastern	Marathahalli	58448.658520
18	18	Eastern	Varthur	36433.267300
19	19	Eastern	Whitefield	44637.984600

Gurgaon population

In [29]:

```
url ="https://raw.githubusercontent.com/DeepentiA/Capstone_Project/master/Gurgaon_population.csv"
df = pd.read_csv(url)
df.head(20)
```

Out[29]:

	Borough	Neighborhoods	Population
0	Central	Cantonment area	866377
1	Central	Domlur	743186
2	Central	Indiranagar	474289
3	Central	Jeevanbheemanagar	527874
4	Central	Malleswaram	893629
5	Central	Pete area	730999
6	Central	Rajajinagar	981362
7	Central	Sadashivanagar	662625
8	Central	Seshadripuram	396862
9	Central	Shivajinagar	77836
10	Central	Ulsoor	656726
11	Central	Vasanth Nagar	942711
12	Eastern	Bellandur	208094
13	Eastern	CV Raman Nagar	122714
14	Eastern	Hoodi	330409
15	Eastern	Krishnarajapuram	351936
16	Eastern	Mahadevapura	905568
17	Eastern	Marathahalli	249182
18	Eastern	Varthur	546186
19	Eastern	Whitefield	83029

FourSquare API

In [31]:

```
```

In [35]:

```
---------------------------------------------------------------------------
ValueError                                Traceback (most recent call last)
<ipython-input-35-aacac5b36d8c> in <module>
      7             columns=['City', 'Borough'],
      8             key_on='feature.id',
----> 9             fill_color='YlGnBu', fill_opacity=0.7, line_opacity=0.2,
     10             )

~/conda/envs/python/lib/python3.6/site-packages/folium/folium.py in choroplet
h(self, geo_data, data, columns, key_on, threshold_scale, fill_color, fill_op
acity, line_color, line_weight, line_opacity, name, legend_name, topojson, re
set, smooth_factor, highlight)
    325                 style_function=style_function,
    326                 smooth_factor=smooth_factor,
--> 327                 highlight_function=highlight_function if highlight el
se None)
    328
    329         self.add_child(geo_json)

~/conda/envs/python/lib/python3.6/site-packages/folium/features.py in __init_
_(self, data, style_function, name, overlay, control, smooth_factor, highligh
t_function)
    493                 raise ValueError(msg)
    494         else:
--> 495             raise ValueError('Unhandled object {!r}.'.format(data))
    496
    497         if style_function is None:

ValueError: Unhandled object     Borough Neighborhoods         AverageIncome
Unnamed: 3
0          0       Central       Cantonment area  18944.099790
1          1       Central                Domlur  56837.022200
2          2       Central           Indiranagar  41991.817440
3          3       Central     Jeevanbheemanagar   6667.447632
4          4       Central           Malleswaram  53270.063890
..       ...           ...                   ...           ...
59        59       Western            Nagarbhavi  38627.411760
60        60       Western        Nandini Layout  32490.969170
61        61       Western          Nayandahalli  46826.803890
62        62       Western  Rajarajeshwari Nagar  12533.785280
63        63       Western           Vijayanagar  51966.782270

[64 rows x 4 columns].
```

In [ ]: