Capstone-hw5_report

July 18, 2020

1. Introduction/Business Problem section

i) Background

Korean food(Asian food) is getting more and more famous in Seattle lately. We can see the Korean dining is everywhere in United States, especially in Los angela. We are going to find the best location to open in Seattle and do more research about the business opportuntiy in Seattle.

This report will try to gather data about other restaurant localization, competitors and best localization.

ii) Problem

As the goal of this is to create a business plan in the end, we need to make sure data from api are correct. We also need to check that customer could be interested in this specific business.

In order to do so, a survey in Seattle and Los Angeles will be done in addition to data gathering. I'll go in the cities and check at different hours if restaurants are working, if streets are full and so on, and what kind of restaurant works well. This survey will allow to validate the data analysis done here.

2. Data

This notebook is highly inspirated by the template given in the course. I will keep the idea of clustering the city by area and then plot heatmap to find better area.

I will change some data:

Country/City: United States Goal: Open a restaurant/little shop for workers in weekday and maybe saturday So, I will cross data from working days, and localisations.

I will use the following API:

Foursquare API: to find restaurant/venues Google API: reverse geolocalisation

```
Coordinate of Seattle, WA: [47.6038321, -122.3300624]
```

I created a grid of area candidates, equaly spaced, centered around city center and within ~1.5km from Prefecture. Our neighborhoods will be defined as circular areas with a radius of 100 meters, so our neighborhood centers will be 200 meters apart.

To accurately calculate distances we need to create our grid of locations in Cartesian 2D coordinate system which allows us to calculate distances in meters (not in latitude/longitude degrees). Then we'll project those coordinates back to latitude/longitude degrees to be shown on Folium map. So let's create functions to convert between WGS84 spherical coordinate system (latitude/longitude degrees) and UTM Cartesian coordinate system (X/Y coordinates in meters).

```
Coordinate transformation check
-------------------------------
Seattle center longitude=-122.3300624, latitude=47.6038321
Seattle center UTM X=-2652222.8719972586, Y=13773159.99848766
Seattle center longitude=-122.33006239999997, latitude=47.60383210000003
```

create a hexagonal grid of cells: we offset every other row, and adjust vertical row spacing so that every cell center is equally distant from all it's neighbors.

[6]:
```
<folium.folium.Map at 0x7fb2a94797b8>
```

get approximate addresses of those locations

```
Reverse geocoding check
-----------------------
Address of [47.6038321, -122.3300624] is: Seattle City Hall, 600, 4th Avenue,
West Edge, International District/Chinatown, Seattle, King County, Washington,
98104, United States of America

Location data loaded from pickle.
```

[10]:

	Address	Latitude	Longitude	\
0	217, 12th Avenue, Central Business District, Y...	47.604092	-122.316915	
1	300, 10th Avenue, Central Business District, Y...	47.604993	-122.318870	
2	412, Broadway, Central Business District, Yesl...	47.605895	-122.320825	
3	Swedish First Hill Medical Center, 747, Broadw...	47.606796	-122.322781	
4	O'Dea High School, 802, Terry Avenue, Central ...	47.607697	-122.324736	
5	Bloodworks Northwest, 921, Terry Avenue, Centr...	47.608598	-122.326692	
6	Virginia Mason Hospital & Seattle Medical Cent...	47.609500	-122.328648	
7	1215, 8th Avenue, Central Business District, F...	47.610401	-122.330604	
8	Washington State Convention Center, 800, Conve...	47.611302	-122.332560	
9	Hotel Theodore, 1531, 7th Avenue, Central Busi...	47.612203	-122.334516	

	X	Y	Distance from center
0	-2.653123e+06	1.377251e+07	1111.720843
1	-2.652923e+06	1.377251e+07	957.038784
2	-2.652723e+06	1.377251e+07	822.145506
3	-2.652523e+06	1.377251e+07	718.277964
4	-2.652323e+06	1.377251e+07	660.244828
5	-2.652123e+06	1.377251e+07	660.244828
6	-2.651923e+06	1.377251e+07	718.277964
7	-2.651723e+06	1.377251e+07	822.145506
8	-2.651523e+06	1.377251e+07	957.038784
9	-2.651323e+06	1.377251e+07	1111.720843

3. Methodology

Now that we have our location candidates, let's use Foursquare API to get info on restaurants in each neighborhood.

We're interested in venues in 'food' category, but only the ones who can be competitors, this mean food truck, quick food, take away, healthy, not restaurant taking too long.

```
Restaurant data loaded.

Total number of restaurants: 309
Total number of Asian restaurants: 165
Percentage of Asian restaurants: 53.40%
Average number of restaurants in neighborhood: 1.4636363636363636
```

Let's now see all the collected restaurants in our area of interest on map, and let's also show Asian restaurants in different color.

[16]: `<folium.folium.Map at 0x7fb2a96a90b8>`

So, on this map we can see all potential competitors, take away restaurant, healthy restaurant. Asian restaurant are in red.

No we need to analyse companies/university localisation and cluster and cross both analysis.

We need to remember that the target are worker wanting to buy healthy fast food for breakfast and lunch, we don't aim the evening. Also, another target can be universities.

```
Universities data loaded.
Total number of universities: 67
Average number of universities in neighborhood: 0.41818181818181815
```

[17]: `<folium.folium.Map at 0x7fb2a9a28390>`

Companies visualization

```
Companies data loaded.
Total number of companies: 2752
Average number of companies in neighborhood: 16.318181818181817
```

[18]: `<folium.folium.Map at 0x7fb2a9ea5e80>`

Global map of worker/students versus Restaurants

[19]: `<folium.folium.Map at 0x7fb2ac36fc88>`

green:company yellow:university red:asian restaurant bllue:non-asian restaurant

Creation and visualization of customer group:

```
Companies data loaded.
Total number of customers: 5
Average number of customers in neighborhood: 110.0
```

[20]: `<folium.folium.Map at 0x7fb2b08f36d8>`

4. Analysis

Let's perform some basic explanatory data analysis and derive some additional info from our raw data. First let's count the number of business in every area candidate:

[21]: `110`

```
Average number of customers in every area with radius=100m: 16.490909090909092
```

[22]:
```
                                          Address   Latitude   Longitude  \
0  217, 12th Avenue, Central Business District, Y...  47.604092 -122.316915
1  300, 10th Avenue, Central Business District, Y...  47.604993 -122.318870
2  412, Broadway, Central Business District, Yesl...  47.605895 -122.320825
3  Swedish First Hill Medical Center, 747, Broadw...  47.606796 -122.322781
4  O'Dea High School, 802, Terry Avenue, Central ...  47.607697 -122.324736

              X             Y  Distance from center  Customers in area
0 -2.653123e+06  1.377251e+07           1111.720843                  5
1 -2.652923e+06  1.377251e+07            957.038784                  2
2 -2.652723e+06  1.377251e+07            822.145506                 23
3 -2.652523e+06  1.377251e+07            718.277964                  9
4 -2.652323e+06  1.377251e+07            660.244828                  9
```

top 10 area

[23]:
```
                                          Address   Latitude   Longitude  \
36  Daniels Recital Hall, 801, 5th Avenue, Central...  47.605613 -122.331132
75  Grand Central, 107, Occidental Avenue South, W...  47.600130 -122.333791
6   Virginia Mason Hospital & Seattle Medical Cent...  47.609500 -122.328648
65  The Halal Guys, 105, Yesler Way, West Edge, In...  47.601726 -122.333616
16  782, Madison Street, Central Business District...  47.607904 -122.328824
45  701, 4th Avenue, Central Business District, Fi...  47.604017 -122.331308
22  Harborview Medical Center, 325, 9th Avenue, Ce...  47.603604 -122.323133
37  Central Library, 1000, 4th Avenue, Central Bus...  47.606514 -122.333088
19  U.S. Bank Centre, 1420, 5th Avenue, Central Bu...  47.610607 -122.334692
9   Hotel Theodore, 1531, 7th Avenue, Central Busi...  47.612203 -122.334516

               X             Y  Distance from center  Customers in area
36 -2.652023e+06  1.377303e+07            240.192379                 62
75 -2.652223e+06  1.377372e+07            559.807621                 56
6  -2.651923e+06  1.377251e+07            718.277964                 56
65 -2.652123e+06  1.377355e+07            399.326338                 53
16 -2.652023e+06  1.377268e+07            519.467306                 52
45 -2.652123e+06  1.377320e+07            107.774892                 51
22 -2.652723e+06  1.377285e+07            586.318455                 46
37 -2.651823e+06  1.377303e+07            421.535739                 43
19 -2.651423e+06  1.377268e+07            932.655500                 42
9  -2.651323e+06  1.377251e+07           1111.720843                 41
```

[25]:
```
<folium.folium.Map at 0x7fb2a9a1b128>
```

We can clearly identify two big cluster, on northwest, on southwest.

Let's create another heatmap map showing heatmap/density of restaurants.

[26]:
```
<folium.folium.Map at 0x7fb2939354e0>
```

As expected, the restaurants are also in this area, but if we look at the first map, we can see that we almost have no competitors in the central area.

K mean clustering

Let's do the kmean clustering to see what will be the result.

[27]: `<folium.folium.Map at 0x7fb2a96a49e8>`

```
563 candidate neighborhood centers generated.
```

OK. Now let's calculate two most important things for each location candidate: number of restaurants in vicinity (we'll use radius of 150 meters) and number of customers.

```
Generating data on location candidates... done.
```

[30]:

	Latitude	Longitude	X	Y	Restaurants nearby	\
186	47.605353	-122.331393	-2.652023e+06	1.377307e+07	0	
311	47.602575	-122.334464	-2.652004e+06	1.377350e+07	3	
185	47.604903	-122.330415	-2.652123e+06	1.377307e+07	5	
335	47.601466	-122.333877	-2.652123e+06	1.377359e+07	14	
384	47.599870	-122.334053	-2.652223e+06	1.377376e+07	6	
310	47.602125	-122.333486	-2.652104e+06	1.377350e+07	8	
210	47.604416	-122.331178	-2.652104e+06	1.377315e+07	4	
408	47.598933	-122.333837	-2.652304e+06	1.377385e+07	4	
289	47.604414	-122.336635	-2.651723e+06	1.377341e+07	3	
211	47.604866	-122.332156	-2.652004e+06	1.377315e+07	2	

	Distance to Asian restaurant	Customers
186	113.335435	153
311	91.689117	143
185	30.687867	125
335	14.976102	122
384	72.177583	122
310	39.721710	120
210	26.981858	119
408	176.485481	118
289	104.479281	117
211	56.481524	116

OK. Let us now filter those locations: we're interested only in locations with no more than two restaurants in radius of 250 meters, and no asian restaurants in radius of 100 meters, and more than 10 customers.

```
Locations with no more than two restaurants nearby: 430
Locations with no Asian restaurants within 400m: 362
Locations with more than 10 customers: 351
Locations with both conditions met: 220
```

Let's see how this looks on a map.

[32]: `<folium.folium.Map at 0x7fb2a96a4f60>`

We can identify two mains areas, one south, the other one north as expected.
Let's see heatmap:

[33]: `<folium.folium.Map at 0x7fb2a9b86b70>`

Looking good. What we have now is a clear indication of zones with low number of restaurants in vicinity, and no Asian restaurants at all nearby, and good numbers of customers.

5. Result

Let us now cluster those locations to create centers of zones containing good locations. Those zones, their centers and addresses will be the final result of our analysis.

[34]: `<folium.folium.Map at 0x7fb293e2f240>`

Addresses of those cluster centers will be a good starting point for exploring the neighborhoods to find the best possible location based on neighborhood specifics.

Let's see those zones on a city map without heatmap, using shaded areas to indicate our clusters:

[35]: `<folium.folium.Map at 0x7fb2936651d0>`

Finaly, let's reverse geocode those candidate area centers to get the addresses

```
============================================================
Addresses of centers of areas recommended for further analysis
============================================================

Hambach Building, South King Street, West Edge, International
District/Chinatown, Seattle, King County, Washington, 98104 => 0.8km from
Prefecture
5th and Madison Condos, 909, 5th Avenue, Central Business District, First Hill,
Seattle, King County, Washington, 98164 => 0.3km from Prefecture
1086, Jefferson Street, Central Business District, First Hill, Seattle, King
County, Washington, 98104 => 0.7km from Prefecture
Third and Stewart Garage, 3rd Avenue, Central Business District, Belltown,
Seattle, King County, Washington, 98181 => 1.2km from Prefecture
Seattle King Street, 301, South Jackson Street, West Edge, International
District/Chinatown, Seattle, King County, Washington, 98104 => 0.7km from
Prefecture
741, James Street, Central Business District, First Hill, Seattle, King County,
Washington, 98104 => 0.3km from Prefecture
M Street Medical Building, 910, 8th Avenue, Central Business District, First
Hill, Seattle, King County, Washington, 98104 => 0.5km from Prefecture
61, Columbia Street, West Edge, International District/Chinatown, Seattle, King
County, Washington, 98104 => 0.6km from Prefecture
398, Dilling Way, West Edge, International District/Chinatown, Seattle, King
County, Washington, 98104 => 0.2km from Prefecture
Two Union Square, 601, Union Street, Central Business District, First Hill,
Seattle, King County, Washington, 98101 => 0.8km from Prefecture
House of Hong, 409, 8th Avenue South, West Edge, International
District/Chinatown, Seattle, King County, Washington, 98104 => 0.9km from
Prefecture
Benaroya Hall, 200, University Street, West Edge, Belltown, Seattle, King
County, Washington, 98101 => 0.7km from Prefecture
814, 6th Avenue South, West Edge, International District/Chinatown, Seattle,
King County, Washington, 98134 => 1.1km from Prefecture
Coastal Environmental Systems, 820, 1st Avenue South, West Edge, International
District/Chinatown, Seattle, King County, Washington, 98134 => 1.1km from
```

```
Prefecture
Waterfront Park, 1301, Alaskan Way, Pike Place Market Area, Belltown, Seattle,
King County, Washington, 98101 => 1.0km from Prefecture
```

6. Discussion

From the above result, we can know the different address of our choice of the best location of restaurant. Seattle is a large city, so the number and density of restaurant is quite high, in this analysis I tried to corrolate the number of restaurant and quantity of potential customer.

the Prefecture area is distributed near the center of Seattle.

We must just take care of one thing, I thing the api didn't return all data, we are missing a lot of companies, the map is still good, and the result can be trusted, but we should cross check data with other data source.

In order to be more accurate, it could be possible to give a weight to customers for example, a university with 1000 student would then weight more than a hair cut company with two employees.

7.Conclusion

The idea of this project came from the homework requirement and one data case from kaggle. I applied my own data source and methdology to it.Also, I have created/modify a huge quantity of function in order to adapt.

It's very far from being perfect, a lot of work can be done, other source of data can be found, but in the end the result seams to correlate with the real world, when we know the city, the area predicted seams correct.