IBM APPLIED DATA SCIENCE CAPSTONE PROJECT

Which city would be best to open up a restaurant and what type of restaurant should one open?

“WHICH CITY WOULD BE BEST TO OPEN UP A RESTAURANT AND WHAT TYPE OF RESTAURANT SHOULD ONE OPEN?"

The project intends to answer this question.

In particular, does it make sense to establish a Halal restaurant in Seoul?

METHODOLOGY

1 Data retrieval and cleaning

- First stage of the project will focus on retrieving data from the data sources and clean the data so that they can be processed with Pandas data frame. Data was scrapped using BeautifulSoup4. Sc rapped data usually contains unneeded characters and signs and should be cleaned.

2 Venue query in a target location

- Second stage of the project will focus on retrieving geolocation data from Foursquare and look for restaurants in the target city. The project will count the number of restaurants, broken down into c ategories. The project will determine a type of restaurant that is least present in the area.

3 Look for opportunities

- Third stage of the project will seek for additional information that will reinforce the recommendatio n for a restaurant in the area from the above.

INITIAL DATA RETRIEVAL FROM WIKIPEDIA

- Scrapped data requires cleaning
- Special care should be taken to also remove hidden non-UTF characters in the string

	City	Nation	Population (Proper)	Population (Metro)	Population (Urban)
0	Chongqing	China\n	30,751,600[8]\n	17,000,000[9]\n	8,165,500[a]\n
1	Shanghai	China\n	24,256,800[11]\n	24,750,000[12]\n	23,416,000[b]\n
2	Beijing	China\n	21,516,000[13]\n	24,900,000[14]\n	21,009,000\n
3	Lagos	Nigeria\n	16,060,303[c]\n	21,000,000[17]\n	13,123,000\n
4	Dhaka	Bangladesh\n	8,906,039[18]\n	20,000,000[19]\n	\n

CLEANING DATA

- Data cleaned and column types corrected for mathematical computations
- Notice that the population columns are now showing as integers

	City	Nation	Population (Proper)	Population (Metro)	Population (Urban)	Population
10	Tokyo	Japan	13839910	38140000	38505000	38505000
13	São Paulo	Brazil	12038175	21090791	36842102	36842102
0	Chongqing	China	30751600	17000000	8165500	30751600
20	Jakarta	Indonesia	10075310	30539000	30075310	30539000
19	Seoul	Korea, South	10197604	12700000	25520000	25520000

ADDING MORE DATA

- Cities' GDP data has been downloaded and cleaned
- Every city has at least 1 data from the 5 estimates

	City	Nation	Estimate 1	Estimate 2	Estimate 3	Estimate 4	Estimate 5	GDP ($bn)
340	Tokyo	Japan	1893.000	1617.0	1479.0	1874.7	1997.5	1893.000
241	New York	United States	1717.712	1403.0	1406.0	1180.3	1056.4	1717.712
198	Los Angeles	United States	1043.735	860.5	792.0	731.8	632.4	1043.735
305	Seoul	South Korea	738.600	845.9	291.0	233.3	0.0	738.600
256	Paris	France	724.000	715.1	564.0	764.2	0.0	724.000

MERGING DATAFRAMES

- Dataframe was merged between population and GDP
- Needs careful analysis of data to verify that the city and nation names are identical between the two data frames

	City	Nation	Population	GDP ($bn)
0	Tokyo	Japan	38505000	1893.000
1	São Paulo	Brazil	36842102	582.079
2	Chongqing	China	30751600	288.800
3	Jakarta	Indonesia	30539000	186.000
4	Seoul	Korea, South	25520000	738.600

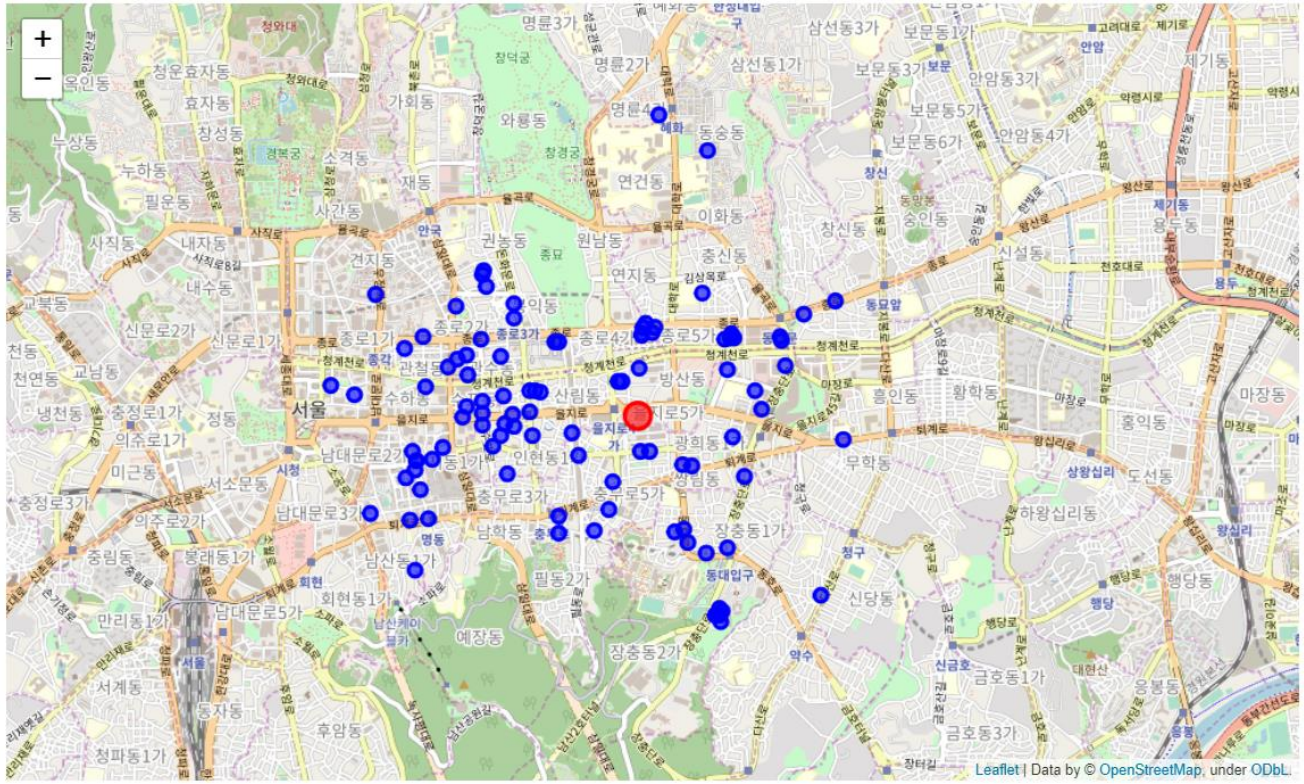
ADDING GEOSPATIAL DATA FOR CITIES

	City	Nation	Population	GDP ($bn)	GDP/capita ($)	lat	lng
0	Singapore	Singapore	5535000	349.5	63144	1.2930	103.8558
1	Paris	France	12405426	724.0	58362	48.8667	2.3333
2	Sydney	Australia	5230330	302.7	57874	-33.9200	151.1852
3	Vienna	Austria	2600000	131.9	50731	48.2000	16.3666
4	Tokyo	Japan	38505000	1893.0	49162	35.6850	139.7514
5	Toronto	Canada	6346088	303.0	47746	43.7000	-79.4200
6	Taipei	Taiwan	7045488	327.3	46455	25.0358	121.5683
7	London	United Kingdom	14040163	595.7	42428	51.5000	-0.1167
8	Rome	Italy	4353775	166.8	38312	41.8960	12.4833
9	Berlin	Germany	5871022	215.2	36655	52.5218	13.4015
10	Madrid	Spain	6378297	225.9	35417	40.4000	-3.6834
11	Warsaw	Poland	3100844	100.0	32249	52.2500	21.0000
12	Auckland	New Zealand	1614300	49.5	30663	-36.8481	174.7630
13	Seoul	Korea, South	25520000	738.6	28942	37.5663	126.9997
14	Dubai	United Arab Emirates	2865560	82.9	28930	25.2300	55.2800
15	Riyadh	Saudi Arabia	5676621	163.5	28802	24.6408	46.7727
16	Prague	Czechia	2619000	73.0	27873	50.0833	14.4660
17	Santiago	Chile	6683852	171.4	25644	-33.4500	-70.6670
18	Kuala Lumpur	Malaysia	7200000	171.8	23861	3.1667	101.7000
19	Istanbul	Turkey	14657000	348.7	23791	41.1050	29.0100

- Data downloaded from https://simplemaps.com/data/world-cities to retrieve the latitude and longitude coordinates for each major cities for the top 20 cities with the highest GDP per capita in US$

RETRIEVING VENUES RELATED TO FOOD FROM FOURSQUARE

- Seoul was selected randomly for further analysis
- Blue dots show locations for 50 food related venues retrieved from Foursquare query
- Red dot depicts the central point for Seoul's coordinate



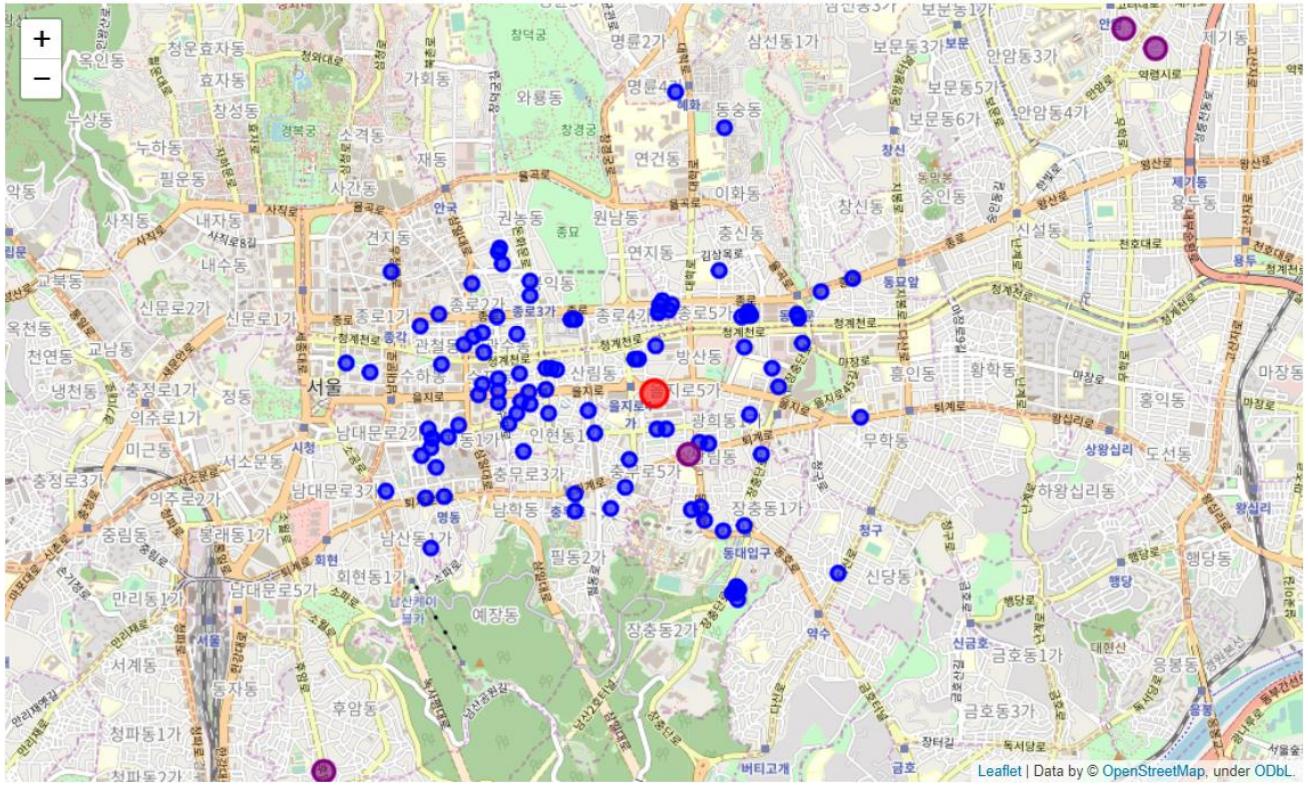
THE INITIAL RESULT

- Most international food restaurants are present from the query
- Notice that there are no Islam food restaurants in this list
- It may be due to the lack of demand, or could this be a sign for opportunity?

categories	Count
Korean Restaurant	35
Café	10
Noodle House	9
Bakery	6
Chinese Restaurant	5
BBQ Joint	4
Bistro	3
Seafood Restaurant	3
Japanese Restaurant	3
Italian Restaurant	2
Sushi Restaurant	2
Sandwich Place	2

categories	Count
Restaurant	2
Indian Restaurant	2
Fried Chicken Joint	2
Buffet	2
French Restaurant	1
Food Court	1
Eastern European Restaurant	1
Samgyetang Restaurant	1
Dumpling Restaurant	1
Burger Joint	1
Tibetan Restaurant	1
Vietnamese Restaurant	1

CONCLUSION



- It turns out that there are two mosques within the vicinity of the search area and five when you consider distance of 30 minutes' car drive
- The project concludes that there is potential for prospective investors or entrepreneurs to open up a Halal food restaurant in Seoul in the norther district

THANK YOU