The Battle of Neighborhoods

IBM-Data Science Specialization – Capstone Project

Nithu Singh

Table of Contents

Introduction

Problem Background

The famous Barbeque Nation owned by Sayaji Hotels Ltd is one of the high-end buffet restaurants which takes you into the world of good food with happy vibes. This place was founded in 2006 by one of their Promoters. They currently own and operate 138 outlets in India, 5 outlets in UAE, 1 outlet in Malaysia and 1 outlet in Oman.

One of the leading casual dining chains in India, Barbeque Nation pioneered the concept of “over the table barbeque” live grills embedded in dining tables – allowing guests to grill their own barbecue’s right at their tables.

Indian cuisine is a natural part of the diet in the markets in which the Barbeque Nation Restaurants operate. They continue to offer new menu options, both vegetarian and non-vegetarian, largely based on Indian cuisine and use seasonal customer preferences to introduce new dishes.

This is just an example, there are many such food chains and restaurants across India that could go international and would be open to the idea of setting up business in Ireland.

Problem Description

As mentioned above, there are many popular restaurant chains are as of now only serving across India, but if they were to go international, specifically open their first international chain of restaurants in Ireland. There are various factors that they'll have to look into.

Example of few such factors are,

* Popular cuisines in Ireland
* How popular is Indian food in Ireland?
* What are existing Indian Restaurants chains in Ireland?
* Which county is more preferable to open their first restaurant ?
* Are there "similar" restaurants? If so, what specialty do that have?

This will be more of a comparative analysis of existing food chains/restaurants across few counties in Ireland taking into account data from foursquare API.

Targeted Audience

This is very generic analysis on existing food restaurants chains across few counties in Ireland. Specific for research purpose, this analysis could also be extended towards restaurant and food chain owners with specific cuisines looking to open their businesses in Ireland and want to understand the current situation about people preferences, food popularity and existing businesses across required neighborhood.

Data

Given the problem at hand, we'll need a variety of data to perform our analysis.
For starter,

1. To lock down on the counties for analysis, I'll need population and ethnic dataset of different counties in Ireland. An assumption is set that, the more the Asian Population, the more Indian Restaurants in the area.
2. For selected counties, we require geographical co-ordinates to query foursquare API to collect data on restaurant serving different existing cuisines and their details.
3. Once all the data is obtained we can perform elaborate analysis on the venues across selected neighborhoods.

Initial Datasets

Based on our problem definition below are the sources from where each datasets were obtained,

To determine the counties and neighborhoods for analysis, the population dataset available on Central Statistics Office website give us a breakdown of population across Ireland based on multiple themes.
https://www.cso.ie/en/census/census2016reports/census2016smallareapopulationstatistics/

The shapefile used in the project was obtained from OSi Website
Link:

https://data-osi.opendata.arcgis.com/datasets/0d5984f732c54246bd087768223c92eb_0?geometry=-23.747%2C51.281%2C8.927%2C55.844

To analyze a suitable location for restaurant, neighborhood's Latitude and Longitude is to be known so that we can point at its coordinates and create a map displaying all the current restaurants with its labels respectively.

The geographical co-ordinates were copied manually from following website - https://latitude.to/map/ie/ireland/regions

The list of international cuisines were extracted from https://developer.foursquare.com/docs/build-with-foursquare/categories/

The ratings and price for different restaurants were obtained using google places API key.
Read more: https://developers.google.com/places/web-service/overview

Methodology

To keep the data obtained and analysis consistent and relevant, all the location data and query to foursquare API and google API was done for each administrative aeras within Republic of Ireland. This was the data that could be accurately obtained at the highest level. Having data per administrative area level could also be used for future analysis of different administrative areas for various statistics and research.

Data Importing, Cleaning and Manipulation

Most of the data cleaning was required to clean and merge both shape files with the population data to plot a geographical population heat map of counties in Ireland.

The data obtained from CSO website, are divided into population data and themes files describing it's header info.

Dataset obtained from Central Statistics Office website give us a breakdown of population across Ireland based on multiple themes.

	GUID	GEOGID	GEOGDESC	T1_1AGE0M	T1_1AGE1M	T1_1AGE2M	T1_1AGE3M	T1_1AGE4M	T1_1AGE5M	T1_1AGE6M	T1_1AGE7M
0	EB9428B8-B443-407E-977B-9393FD73E1D0	CTY31_CC	Cork City	720	614	613	645	602	643	647	598
1	84B85B58-010E-45DB-BC55-E137BBE627E4	CTY31_CE	Clare	762	759	780	842	868	843	929	954
2	8A998E65-D3BE-4620-98DE-EBF44DE0285B	CTY31_CK	Cork County	2957	3154	3174	3324	3365	3579	3628	3543
3	F1CE5AB6-2A36-4E32-B04A-E79CCD016BCE	CTY31_CN	Cavan	539	540	566	636	608	671	635	680
4	D3A82C01-E07A-4554-8FE4-355EBB161F5A	CTY31_CW	Carlow	349	440	383	426	459	475	442	455

And a separate file contains header info

```
header_desc.head()
```

	THEME	TABLES WITHIN THEMES	FIELD NAME WITHIN SAPS TABLES	DESCRIPTION OF FIELD
0	Theme 1 Sex, Age and Marital Status	Population aged 0-19 by sex and year of age, p...	T1_1AGE0M	Age 0 Males
1	Theme 1 Sex, Age and Marital Status	Population aged 0-19 by sex and year of age, p...	T1_1AGE1M	Age 1 Males
2	Theme 1 Sex, Age and Marital Status	Population aged 0-19 by sex and year of age, p...	T1_1AGE2M	Age 2 Males
3	Theme 1 Sex, Age and Marital Status	Population aged 0-19 by sex and year of age, p...	T1_1AGE3M	Age 3 Males
4	Theme 1 Sex, Age and Marital Status	Population aged 0-19 by sex and year of age, p...	T1_1AGE4M	Age 4 Males

Because our assumption is that the more the Asian Population, the more Indian Restaurants in the area, the group we're particularly interested in is Asian[Indian] or Asian Irish[Indian] and Total Population. People when living abord/away from crave for home country style food, and because this study is focused on Indian food. We're assuming them to be our largest customer base. This assumption is later tested in the analysis.

Some data manipulation that were performed are:

```
pop_shp.ADMINISTRATIVE_AREA = pop_shp.ADMINISTRATIVE_AREA.apply(lambda x: x if "CITY" in x else x.replace(' COUNTY', ''))
pop_data = pop_shp.merge(roi_pop, left_on = "ADMINISTRATIVE_AREA", right_on = "ADMINISTRATIVE_AREA", how = 'left')
pop_data[pd.isnull(pop_data).any(axis=1)]
```

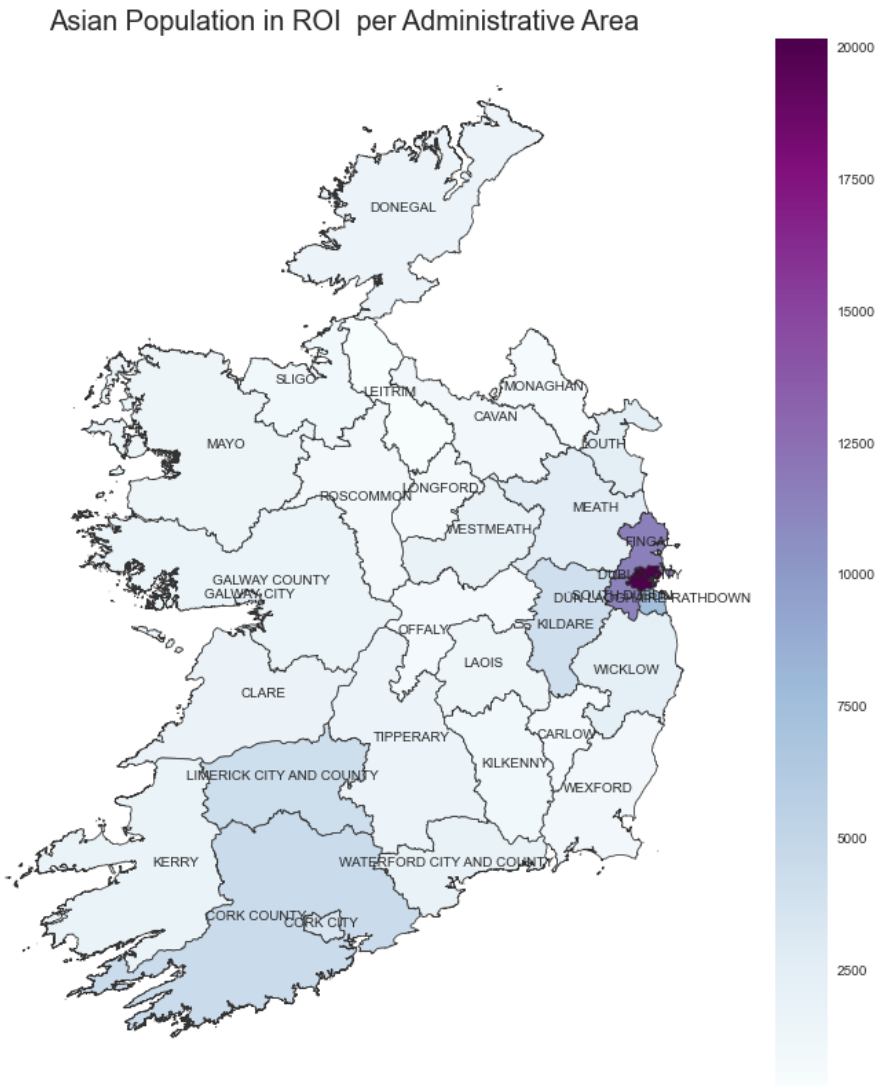
	ADMINISTRATIVE_AREA	COUNTY	PROVINCE	geometry	ASIAN_POPULATION	TOTAL_POPULATION
21	DUN LAOGHAIRE-RATHDOWN	DUBLIN	Leinster	MULTIPOLYGON (((726029.569 724958.910, 725964....	NaN	NaN
22	GALWAY	GALWAY	Connacht	MULTIPOLYGON (((587504.376 731428.868, 587456....	NaN	NaN
24	CORK	CORK	Munster	MULTIPOLYGON (((539837.098 620406.502, 539702....	NaN	NaN

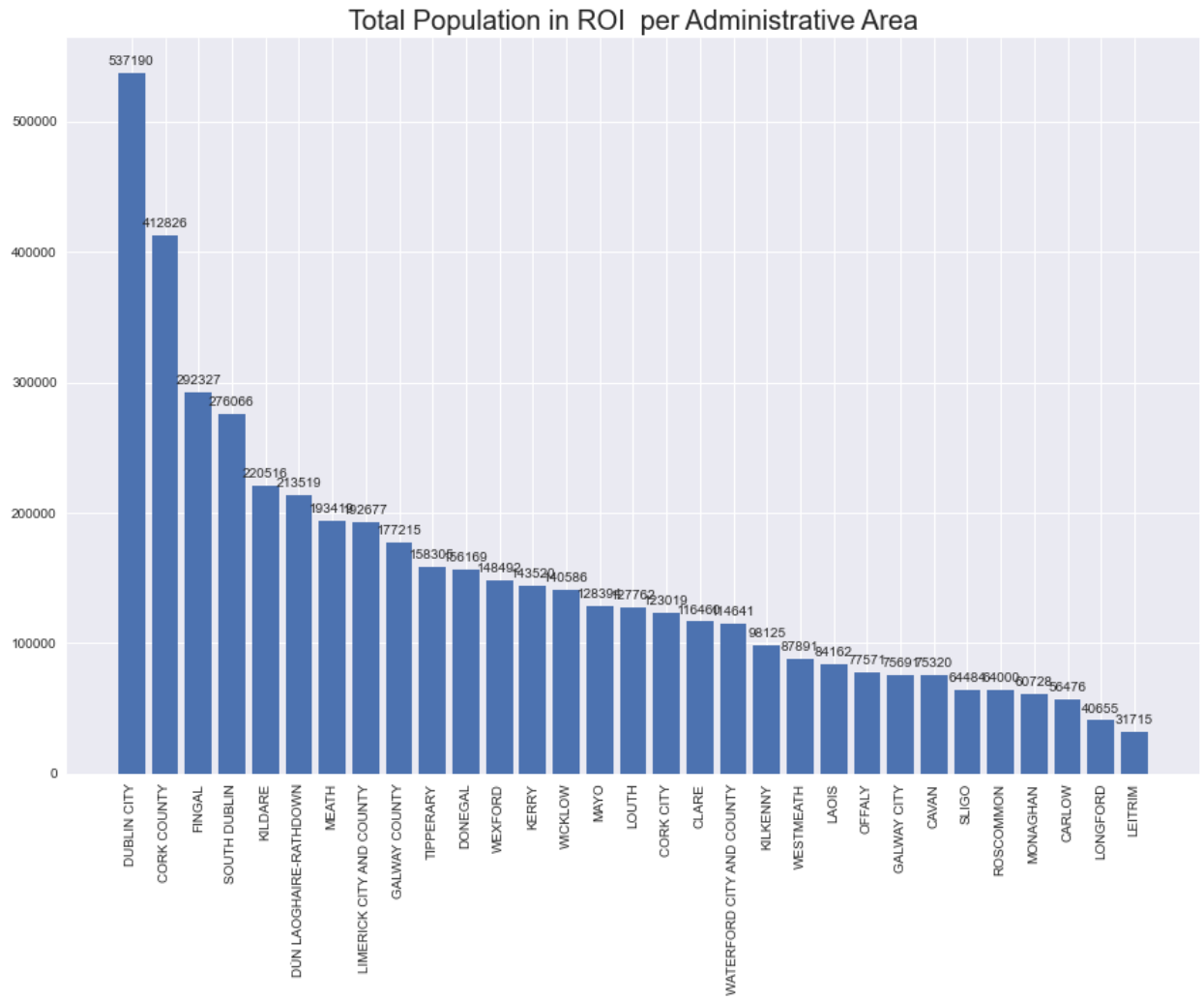
```
pop_shp.loc[24,"ADMINISTRATIVE_AREA"] = "CORK COUNTY"
pop_shp.loc[22,"ADMINISTRATIVE_AREA"] = "GALWAY COUNTY"
pop_shp.loc[21,"ADMINISTRATIVE_AREA"] = "DÚN LAOGHAIRE-RATHDOWN"
```

```
pop_data = pop_shp.merge(roi_pop, left_on = "ADMINISTRATIVE_AREA", right_on = "ADMINISTRATIVE_AREA", how = 'left')
```

Here, the columns 'ASIAN_POPULATION' and 'TOTAL_POPULATION' are after merge are Nan due to mismatch in Administrative Area columns.

Based on all of the data manipulation above, see below graph of Asian and Total population across different Administrative Areas in ROI





FOURSQUARE API

The data above is combined with geographical dataset to obtain the following table after which top 10 counties are locked for analysis.

Even thou an assumption is set that, the more the Asian Population, the more Indian Restaurants in the area I am analyzing top 10 counties by total population.

```
top_aa.head()
```

ADMINISTRATIVE_AREA	COUNTY	PROVINCE	ASIAN_POPULATION	TOTAL_POPULATION	Latitude	Longitude	ll	radius
DUBLIN CITY	DUBLIN	Leinster	20172	537190	53.35512	-6.24922	53.35511999999999,-6.249219999999999	4500
CORK COUNTY	CORK	Munster	4436	412826	51.96667	-8.58333	51.96667,-8.58333	35000
FINGAL	DUBLIN	Leinster	11514	292327	53.45909	-6.21942	53.45909,-6.2194199999999995	6300
SOUTH DUBLIN	DUBLIN	Leinster	11334	276066	53.28595	-6.37739	53.28595,-6.37739	5700
KILDARE	KILDARE	Leinster	4064	220516	53.16667	-6.75000	53.166669999999996,-6.75	16000

In the data frame above, radius for analysis for each administrative area was obtained by using maps.ie to Measure Circle / Radius on a map to determine approx. shortest radius of each administrative area without overlap.

Inspired by https://github.com/gbarreiro/best_cuisines

For next part of the analysis, i.e., to obtain the number of restaurants serving each cuisine and their details the ***Foursquare API*** is used. As a user, you need a Developer Account, in order to get the credentials for authenticating your HTTP requests. For each administrative are selected above, we call the API for the restaurants of each kind of international cuisine, one by one, and store the results in a Pandas DataFrame, in order to process it later and perform some interesting analytics.
From the data, we manually extract the categories under food corresponding to national cuisines (Spanish, Indian, Thai...), in order to be able of filtering the restaurants only by the nationality, and not other irrelevant criteria, like vegan, halal, burger place.

For each administrative area, call to API is made for the restaurants of each cuisine. i.e, query.get('https://api.foursquare.com/v2/venues/explore?&client_id={}&client_secret={}&v={}&limit={}&radius={radius}&ll={lat_lon}&categoryId={category_id}'.format())

The results for which are comes in json format for which required information is queried and stored in corresponding data frames.

restaurants_summary –

```
restaurants_summary
```

	Administrative_Area	Cuisine	Number of restaurants
0	DUBLIN CITY	Afghan	2
1	DUBLIN CITY	African	4
2	DUBLIN CITY	American	27
3	DUBLIN CITY	Burmese	0
4	DUBLIN CITY	Cambodian	0
5	DUBLIN CITY	Chinese	53
6	DUBLIN CITY	Filipino	1
7	DUBLIN CITY	Himalayan	1
8	DUBLIN CITY	Indonesian	1
9	DUBLIN CITY	Japanese	39
10	DUBLIN CITY	Korean	10
11	DUBLIN CITY	Malay	5

restaurants_details -

```
restaurants_details
```

	Administrative_Area	Cuisine	Restaurant_id	Restaurant_Name	Restaurant_Lat	Restaurant_Lon
0	DUBLIN CITY	Afghan	57f544f4498e9282a9682873	Afghan Restaurant	53.338610	-6.265878
1	DUBLIN CITY	African	4e398411b61c438b5487b701	Akwaaba	53.352742	-6.244516
2	DUBLIN CITY	African	4afbcb53f964a520141f22e3	Il Caffè di Napoli	53.342692	-6.250147
3	DUBLIN CITY	African	4c13966ba5eb76b0f01ec0b7	Akanchawa's Honey Pot	53.354090	-6.269668
4	DUBLIN CITY	American	4cdd9beaf8a4a1431027d1bc	The Butcher Grill	53.324288	-6.252728
5	DUBLIN CITY	American	4b8d5387f964a520c1f432e3	TriBeCa	53.324367	-6.252996
6	DUBLIN CITY	American	4ade0f09f964a520de7021e3	Hard Rock Cafe Dublin	53.345878	-6.260866
7	DUBLIN CITY	American	4fc62996e4b0505c6e36dba1	TGI Fridays	53.345863	-6.259748
8	DUBLIN CITY	American	51acdfab498ea50be7f3de5f	Bunsen	53.337270	-6.265669
9	DUBLIN CITY	American	59eb6000f62e0902f051ff72	Ruby's	53.348447	-6.227838
10	DUBLIN CITY	American	4af2d5a7f964a520abe821e3	Herbstreet	53.344294	-6.237792
11	DUBLIN CITY	American	4b19a00df964a520ade023e3	Eddie Rocket's	53.347805	-6.259827

After having all the data downloaded, the DataFrame had to be transformed, in order to be ready for getting the information of interest. Through a one-hot encoding of the cuisines and a grouping by city, a new DataFrame was created, where each row was a city, and each column the number of restaurants of each cuisine. Finally, the rows were normalized, so each column represented the percentage of restaurants of each cuisine for each city. This was done because depending on the size, lifestyle and other factors, in some cities there might be a higher number of restaurants than in others.

Google Places API

After preprocessing this data I further went on to use the ***Google Places API*** to determine the ratings and price of each restaurant that is available to us in the resturants_details dataframe. It was a surprise that only a few restaurants out of the total had ratings and price details available for them.

After this further analysis were done to answer a few question we defined in our problem definition.

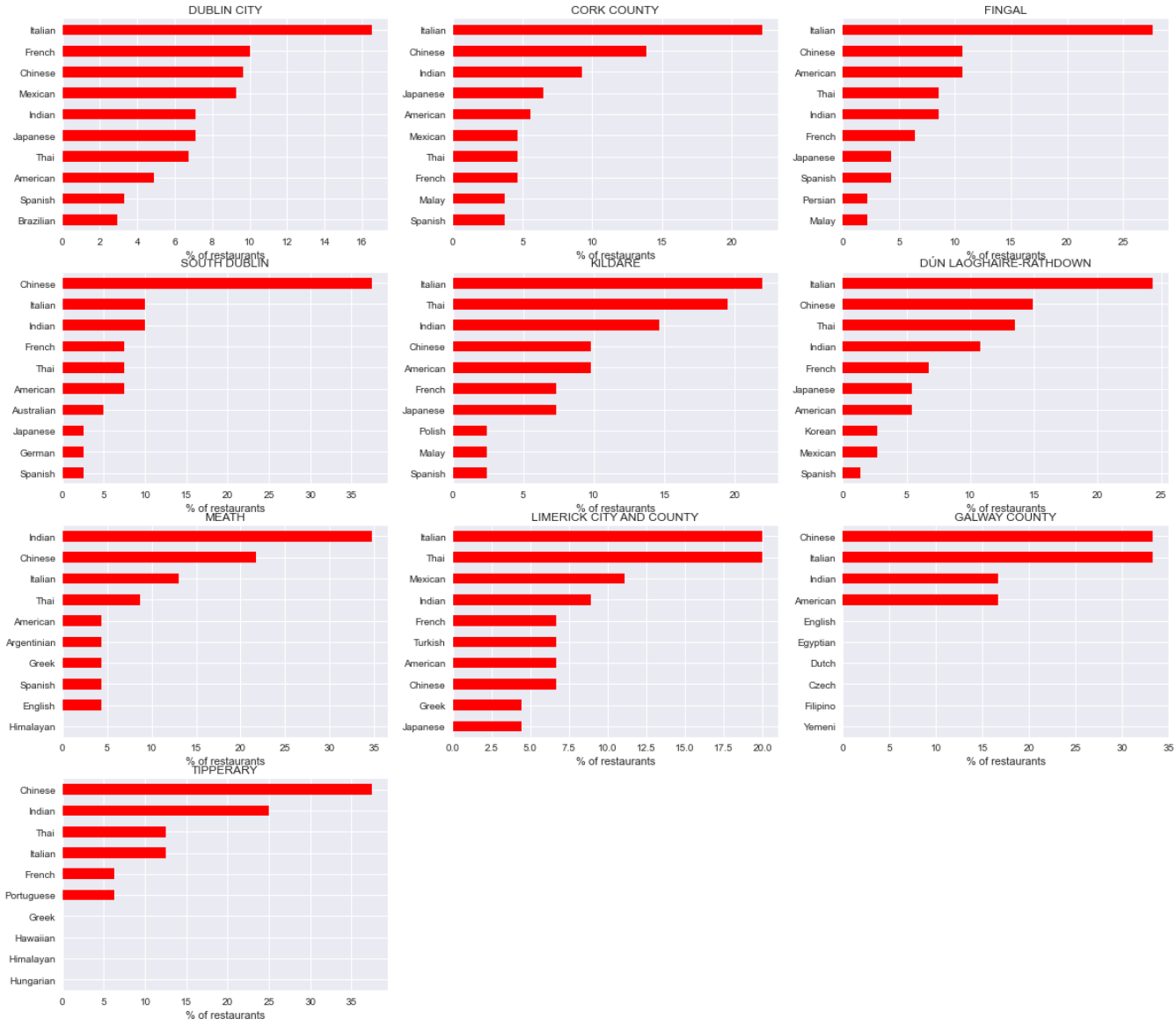
Results

Popular cuisines in Ireland - how popular is Indian food in Ireland?

Before looking at the results, it's important to make following observations. In this project, we're working with 65 different types of cuisines, those available in Foursquare corresponding to a country. Since the results shown here are in percentages (over 100), in an equative division, a 1.54% would correspond to each cuisine . I think this is an important clarification, since without it, it might seem that a percentage of 20% for some cuisine in some administrative are might not be too much, when actually it would be a really high percentage.

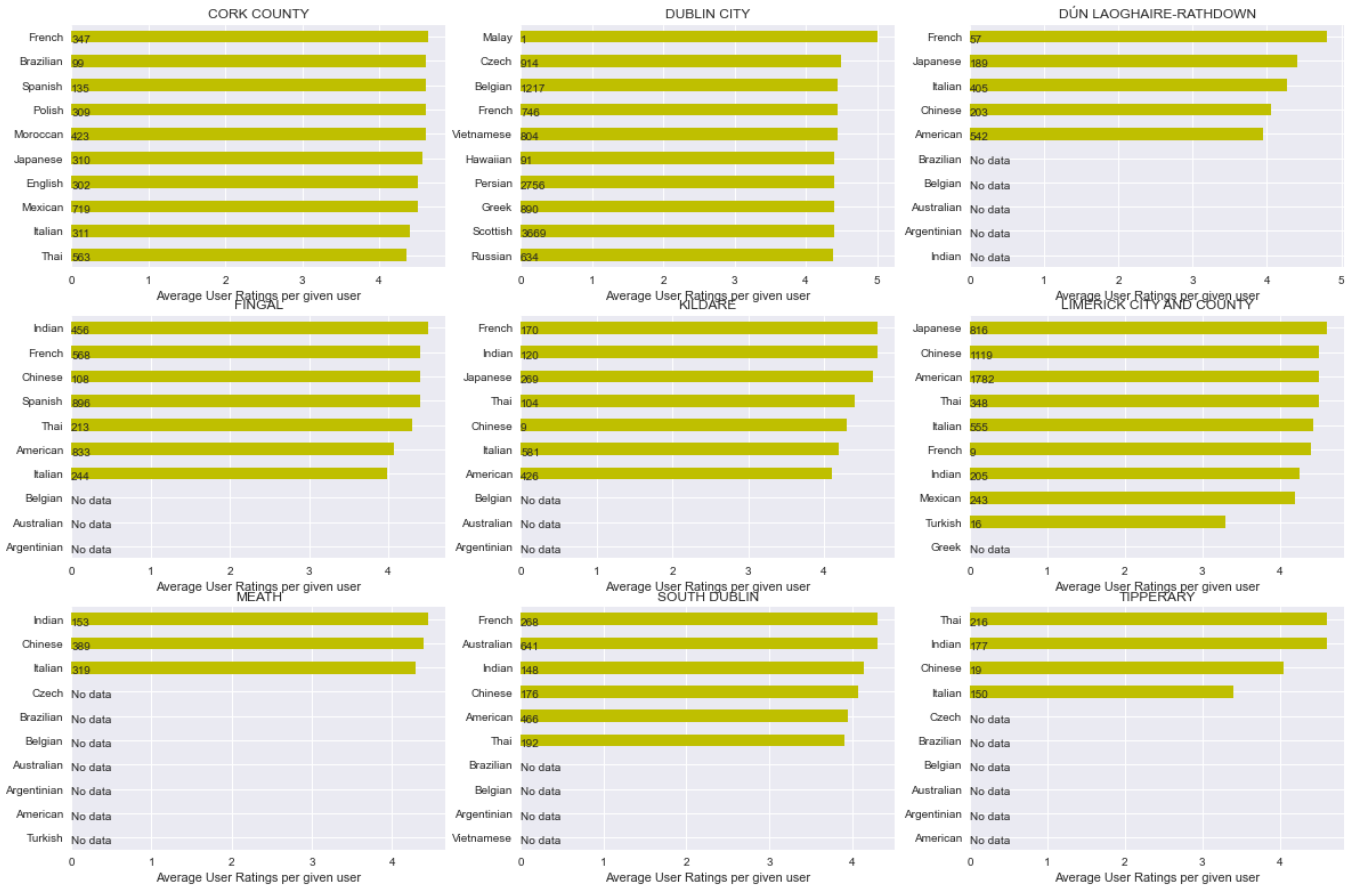
The first analysis, the **top 10 cuisines for each administrative area**, showed that in most of the areas, the Italian was the winner. However, in some area like South Dublin, Galway and Tipperary Chinese is the winner. Indian however is within top 5 popular cuisines with it being top choice of cuisine in Meath.

Based on number of restaurant serving particular cuisines, Top 10 cuisines in each administrative area are,



Thou the restaurants data for which user ratings were available were few, based on the data collected,

Popular 10 cuisine based on average restaurant ratings serving particular cuisine in particular administrative areas are:



In the above plot you can clearly see that there is data missing for administrative area Galway and that the average ratings per cuisine is also fluctuating across different administrative area.

But in Meath, the top three cuisines still remain the same based on user ratings too.

Which county is preferable to open their first restaurant ?

Based on the popularity above, Meath County seems most preferable administrative area to open new Indian Restaurant.

But to explore more, I next went on to merge population dataset with restaurants summary and restaurants details for restaurants serving Indian Cuisine to test my assumption that the more the Asian Population, the more Indian Restaurants, there is a higher positive correlation between the two.

```
data_for_analysis.head()
```

	Administrative_Area	Restaurant_Data_Available	Cuisine	Number of restaurants	ADMINISTRATIVE_AREA	ASIAN_POPULATION	TOTAL_POPULATION
0	CORK COUNTY	9	Indian	10	CORK COUNTY	4436	412826
1	DUBLIN CITY	38	Indian	39	DUBLIN CITY	20172	537190
2	DÚN LAOGHAIRE-RATHDOWN	7	Indian	8	DÚN LAOGHAIRE-RATHDOWN	7571	213519
3	FINGAL	3	Indian	4	FINGAL	11514	292327
4	GALWAY COUNTY	0	Indian	1	GALWAY COUNTY	1591	177215

Following was the result obtained.

	Restaurant_Data_Available	Number of restaurants	ASIAN_POPULATION	TOTAL_POPULATION
Restaurant_Data_Available	1.000000	1.000000	0.765533	0.841339
Number of restaurants	1.000000	1.000000	0.765533	0.841339
ASIAN_POPULATION	0.765533	0.765533	1.000000	0.788330
TOTAL_POPULATION	0.841339	0.841339	0.788330	1.000000

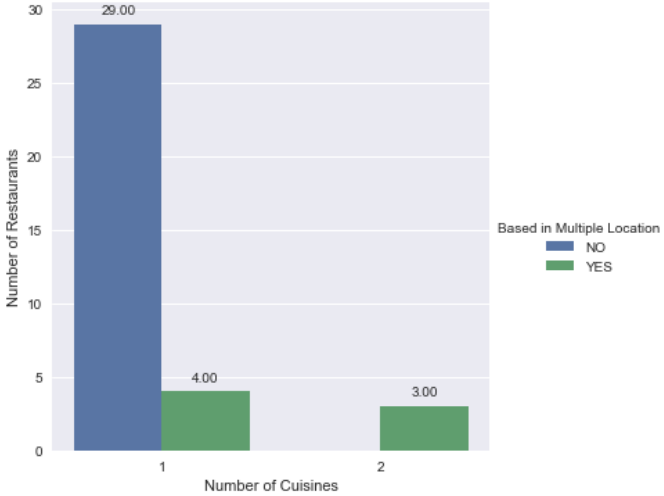
The data frame above suggested that there is a higher positive correlation between total population and number of restaurants serving Indian than Asian population and number of restaurants serving Indian.

And based on above analysis, the top two administrative areas for openings new restaurants should be Dublin City and County Dublin.

Are there existing Indian Restaurants chains in Ireland if any?

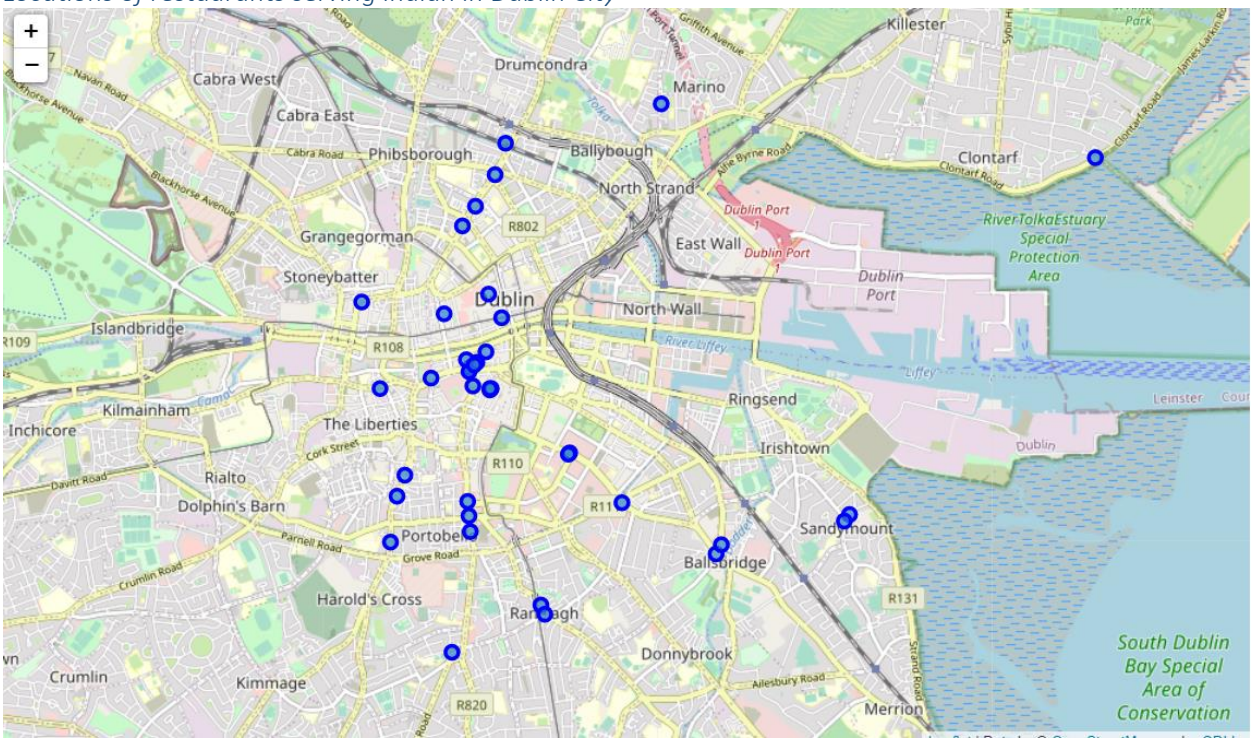
With the above, for final parts of analysis, I only looked at two administrative Areas Dublin City and County Cork to analyze the restaurants there. And based on further data transformations and aggregation, the following was obtained.

Summary of restaurants serving Indian in Dublin City

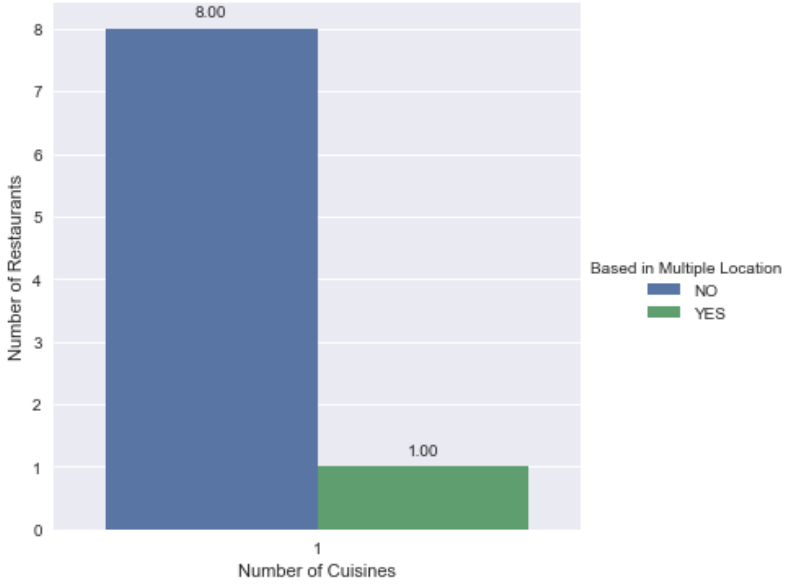


There are three Indian Restaurants in Dublin City based in multiple locations. They are Bombay Pantry, Jewel in the Crown and Indie Spice, and these could be analyzed further.

Locations of restaurants serving Indian in Dublin City

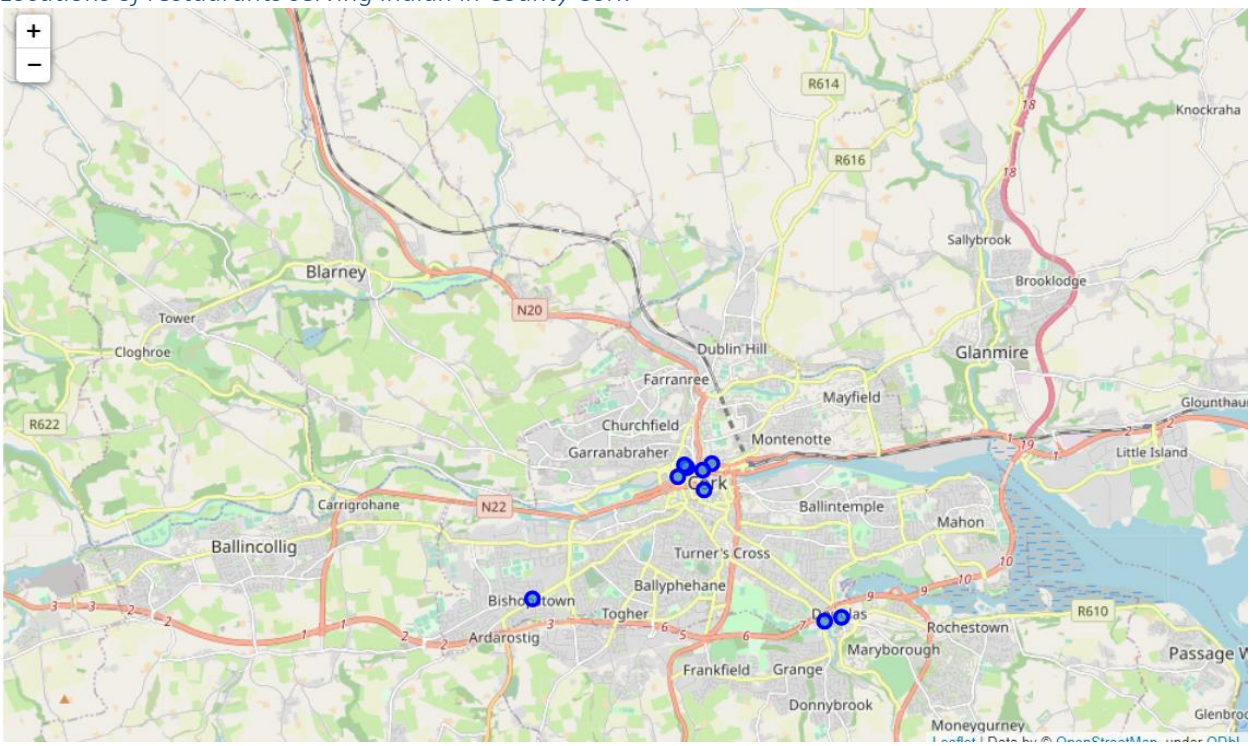


Summary of restaurants serving Indian in County Cork



There are no restaurant chains in county cork that serve other administrative area too.

Locations of restaurants serving Indian in County Cork



Future Recommendation

The results obtained in this project are very interesting, especially for those interested further analyzing cuisines and restaurants across Ireland. However, for doing a deeper research, probably all administrative area should be explored.

Another recommendation would be to consider the ratings and price of the restaurants, using both Foursquare API and google API. This will probably increase the number of datasets we will be looking at.

Conclusion

In this project, I have used data from multiple sources including Foursquare API and Google Places API to perform an analysis of cuisines popularity along selected administrative areas across Republic of Ireland. The reason for basing our analysis on administrative area was because eventually it is each administrative council that that must be reached out establish a new business in the area. With geographical heat map and bar graphs, the results were presented in an attractive and easy to understand way, allowing to better understand results at each stages of our analysis. This project showcases how useful a geographical data API like Foursquare and Google places API can used be to carry studies about very diverse topics, and also how powerful are the Python-based open-source tools.