CAPSTONE PROJECT - THE BATTLE OF NEIGHBORHOODS (WEEK 2)

Market Entry Research

MARKET ENTRY RESEARCH FOR A NEW RESTAURANT IN DUBLIN, IRELAND

Market Entry research project for XYZ Pvt. Ltd prior to start their chain of restaurants in Dublin, Ireland

Benny Jose
Coursera Capstone Project

Table of Contents

1. Introduction 2
1.1 Background 2
1.2 Business Problem 2
2. Data 2
2.1 Data sources and Acquisition 2
2.2 How Data Intend to for the analysis 2
3. Exploratory Analysis 3
3.1 Identification and formatting data within the scope of analysis. 3
3.2 Remove rows to clean data without impacting the quantitative analysis of business problem 4
Some records contain 'NaN' values and was removed. 4
3.3 Geospatial Map for better planning of Restaurant location 4
3.4 Extract and filter venues with restaurants 4
4. Results Section 5
4.1 Geospatial Results of Dublin 5
4.2 Restaurants population/Saturation at each postal district 5
1.1 Generate graphs representing types of restaurants in Dublin 6
5. Discussion Section 7
6. Conclusion 8

1. Introduction

1.1 Background

A restaurant chain company XYZ Pvt. Ltd planning to open their outlets in Europe, Ireland. They selected Dublin city as their target market for starting the business. Since they do not have any business in Dublin, they want to do a market analysis prior to market entry.

1.2 Business Problem

XYZ Pvt Ltd. have expertise in operating various types of restaurants by giving more specialization to the taste of community where they start the restaurant. Here the business problem is to identify and analyses on below points so that they can make better decisions.

1) What are the types of restaurants currently available in Dublin?
   Eg: How many America, Indian, Italian restaurants in each postal district in Dublin
2) Number of each type of restaurants in the locality.
3) Identify areas with less saturation of restaurants to minimize competition.

2. Data

2.1 Data sources and Acquisition

As part geographical special and statistical analysis following data sources used.

Autoaddress:

Auto address is an address management consultancy company in Ireland. They have all the postal district addresses and other location details of entire Ireland.
Postal district details of Ireland can be extracted from developer center section of resources using below url.
https://www.autoaddress.ie/support/developer-centre/resources/routing-key-boundaries
Used python packages to get coordinates of Dublin postal districts.

Foursquare:

Using coordinates as input extract all the venues from Foursquare and filter to shortlist names and types of Restaurants in Dublin.

2.2 How Data Intend to for the analysis

Autoaddress will return all the postal district in Ireland along with post code key. Postcode key can be ignored. All postal district in Dublin named as Dublin 1, Dublin 2 , up to Dublin 24. With this info we can get full list of postal districts. Using Python packages coordinates of each Dublin district can be retrieved and will be stored as a panda data frame.

The above data will feed to Foursquare to get all the venues in Dublin in each postal district.
A filtering will be required to identify all the restaurants and the type of restaurants from the list.
Once we have this information grouping and categorization can be done. With these details a market research can be conducted.
These details will contribute to Market entry discussion and to make a decision on area and type of restaurant sot start.

3. Exploratory Analysis

During exploratory analysis of available data following steps are done.

3.1 Identification and formatting data within the scope of analysis.

Postal district data downloaded from https://www.autoaddress.ie/support/developer-centre/resources/routing-key-boundaries

BeautifulSoup scrapping library in panda used to get data from URL and to parse.

```
In [3]: from bs4 import BeautifulSoup # scraping library

In [4]: url = "https://www.autoaddress.ie/support/developer-centre/resources/routing-key-boundaries"

        text_result = requests.get(url).text
        html_parsed_result = BeautifulSoup(text_result, 'html.parser')

        neightborhood_info_table = html_parsed_result.find('table', class_ = 'three')
        neightborhood_rows = neightborhood_info_table.find_all('tr')

        # extract the info ('Postcode', 'District Post code')
        neightborhood_info = []
        for row in neightborhood_rows:
            info = row.text.split('\n')[1:-1] # remove empty str (first and last items)
            neightborhood_info.append(info)

        neightborhood_info[0:10]
        # print('Data downloaded!')

Out[4]: [['ROUTING KEY', 'DESCRIPTOR'],
         ['Y14', 'ARKLOW'],
         ['A84', 'ASHBOURNE'],
         ['H65', 'ATHENRY'],
         ['N37', 'ATHLONE'],
         ['R14', 'ATHY'],
         ['K32', 'BALBRIGGAN'],
         ['F26', 'BALLINA'],
         ['H53', 'BALLINASLOE'],
         ['P31', 'BALLINCOLLIG']]
```

To get longitude and latitude of postal district *geopy.extra.rate_limiter* used

```
In [10]: geolocator = Nominatim(user_agent="dublin_explorer")
         from geopy.extra.rate_limiter import RateLimiter
         # 1 - conveneint function to delay between geocoding calls
         geocode = RateLimiter(geolocator.geocode, min_delay_seconds=1)
         # 2- - create location column
         cleandf3['postal_dist_coords'] = cleandf3['Neighbourhood'].apply(geocode)

In [11]: # 3 - create longitude, laatitude and altitude from location column (returns tuple)
         cleandf3['point'] = cleandf3['postal_dist_coords'].apply(lambda loc: tuple(loc.point) if loc else None)

In [12]: # 4 - split point column into Latitude, longitude and altitude columns
         cleandf3[['latitude', 'longitude', 'altitude']] = pd.DataFrame(cleandf3['point'].tolist(), index=cleandf3.index)

In [13]: cleandf3.head(100)
```

Out[13]:

	Key	Neighbourhood	postal_dist_coords	point	latitude	longitude	altitude
51	D01	DUBLIN 1	(North City ED, Dublin 1, Dublin, County Dubli...	(53.3524881, -6.25664568972183, 0.0)	53.352488	-6.256646	0.0
52	D02	DUBLIN 2	(Dublin 2, Dublin, County Dublin, Leinster, Ir...	(53.33894015, -6.25271282175961, 0.0)	53.338940	-6.252713	0.0
53	D03	DUBLIN 3	(Clontarf East B ED, Dublin 3, Dublin, County ...	(53.3612231, -6.18546680600004, 0.0)	53.361223	-6.185467	0.0
54	D04	DUBLIN 4	(Pembroke East E ED, Dublin 4, Dublin, County ...	(53.32746935, -6.22753692446213, 0.0)	53.327469	-6.227537	0.0
55	D05	DUBLIN 5	(Edenmore ED, Dublin 5, Dublin, County Dublin,...	(53.3834538, -6.18192324547357, 0.0)	53.383454	-6.181923	0.0
56	D06	DUBLIN 6	(Rathmines West C ED, Dublin 6, Dublin, County...	(53.3176976, -6.25952513256976, 0.0)	53.317698	-6.259525	0.0
57	D6W	DUBLIN 6W	(Terenure C ED, Dublin 6W, Dublin, County Dubl...	(53.30928205, -6.29943489174728, 0.0)	53.309282	-6.299435	0.0
58	D07	DUBLIN 7	(Cabra East C ED, Dublin 7, Dublin, County Dub...	(53.3605505, -6.28446550135915, 0.0)	53.360551	-6.284466	0.0
59	D08	DUBLIN 8	(Phoenix Park ED, Dublin 8, Dublin, County Dub...	(53.3505559, -6.3205769204553, 0.0)	53.350556	-6.320577	0.0
60	D09	DUBLIN 9	(Whitehall B ED, Dublin 9, Dublin, County Dubl...	(53.3860497, -6.24557708531776, 0.0)	53.386050	-6.245577	0.0

3.2 Remove rows to clean data without impacting the quantitative analysis of business problem

Some records contain 'NaN' values and was removed.

```
In [14]: clean_dub_data = cleandf3[cleandf3.Key != 'T34']

In [15]: clean_dub_data.head(200)
```

Out[15]:

	Key	Neighbourhood	postal_dist_coords	point	latitude	longitude	altitude
51	D01	DUBLIN 1	(North City ED, Dublin 1, Dublin, County Dubli...	(53.3524881, -6.25664568972183, 0.0)	53.352488	-6.256646	0.0

3.3 Geospatial Map for better planning of Restaurant location.

Geospatial map of Dublin generated for identification of locations for restaurant. This will help on top of other statistical data during market research.

Create a map of Dublin with neighborhoods superimposed on top.

```
In [16]: address = 'Dublin, Ireland'

geolocator = Nominatim(user_agent="dbln_explorer")
dblocation = geolocator.geocode(address)
dblatitude = dblocation.latitude
dblongitude = dblocation.longitude
print('The geograpical coordinate of Dublin are {}, {}.'.format(dblatitude, dblongitude))
```

The geograpical coordinate of Dublin are 53.3497645, -6.2602732.

```
In [17]: # create map of New York using latitude and longitude values
map_dublin = folium.Map(location=[dblatitude, dblongitude], zoom_start=10)

# add markers to map
for latitude, longitude, Key, Neighbourhood in zip(clean_dub_data['latitude'], clean_dub_data['longitude'], clean_dub_data['Key'], clean_dub_data['Neighbourhood']):
    label = '{}, {}'.format(Neighbourhood, Key)
    label = folium.Popup(label, parse_html=True)
    folium.CircleMarker(
        [latitude, longitude],
        radius=5,
        popup=label,
        color='blue',
        fill=True,
        fill_color='#3186cc',
        fill_opacity=0.7,
        parse_html=False).add_to(map_dublin)

map_dublin
```

Out[17]:

3.4 Extract and filter venues with restaurants

For this iterate foursquare API call for each postal district. Defined getNearbyVenues function to enable this. Once extracted did a search with keyword restaurants to filter rows with restaurants.

```
dublin_venues = dublin_venues[dublin_venues['Venue Category'].str.contains('Restaurant')]
dublin_venues.head()
```

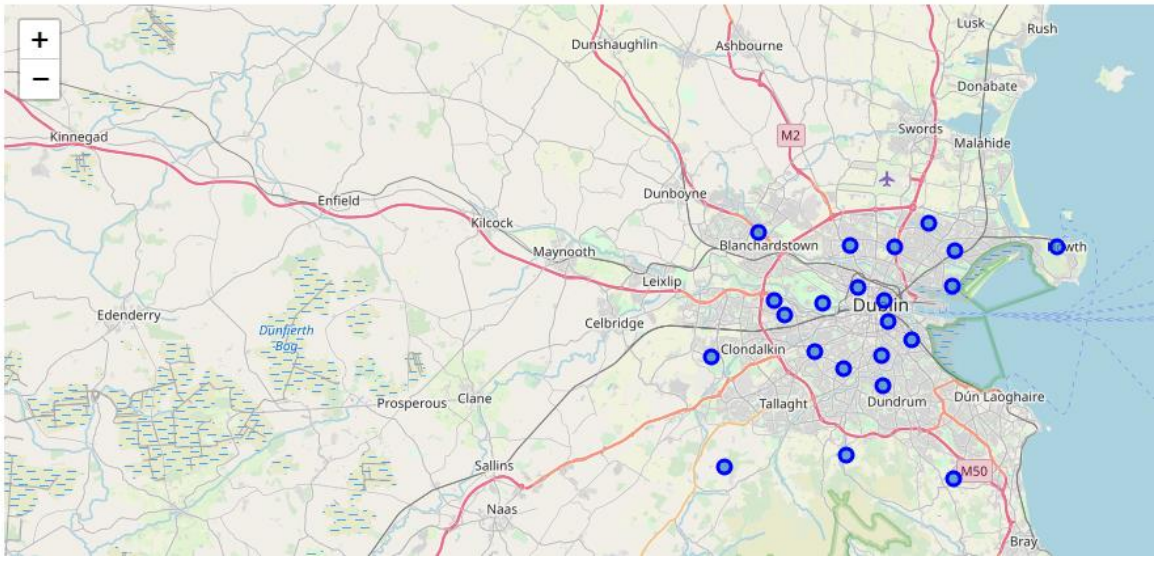
3]:

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
2	DUBLIN 1	53.352488	-6.256646	The Hop House (Kim Chi)	53.352981	-6.260772	Korean Restaurant
7	DUBLIN 1	53.352488	-6.256646	El Grito Mexican Taqueria	53.357390	-6.256618	Mexican Restaurant
16	DUBLIN 1	53.352488	-6.256646	The Winding Stair	53.346596	-6.263784	Restaurant
24	DUBLIN 1	53.352488	-6.256646	Il Vicoletto	53.344632	-6.263744	Italian Restaurant
29	DUBLIN 1	53.352488	-6.256646	The Port House Pintxo	53.345162	-6.264965	Tapas Restaurant

4. Results Section

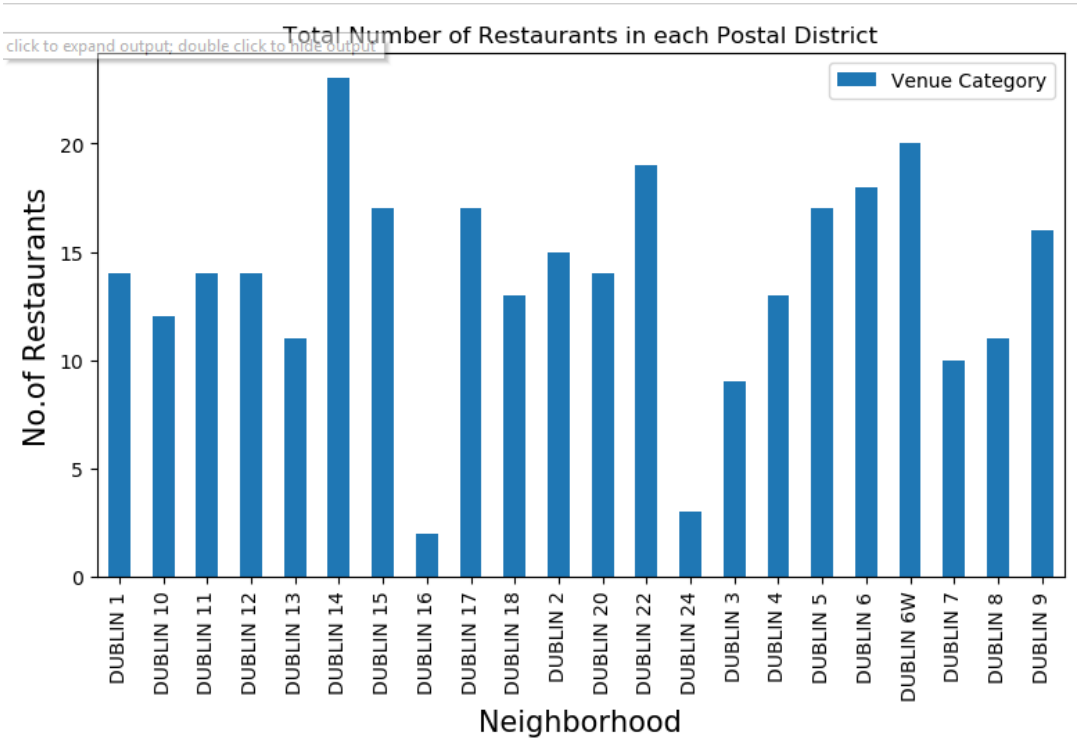
4.1 Geospatial Results of Dublin

Out[17]:



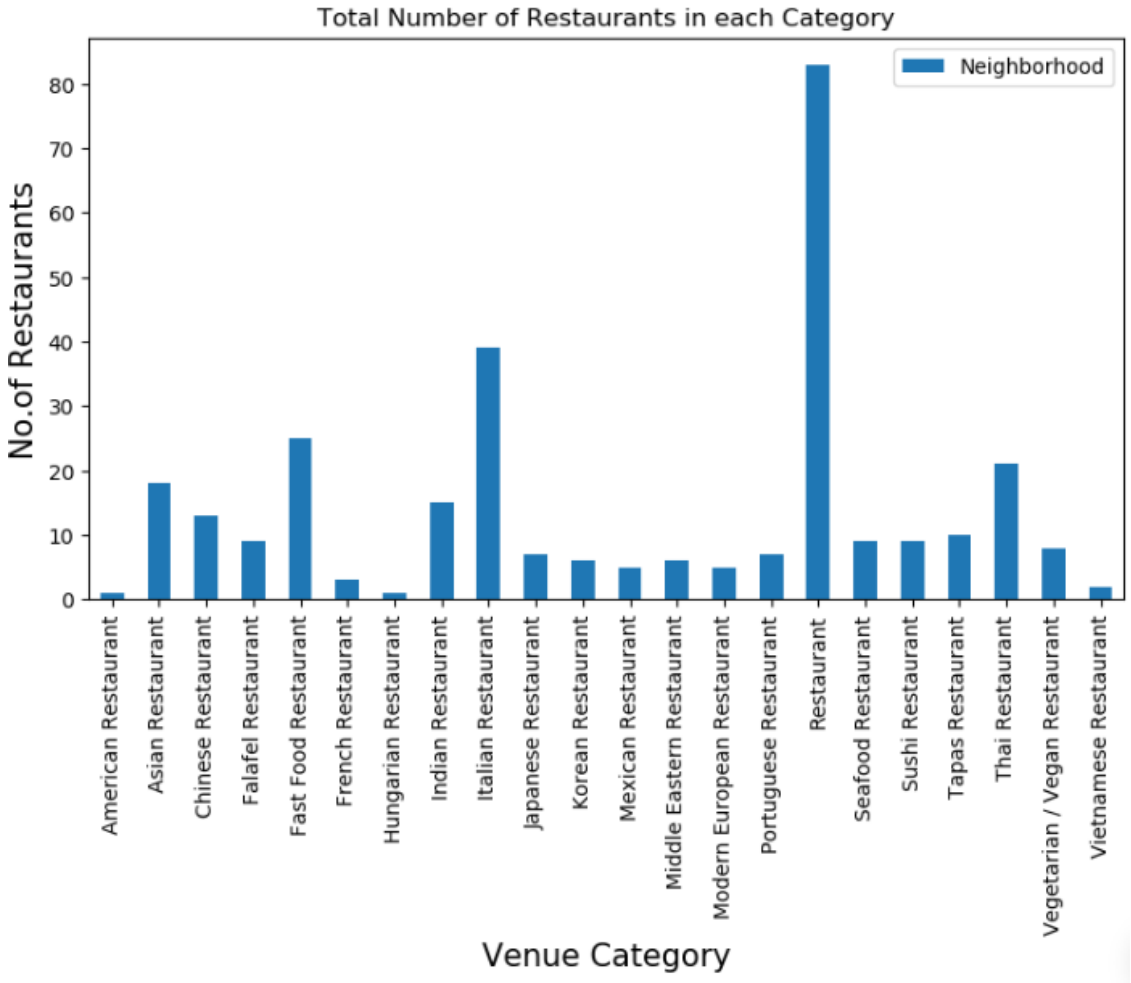
4.2 Restaurants population/Saturation at each postal district

Now we need to do categorical analysis to identify number of restaurants in each postal district. This will help to identify saturation in each area which will help to decide about locality of new restaurants. Results from data frame categorized based on postal district for this.



1.1 Generate graphs representing types of restaurants in Dublin

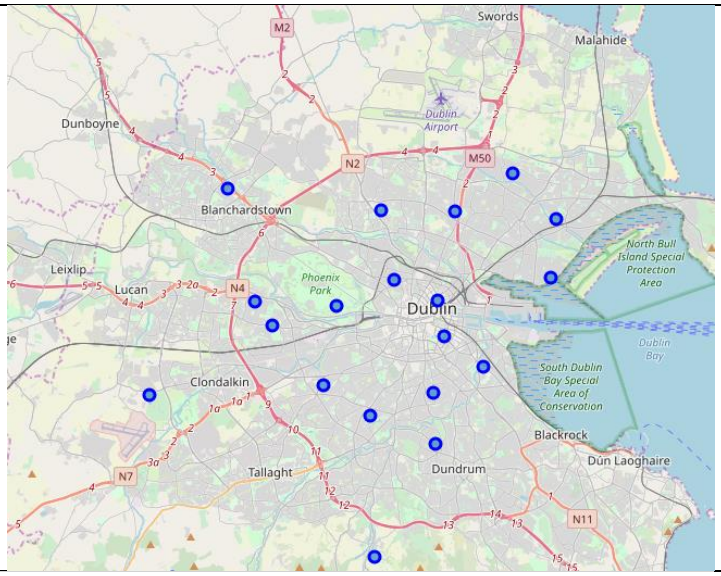
As part of analysis we need types of restaurants in each sector of Dublin so that it will be easier to identify the type of restaurant with minimum competition.

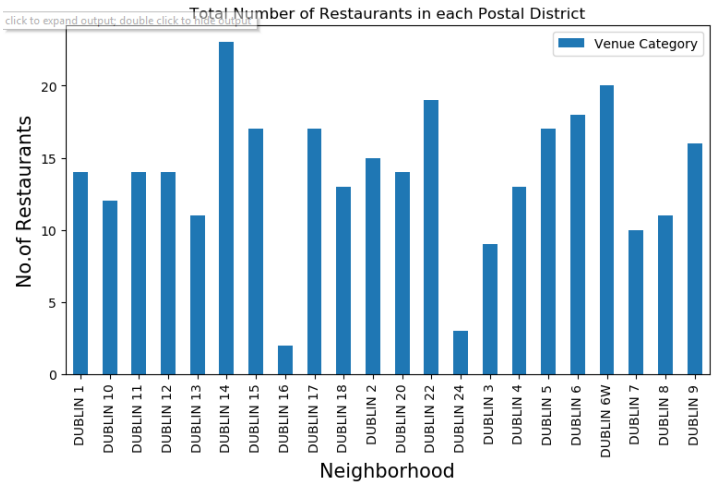


5. Discussion Section

As per the requirement from XYZ Pvt. Ltd they want to identify right location to start their restaurant. They also want to know what type of restaurant will be suitable for each location. From the results posted earlier to reach in a right conclusion we need to analyze the location with its sectorization.

Postal districts with lesser number are closer to the town center and with high population or with more relevant location for restaurants.





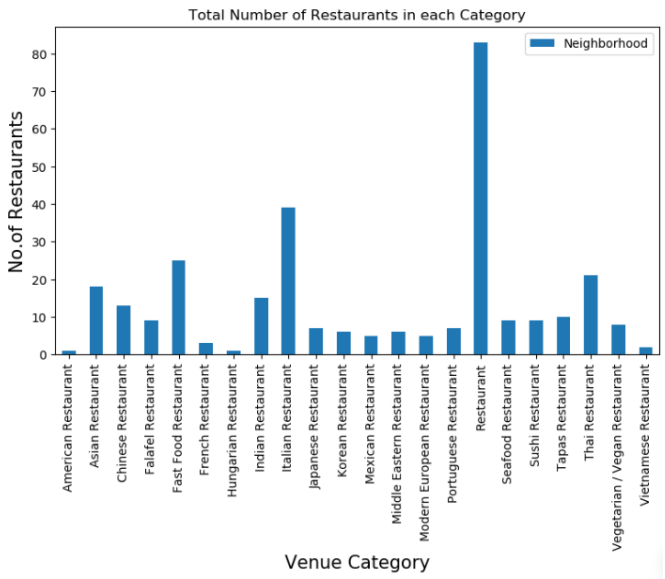
We need to compare both images together. If we consider first 10 postal districts close to city center Dublin 3 has the lowest number of restaurants closest to city center.

Across all postal districts, Dublin 16 got a smaller number of restaurants. But it may not be advisable to start a restaurant there as it away from city center.

As we move away from city center the density of offices and schools will be lesser and will have less scope for restaurants.

Though Dublin 14 is much away from city center this postal district got maximum number of restaurants. This could be due to high number of offices towards south side of Dublin.

We can have more understanding on ideal location for a restaurant only after the analysis of type of restaurant in that area. This we will in the next section.



As you can see on the type of the restaurant, highest number of restaurants is general in nature. They don't specific to any type of special menu.

In other type of restaurants Italian restaurants are more. This indicates that Italian food is more popular in Dublin if we go with specific type of menu.

Conclusion on results will make in the next section

6. Conclusion

As per the results following are the conclusion made from the results. XYZ Pvt. Ltd will have optimized result from business if they enter in to market considering below points.

a. Dublin 14 and Dublin 3 are ideal locations with high potential.

b. Dublin 3 has high potential for general restaurants. Also, will have less competition and will attract general customers who don't a have any specific choice. Since it is close to city center customers probably look for general restaurants. So, opening a restaurant with general menu would be advisable for Dublin 3.

c. Dublin 14 also has high potential, but this location is away from City center. So mostly this area will have regular customers than one off visitors. So specific menus will have more impact. As you can see Italian is the most favorable menu across all postal district. So, opening an Italian restaurant at Dublin 14 will be advisable.