Capstone Project

IBM Data Science Coursera Module

Opening a Coffee Shop in Cork City, Ireland

Table of Contents

1. Introduction

The report is part of the Coursera Capstone Module within the IBM Data Science Professional Certificate. As part of this project, I intend to leverage Foursquare location data to solve a problem for a particular audience.

1.1 Business Problem

The objective is to find a suitable location to open a coffee shop within Cork City which will be useful for potential investors. Using Foursquare data, I will explore the neighbourhoods in Cork City and determine where is the best place to set up a new restaurant.

The population of Cork is growing at a rapid pace, and following Brexit, Cork will be the second largest English-speaking city in Europe (after Dublin). Opening a restaurant in the city would an attractive long-term investment due to the rapidly increasing population. However, an investor would need to consider a number of factors prior to opening a coffee shop e.g. city centre location, competing coffee shops / restaurants nearby and parking facilities.

1.3 Cork City

The city of Cork covers a total surface area of 37.3 square kilometers - in the city proper. The population density is approximately 3,300 individuals per square kilometer in the city. The population of Cork is quite substantial when compared to other areas of Ireland. Cork city is the only city in the county and the second most populous city in the Republic of Ireland, with a population of 125,657 according to the 2016 census.

2. Data Description

As per the Coursera assignment, the recommended Data Source will be Foursquare Location Data for geographical data.

2.1 Data Source:

In order to support my analysis, my main data sources will be: 1) Foursquare API Location Data for restaurants/coffee shops & 2) Data.Gov.ie for car parking data.

This project will focus only on the Cork City (as opposed to Cork County/Region). I will collect the relevant location data from Foursquare and and additional data from Data.Gov.ie. I will obtain the recommended venues within 3km radius of the centre of Cork city neighbourhood.

3. Methodology

The key steps will be as follows:

1. Import necessary libraries to Python

2. Connect to the Foursquare API using my Client ID

3. Obtain coordinates of Cork City

4. Search for Coffee Shops in Cork City and examine/visualise locations

5. Search for Restaurants in Cork City and examine/visualise locations

6. Investigate nearby Parking Spaces using publicly available Irish Government data

7. Examine results and choose appropriate location to open coffee shop

To begin, I shall import the libraries to Python

1. Import necessary libraries

```
import requests # library to handle requests
import pandas as pd # library for data analsysis
import numpy as np # library to handle data in a vectorized manner
import random # library for random number generation

!conda install -c conda-forge geopy --yes
from geopy.geocoders import Nominatim # module to convert an address into latitude and longitude values

# libraries for displaying images
from IPython.display import Image
from IPython.core.display import HTML

# tranforming json file into a pandas dataframe library
from pandas.io.json import json_normalize

!conda install -c conda-forge folium=0.5.0 --yes
import folium # plotting library

print('Folium installed')
print('Libraries imported.')
```

```
Collecting package metadata: done
Solving environment: done
```

2. Connect to Foursquare API

```
CLIENT_ID = 'H3ZBNJQJOUYUV5YF3PKK1JNSMJ5SEFKV3TDSBSZGTYT34JGX' # your Foursquare ID
CLIENT_SECRET = 'U5ARHRIRB2OSHZZTQ353HV1C0S0DFUDC2CY2WM2PBSFLBI4V' # your Foursquare Secret
VERSION = '20180604'
LIMIT = 30
print('Your credentails:')
print('CLIENT_ID: ' + CLIENT_ID)
print('CLIENT_SECRET:' + CLIENT_SECRET)
```

```
Your credentails:
CLIENT_ID: H3ZBNJQJOUYUV5YF3PKK1JNSMJ5SEFKV3TDSBSZGTYT34JGX
CLIENT_SECRET:U5ARHRIRB2OSHZZTQ353HV1C0S0DFUDC2CY2WM2PBSFLBI4V
```

3. Obtain coordinates of Cork City

```
address = 'Cork City'

geolocator = Nominatim(user_agent="foursquare_agent")
location = geolocator.geocode(address)
latitude = location.latitude
longitude = location.longitude
print(latitude, longitude)
```

```
51.8979282 -8.4705806
```

4. Search for Coffee Shops within 3km of Cork City Centre ¶

```
search_query = 'Coffee'
radius = 3000
print(search_query + ' .... OK!')
```

```
Coffee .... OK!
```

Corresponding URL must be defined.

```
url = 'https://api.foursquare.com/v2/venues/search?client_id={}&client_secret={}&ll={},{}&v={}&query={}&
url
```

```
'https://api.foursquare.com/v2/venues/search?client_id=H3ZBNJQJOUYUV5YF3PKK1JNSMJ5SEFKV3TDSBSZGTYT34JGX
&client_secret=U5ARHRIRB2OSHZZTQ353HV1C0S0DFUDC2CY2WM2PBSFLBI4V&ll=51.8979282,-8.4705806&v=20180604&que
ry=Coffee&radius=3000&limit=30'
```

GET Request must be sent and results examined.

```
results = requests.get(url).json()
results
```

Get relevant part of JSON and transform it into a pandas dataframe. For purposes of this report, the coffee data which is the key dataframe will be called "dataframe". ¶

[7]:
```
# assign relevant part of JSON to venues
venues = results['response']['venues']

# tranform venues into a dataframe
dataframe = json_normalize(venues)
dataframe.head()
```

Information of interest to be defined and dataframe filtered.

]:
```
# keep only columns that include venue name, and anything that is associated with location
filtered_columns = ['name', 'categories'] + [col for col in dataframe.columns if col.startswith('locatic
dataframe_filtered = dataframe.loc[:, filtered_columns]

# function that extracts the category of the venue
def get_category_type(row):
    try:
        categories_list = row['categories']
    except:
        categories_list = row['venue.categories']

    if len(categories_list) == 0:
        return None
    else:
        return categories_list[0]['name']

# filter the category for each row
dataframe_filtered['categories'] = dataframe_filtered.apply(get_category_type, axis=1)

# clean column names by keeping only last term
dataframe_filtered.columns = [column.split('.')[-1] for column in dataframe_filtered.columns]

dataframe_filtered
```

Next, I will visualise the coffee shops in the dataframe within Cork City.

```
]: dataframe_filtered.name
```

```
]: 0                                    Nectar Coffee
   1                                     Costa Coffee
   2                                  Coffee Central
   3                                  Coffee Station
   4                                Priory Coffee Co
   5                                     Costa Coffee
   6                          The Bookshelf Coffee Shop
   7                                  Mahers Coffee
   8                                     Costa Coffee
   9                               Three Fools Coffee
   10                             Cork Coffee Roasters
   11                                       Coffee Pod
   12                                    Costa Coffee
   13                             Cork Coffee Roasters
   14                             Dukes Coffee Company
   15                                    Costa Coffee
   16                              Soma Coffee Company
   17                            Gloria Jean's Coffees
   18                         The bagel BAR Coffee House
   19                                 Coffee Station
   20                               Warren Allen Coffee
   21                                    Candy & Coffee
```

```
: venues_map = folium.Map(location=[latitude, longitude], zoom_start=13) # generate map centred around the

# add a red circle marker to represent Cork City
folium.features.CircleMarker(
    [latitude, longitude],
    radius=10,
    color='red',
    popup='Cork City',
    fill = True,
    fill_color = 'red',
    fill_opacity = 0.6
).add_to(venues_map)

# add Cofee Shops as blue circle markers
for lat, lng, label in zip(dataframe_filtered.lat, dataframe_filtered.lng, dataframe_filtered.categories
    folium.features.CircleMarker(
        [lat, lng],
        radius=5,
        color='blue',
        popup=label,
        fill = True,
        fill_color='blue',
        fill_opacity=0.6
    ).add_to(venues_map)

# display map
venues_map
```

Image: Coffee Shops in Cork City

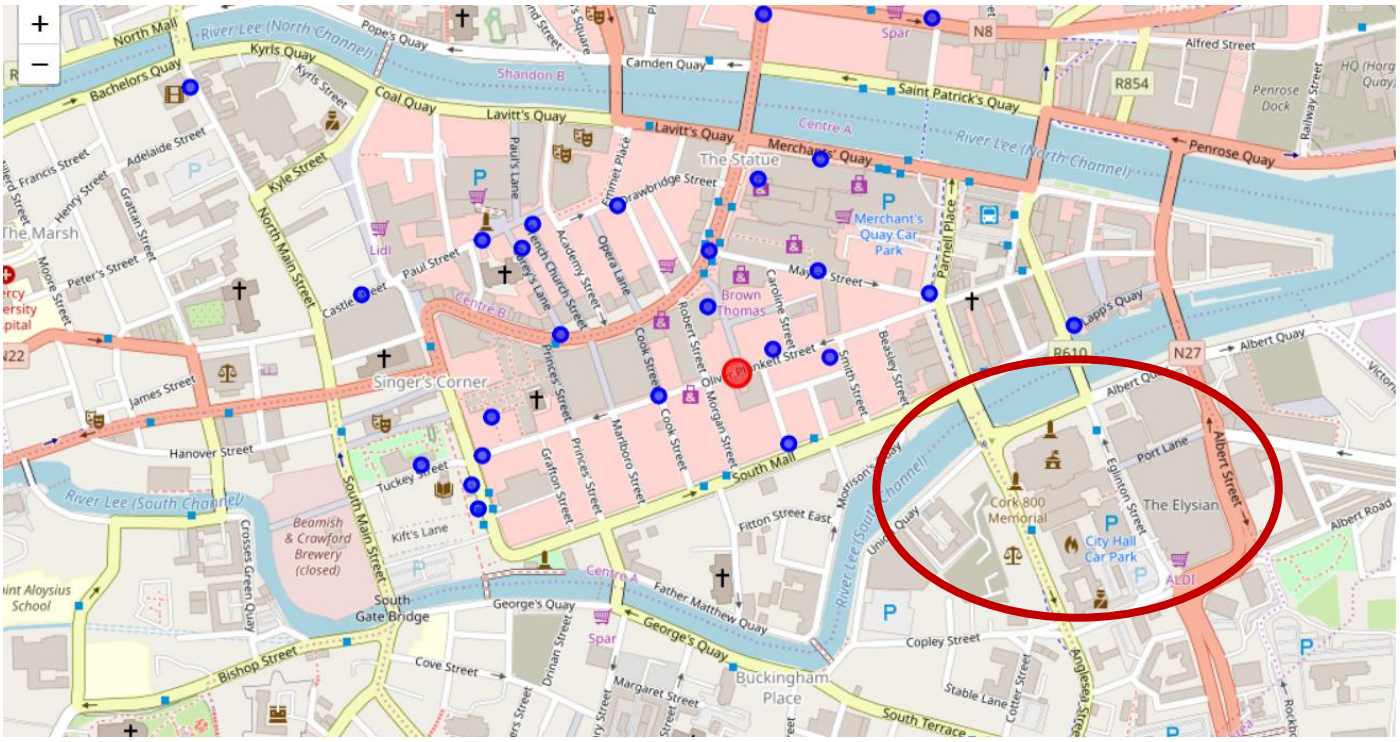
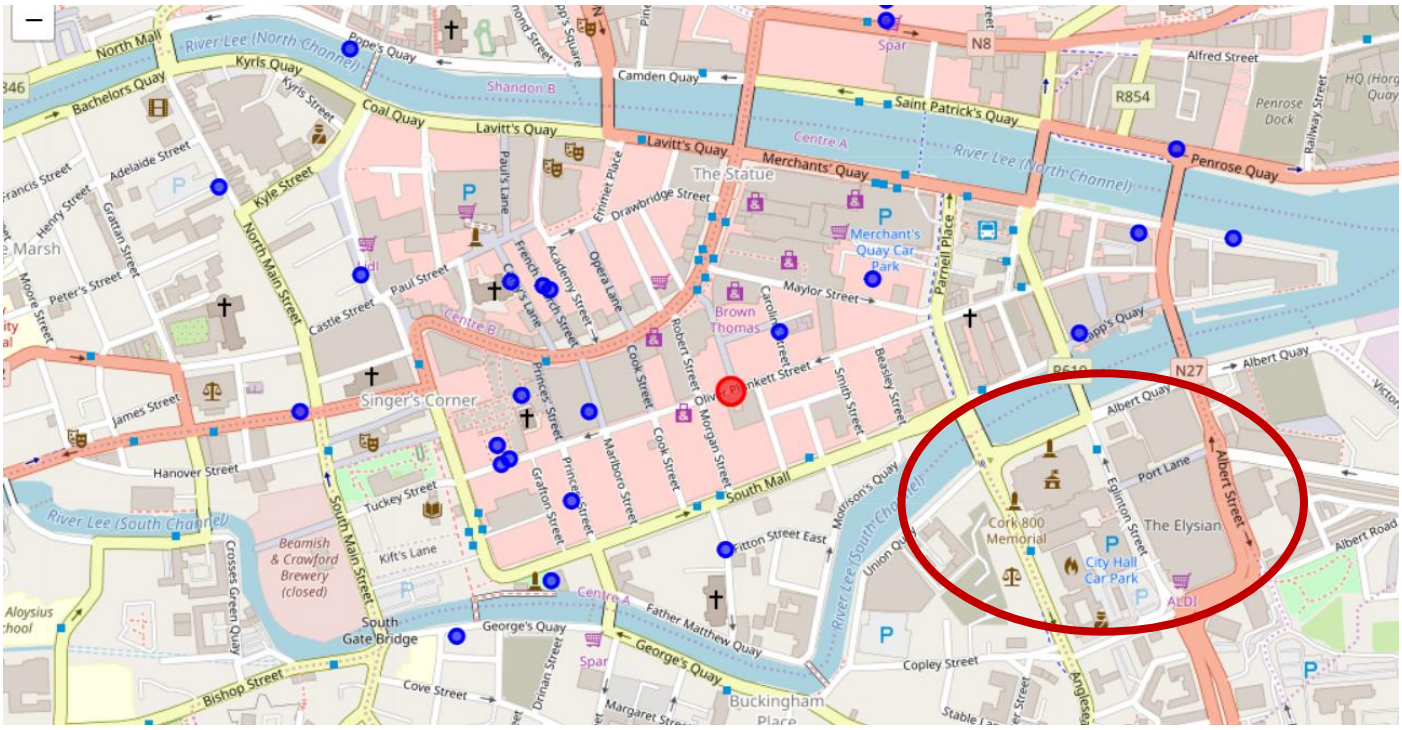


Image: Restaurants in Cork City



5. Investigate nearby Parking Spaces using publicly available Irish Government data

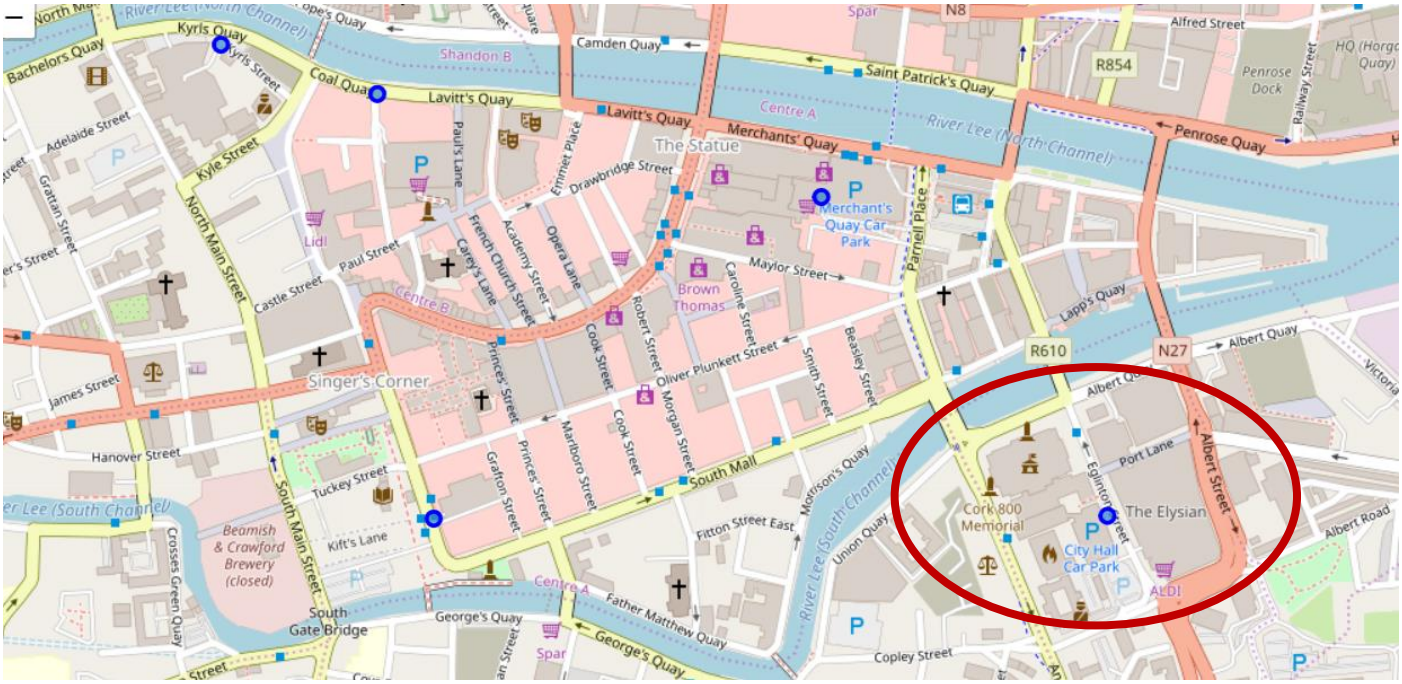
```
df_parking = pd.read_csv("http://data.corkcity.ie/datastore/dump/6cc1028e-7388-4bc5-95b7-667a59aa76dc")
```

```
map_cork = folium.Map(location=[latitude, longitude], zoom_start=13)

# add markers to map
for lat, lng, name, space, freespace in zip(df_parking['latitude'], df_parking['longitude'], df_parking[
    label = '{}, {}'.format(space, freespace)
    label = folium.Popup(label, parse_html=True)
    folium.CircleMarker(
        [lat, lng],
        radius=5,
        popup=label,
        color='blue',
        fill=True,
        fill_color='#3186cc',
        fill_opacity=0.7,
        parse_html=False).add_to(map_cork)

map_cork
```

Image: Parking in Cork City



Results & Conclusion

Albert Quay appears to be the most suitable location to open a coffee shop within Cork City for potential investors. Using Foursquare data, I explore the neighbourhoods in Cork City and determined it is an appropriate place to set up a new coffee shop.

1) Very few coffee shops nearby
2) No competing restaurants nearby
3) There very few cark parking locations in the city. This location is right next to one of the few parking spots.
4) A number of offices setting up near Albert Quay and Cork city population is increasing.