The Battle of Neighborhoods

Where to open a restaurant in Glasgow

Applied Data Science Capstone – Report

Submission by : Kalirajan Natarajan

Introduction

Glasgow is the largest and most populous city in Scotland, United Kingdom.

It is known for its vibrant nature. It is always welcoming new businesses. As it is touted as a business capital of Scotland, there is no shortage of tourists and business visitors.

Target Audience

Target audience for this report are the restaurateurs, who are looking for a suitable place to open a restaurant in Glasgow wards

Not all the constituencies are having equal distribution of restaurants. So this is an attempt to find out how where the restaurants are available in Glasgow constituencies, what is the ratio of restaurants to population and suggesting which ward / constituency would be better to open the restaurant in Glasgow.

Data section

In order for the study we need data related to the various neighbourhoods available in Glasgow. While searching for the data, I happened to stumble upon the following website.

https://www.doogal.co.uk/PostcodeDownloads.php

This website has a list of postcodes of entire UK region. Data is available in CSV is an additional benefit.

I have chosen a CSV file containing all area information for Scotland. I have uploaded the CSV file to IBM Cloud's Object storage as well. It is available in the following link.

https://kali-capstone-assignment.s3.eu-gb.cloud-object-storage.appdomain.cloud/scotland.csv

Here is the snapshot of how the CSV file looks like

```
Postcode,In Use?,Latitude,Longitude,Easting,Northing,Grid Ref,County,District,Ward,District Code,Ward Code,Country,County Code,Constituency,Introduced,Terminated,Parish,Nation
AB1 0AA,No,57.101474,-2.242851,385386,801193,NJ853011,"","Aberdeen City","Lower Deeside",S12000033,S13002843,Scotland,S99999999,"Aberdeen South",1980-01-01,1996-06-01,"","",,,
AB1 0AB,No,57.102554,-2.246308,385177,801314,NJ851013,"","Aberdeen City","Lower Deeside",S12000033,S13002843,Scotland,S99999999,"Aberdeen South",1980-01-01,1996-06-01,"","",,,
AB1 0AD,No,57.100556,-2.248342,385053,801092,NJ850010,"","Aberdeen City","Lower Deeside",S12000033,S13002843,Scotland,S99999999,"Aberdeen South",1980-01-01,1996-06-01,"","",,,
AB1 0AE,No,57.084444,-2.255708,384600,799300,NO845992,"","Aberdeenshire","North Kincardine",S12000034,S13002864,Scotland,S99999999,"West Aberdeenshire and Kincardine",1994-02-
AB1 0AF,No,57.096656,-2.258102,384460,800660,NJ844006,"","Aberdeen City","Lower Deeside",S12000033,S13002843,Scotland,S99999999,"Aberdeen South",1990-12-01,1992-07-01,"","",,,
AB1 0AG,No,57.097085,-2.267513,383890,800710,NJ838007,"","Aberdeen City","Lower Deeside",S12000033,S13002843,Scotland,S99999999,"Aberdeen South",1990-12-01,1992-07-01,"","",,,
AB1 0AJ,No,57.099011,-2.252854,384779,800921,NJ847009,"","Aberdeen City","Lower Deeside",S12000033,S13002843,Scotland,S99999999,"Aberdeen South",1980-01-01,1996-06-01,"","",,,
AB1 0AL,No,57.101765,-2.254688,384669,801228,NJ846012,"","Aberdeen City","Lower Deeside",S12000033,S13002843,Scotland,S99999999,"Aberdeen South",1980-01-01,1996-06-01,"","",,,
AB1 0AN,No,57.097553,-2.245483,385225,800757,NJ852007,"","Aberdeen City","Lower Deeside",S12000033,S13002843,Scotland,S99999999,"Aberdeen South",1980-01-01,1996-06-01,"","",,,
AB1 0AP,No,57.098244,-2.245768,385208,800834,NJ852008,"","Aberdeen City","Lower Deeside",S12000033,S13002843,Scotland,S99999999,"Aberdeen South",1980-01-01,1996-06-01,"","",,,
AB1 0AQ,No,57.098554,-2.252504,384800,800870,NJ848008,"","Aberdeen City","Lower Deeside",S12000033,S13002843,Scotland,S99999999,"Aberdeen South",1980-01-01,1996-06-01,"","",,,
AB1 0AR,No,57.091357,-2.224831,386474,800063,NJ864000,"","Aberdeenshire","North Kincardine",S12000034,S13002864,Scotland,S99999999,"West Aberdeenshire and Kincardine",1980-01-
AB1 0AS,No,57.083838,-2.234437,385889,799228,NO858992,"","Aberdeenshire","North Kincardine",S12000034,S13002864,Scotland,S99999999,"West Aberdeenshire and Kincardine",1980-01-
AB1 0AT,No,57.089299,-2.239768,385568,799837,NO855998,"","Aberdeenshire","North Kincardine",S12000034,S13002864,Scotland,S99999999,"West Aberdeenshire and Kincardine",1980-01-
AB1 0AX,No,57.095679,-2.24788,385079,800549,NJ850005,"","Aberdeenshire","North Kincardine",S12000034,S13002864,Scotland,S99999999,"West Aberdeenshire and Kincardine",1980-01-0
AB1 0BA,No,57.090696,-2.255866,384593,799996,NO845999,"","Aberdeenshire","North Kincardine",S12000034,S13002864,Scotland,S99999999,"West Aberdeenshire and Kincardine",1980-01-
AB1 0BB,No,57.086315,-2.258591,384426,799509,NO844995,"","Aberdeenshire","North Kincardine",S12000034,S13002864,Scotland,S99999999,"West Aberdeenshire and Kincardine",1980-01-
AB1 0BD,No,57.079125,-2.268851,383801,798711,NO838987,"","Aberdeenshire","North Kincardine",S12000034,S13002864,Scotland,S99999999,"West Aberdeenshire and Kincardine",1980-01-
AB1 0BE,No,57.090611,-2.257483,384495,799987,NO844999,"","Aberdeenshire","North Kincardine",S12000034,S13002864,Scotland,S99999999,"West Aberdeenshire and Kincardine",1991-03-
AB1 0BJ,No,57.081421,-2.262319,384198,798965,NO841989,"","Aberdeenshire","North Kincardine",S12000034,S13002864,Scotland,S99999999,"West Aberdeenshire and Kincardine",1980-01-
AB1 0BL,No,57.081938,-2.246567,385153,799019,NO851990,"","Aberdeenshire","North Kincardine",S12000034,S13002864,Scotland,S99999999,"West Aberdeenshire and Kincardine",1980-01-
AB1 0BN,No,57.069311,-2.258011,384454,797616,NO844976,"","Aberdeenshire","North Kincardine",S12000034,S13002864,Scotland,S99999999,"West Aberdeenshire and Kincardine",1980-01-
AB1 0BP,No,57.067251,-2.272212,383592,797390,NO835973,"","Aberdeenshire","North Kincardine",S12000034,S13002864,Scotland,S99999999,"West Aberdeenshire and Kincardine",1980-01-
AB1 0BQ,No,57.067834,-2.284733,382833,797458,NO828974,"","Aberdeenshire","North Kincardine",S12000034,S13002864,Scotland,S99999999,"West Aberdeenshire and Kincardine",1980-01-
AB1 0BR,No,57.062162,-2.278341,383218,796825,NO832968,"","Aberdeenshire","North Kincardine",S12000034,S13002864,Scotland,S99999999,"West Aberdeenshire and Kincardine",1980-01-
AB1 0BS,No,57.051402,-2.265255,384007,795624,NO840956,"","Aberdeenshire","North Kincardine",S12000034,S13002864,Scotland,S99999999,"West Aberdeenshire and Kincardine",1980-01-
```

As the data is for entire region of Scotland, there is a need to cleanse the data and get Glasgow only information. Also there are various columns and only **selected columns** are chosen for the study, and certain columns are used to reduce the data

Postcode	**Ward**	National Park	London zone
In Use?	District Code	**Population**	LSOA Code
Latitude	Ward Code	Households	Local authority
Longitude	Country	Built up area	MSOA Code
Easting	County Code	Built up sub-division	Middle layer super output area
Northing	Constituency	Lower layer super output area	Parish Code
Grid Ref	Introduced	Rural/urban	Census output area
County	Terminated	Region	Constituency Code
District	Parish	Altitude	Index of Multiple Deprivation
Quality	Plus Code	Nearest station	Average Income
User Type	Water company	Distance to station	Postcode district
Last updated	Postcode area	Police force	

Following operations are applied on the dataset, after loading it into Pandas Data frame

1. Filtering Glasgow only data – Using District column
2. Selecting only the data in use – Using “In Use?” column
3. Grouping the resultant data by Wards , with aggregating population
4. Selecting the columns that are relevant for the analysis

Shape of the original data was 224,804 rows and 47 columns. After cleansing the data selected for processing has a shape of 23 rows and 4 columns

```
glasgow_df.head(10)
```

	Ward	Latitude	Longitude	Population
0	Anderston/City/Yorkhill	55.869576	-4.306318	28099.0
1	Baillieston	55.867866	-4.154606	21526.0
2	Calton	55.860109	-4.245577	22806.0
3	Canal	55.926410	-4.294789	26364.0
4	Cardonald	55.860704	-4.379913	29876.0
5	Dennistoun	55.877318	-4.240140	20024.0
6	Drumchapel/Anniesland	55.918846	-4.387423	28598.0
7	East Centre	55.873410	-4.198915	26529.0
8	Garscadden/Scotstounhill	55.900029	-4.385637	30088.0
9	Govan	55.868672	-4.352455	24722.0

```
glasgow_df.shape
```

```
(23, 4)
```

Methodology section

For each of the wards, latitudes and longitudes are available.

Using this information, Four Square APIs are used to obtain the Venues that are closest to 2kms of the wards. From the result, venue types are obtained and stored against the ward

Once these venues are obtained, they are converted to “one hot” format and they are grouped by Average usage per ward.

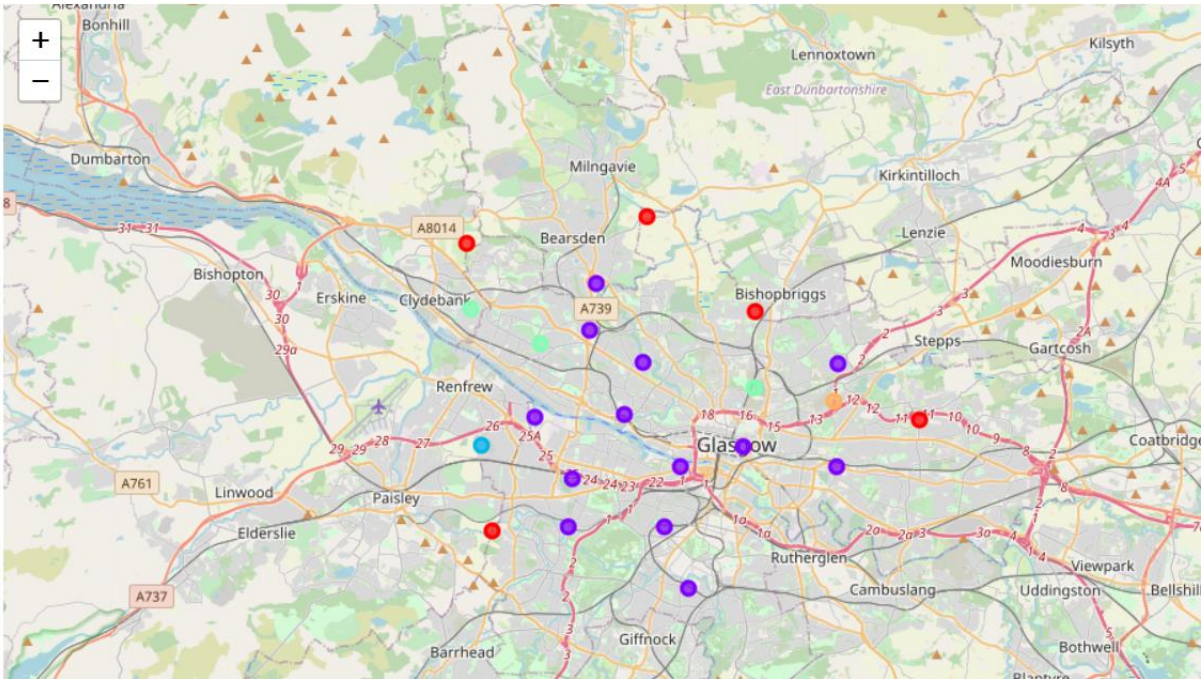
This will result in the top 5 venue types for each are obtained.

Further we would be applying unsupervised learning methodology of K-Means clustering to find cluster for each ward. This is to find out which wards forms part of which cluster.

For the purpose of this study, the value of k used will be 5

Results

Here is the output of cluster information depicted on a map using Folium.



Further to this, here are the outcome of clusters with top 10 venues for each clusters

Cluster 1

```
glasgow_merged.loc[glasgow_merged['Cluster Labels'] == 0, glasgow_merged.columns[[0] + [3] + list(range(5, glasgow_merged.shape[1]))]]
```

	Ward	Population	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
21	Springburn/Robroyston	25721.0	Chinese Restaurant	Whisky Bar	Convenience Store	Discount Store	Electronics Store	English Restaurant	Fast Food Restaurant	Fish & Chips Shop	Furniture / Home Store	Garden Center

Cluster 2

```
glasgow_merged.loc[glasgow_merged['Cluster Labels'] == 1, glasgow_merged.columns[[0] + [3] + list(range(5, glasgow_merged.shape[1]))]]
```

	Ward	Population	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Anderston/City/Yorkhill	28099.0	Café	Discount Store	Chinese Restaurant	Restaurant	Pharmacy	Supermarket	Pub	Bar	Outdoor Supply Store	Sandwich Place
2	Calton	22806.0	Coffee Shop	Café	Cocktail Bar	Seafood Restaurant	Bar	Pub	Italian Restaurant	Plaza	Hotel	Restaurant
9	Govan	24722.0	Hotel	Bistro	Gas Station	Garden Center	Furniture / Home Store	Pier	Scandinavian Restaurant	Restaurant	Auto Garage	Beer Bar
11	Hillhead	22905.0	Whisky Bar	Soccer Field	Convenience Store	Gym	Hotel	Jazz Club	Theater	College Cafeteria	Vegetarian / Vegan Restaurant	Deli / Bodega
12	Langside	28592.0	Supermarket	Train Station	Gas Station	Deli / Bodega	Discount Store	Electronics Store	English Restaurant	Fast Food Restaurant	Fish & Chips Shop	Furniture / Home Store
13	Linn	29478.0	Train Station	Bakery	Tennis Court	Café	Whisky Bar	Garden Center	Discount Store	Electronics Store	English Restaurant	Fast Food Restaurant
14	Maryhill	19718.0	River	Bus Stop	Grocery Store	Park	Deli / Bodega	Discount Store	Electronics Store	English Restaurant	Fast Food Restaurant	Fish & Chips Shop
15	Newlands/Auldburn	22502.0	College Cafeteria	Train Station	Track	Garden Center	Convenience Store	Deli / Bodega	Discount Store	Electronics Store	English Restaurant	Fast Food Restaurant
16	North East	19904.0	Pet Store	Café	Park	Whisky Bar	Discount Store	Electronics Store	English Restaurant	Fast Food Restaurant	Fish & Chips Shop	Furniture / Home Store
17	Partick East/Kelvindale	26887.0	Discount Store	Chinese Restaurant	Pub	Gym / Fitness Center	Auto Garage	Bank	Grocery Store	Electronics Store	Fast Food Restaurant	Supermarket
18	Pollokshields	26800.0	Auto Garage	Train Station	Bakery	Pizza Place	Whisky Bar	Garden Center	Discount Store	Electronics Store	English Restaurant	Fast Food Restaurant
19	Shettleston	26841.0	Clothing Store	Coffee Shop	Shopping Mall	Discount Store	Grocery Store	Bakery	Movie Theater	Sandwich Place	Supermarket	Fast Food Restaurant
20	Southside Central	23588.0	Hotel	American Restaurant	Buffet	Gym	Grocery Store	Movie Theater	Fast Food Restaurant	English Restaurant	Portuguese Restaurant	Coffee Shop

Cluster 3

```
glasgow_merged.loc[glasgow_merged['Cluster Labels'] == 2, glasgow_merged.columns[[0] + [3] + list(range(5, glasgow_merged.shape[1]))]]
```

	Ward	Population	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
4	Cardonald	29876.0	Go Kart Track	Whisky Bar	Garden Center	Deli / Bodega	Discount Store	Electronics Store	English Restaurant	Fast Food Restaurant	Fish & Chips Shop	Furniture / Home Store

Cluster 4

```
glasgow_merged.loc[glasgow_merged['Cluster Labels'] == 3, glasgow_merged.columns[[0] + [3] + list(range(5, glasgow_merged.shape[1]))]]
```

	Ward	Population	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
5	Dennistoun	20024.0	Bakery	Convenience Store	Whisky Bar	Gas Station	Discount Store	Electronics Store	English Restaurant	Fast Food Restaurant	Fish & Chips Shop	Furniture / Home Store
8	Garscadden/Scotstounhill	30088.0	Moving Target	Pharmacy	Whisky Bar	Garden Center	Discount Store	Electronics Store	English Restaurant	Fast Food Restaurant	Fish & Chips Shop	Furniture / Home Store
22	Victoria Park	21618.0	Bakery	Golf Course	Pharmacy	Whisky Bar	Garden Center	Discount Store	Electronics Store	English Restaurant	Fast Food Restaurant	Fish & Chips Shop

Cluster 5

```
glasgow_merged.loc[glasgow_merged['Cluster Labels'] == 4, glasgow_merged.columns[[0] + [3] + list(range(5, glasgow_merged.shape[1]))]]
```

	Ward	Population	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
7	East Centre	26529.0	Indian Restaurant	Italian Restaurant	Deli / Bodega	Discount Store	Electronics Store	English Restaurant	Fast Food Restaurant	Fish & Chips Shop	Furniture / Home Store	Garden Center

Discussion

Based on these results we can find out there are 5 clusters (as that is the number of clusters we used)

Here are the observations based on 5 clusters

Cluster 5 - Consisting of ward East Centre is well served with Restaurants. Top 2 venues are Indian restaurant and Italian Restaurant.

Cluster 1 - Consisting of ward Springburn/Robroyston is also having its fair share of eateries with position 1, 6, 7 and 8.

So clearly these two clusters are not a good choice for starting a new restaurants.

Cluster 3 **-** Consisting of ward Cardonald is not having Restaurants in its top 5. It has a population total of nearly 30,000 people

Cluster 4 - Consisting of ward Dennistoun / Victoria park/ Garscadden is also not having Restaurants in its top 5. Its cumulative population is nearly 71,000 people

Cluster 2 - Consisting of various wards is under served with Restaurants. With the exception of few wards, other wards doesn't have Restaurant in its top 5 . Its cumulative population is nearly 320,000 people

Conclusion

It is therefore suggested to open restaurants in the following five wards(in the specific order)

1. Garscadden – Population of 30088
2. Linn - Population of 29478
3. Langside - Population of 28592
4. Shettleston - Population of 26841
5. Govan - Population of 24722