CAPTSTONE PROJECT: BATTLE OF THE NEIGHBORHOODS

I. Purpose:

This document provides the details of my final peer reviewed assignment for the IBM Data Science Professional Certificate program – Coursera Capstone.

II. Introduction:

Glasgow is the most populous city in Scotland, and the third most populous city in the United Kingdom, as of the 2017 estimated city population of 621,020. Historically part of Lanarkshire, the city now forms the Glasgow City council area, one of the 32 council areas of Scotland;Glasgow is situated on the River Clyde in the country's West Central Lowlands. It is the fifth most visited city in the UK.

Glasgow has a population density of 3,400 people per square kilometer, which makes it the most densely populated city in Scotland. The larger Greater Glasgow area has an estimated population of 1.2 million, while the region surrounding the conurbation has about 2.8 million residents. This represents about 42% of the population of Scotland. Additionally, the city proper takes 175 square kilometers of area (approximately 68 square miles), while the metro area expands out to 3,338 square kilometers (about 1,289 square miles).

Coffee is the most popular drink worldwide with around two billion cups consumed every day. In the UK, we now drink approximately 95 million cups of coffee per day. The coffee industry creates over 210,000 UK jobs. The Gross Value-Added contribution from the UK coffee industry to the economy is estimated to be £9.1 billion, whilst output contribution, including indirect and induced multiplier impacts, of £17.7 billion in 2017.

In this project, we will atempt to use FourSquare and K-Means clustering to find the optimal location for opening a new cafe.

III. Data Acquisition:

This demonstration will make use of the following data sources:

- Greater Glasgow & Clyde areas and their size and population density: https://en.wikipedia.org/wiki/List_of_places_in_Glasgow
- Glasgow Top Venue Recommendations from FourSquare API (FourSquare website: www.foursquare.com)

I will be using the FourSquare API to explore areas in Glasgow. The Foursquare explore function will be used to get the location of coffee shops in each neighborhood, and then use this feature to group the neighborhoods into clusters. The following information are retrieved on the first query. Using these results, we can determine which neighborhood has the most coffee shops, where the most frequented coffee shop locations are and use this to find the best location for opening a new business.

IV. Methodology:

To get the required data, we first extracted the data concerning the areas of Glasgow – such as population density, coordinates and size.

	District	Population	Area (km²)	Density (/km²)
1	Govanhill	9,725	0.86	11,308
2	Pollokshields	9,738	1.59	6,125
3	Partick	8,884	0.85	10,452
4	Hillhead	6,275	0.96	6,536
5	Govan	5,860	1.63	3,595
6	Gorbals	6,030	0.83	7,265
7	Shawlands	7,015	0.52	13,490
8	Langside	4,425	0.46	9,620
Σ	Total	57,952	7.7	7,526

We can see from the table above that Shawlands has the highest population density, followed by Govanhill and Partick. So we can look to these three districts first. A quick google search will show us that Partick would be a better choice, as its close to

the more popular west end living location and also close to the University of Glasgow, making it more likely that a great number of students will live in the area, and it is known that students are avid consumers of coffee.

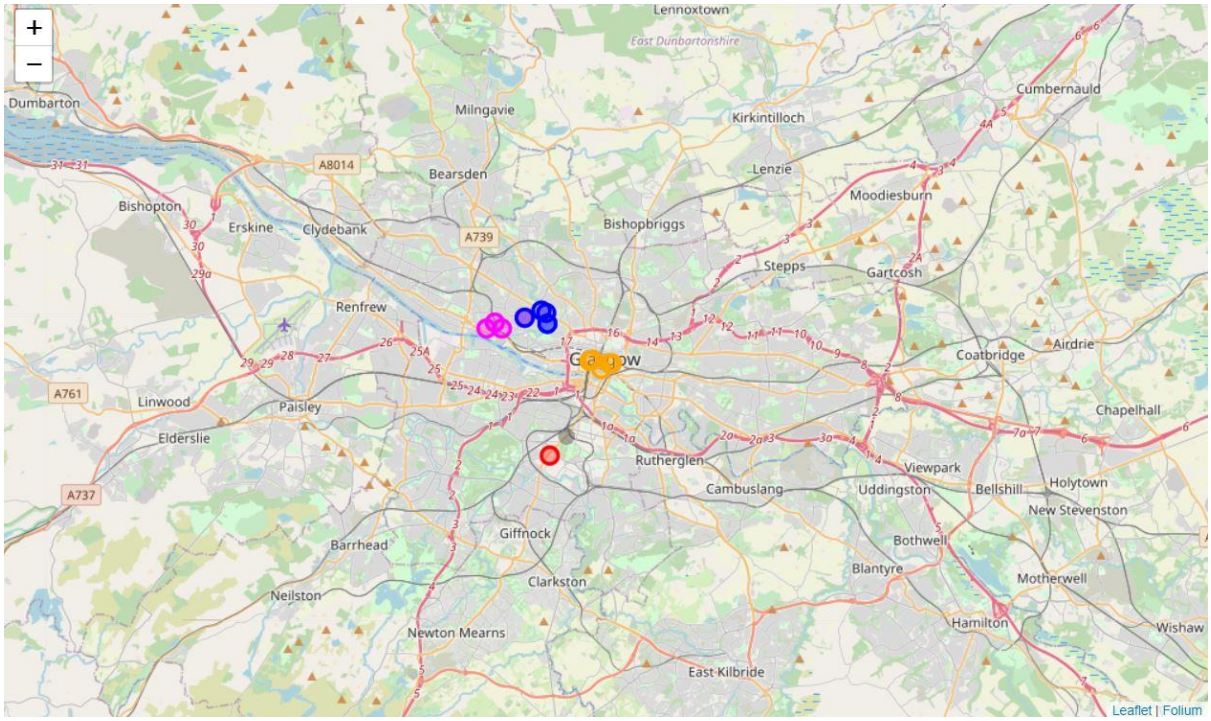
	District	Population	Area (km^2)	Density (/km^2)	District_Coord	Latitude	Longitude
0	Govanhill	9,725	0.86	11,308	(55.8363741, -4.2581531)	55.836374	-4.258153
1	Pollokshields	9,738	1.59	6,125	(55.8422663, -4.2849973)	55.842266	-4.284997
2	Partick	8,884	0.85	10,452	(55.8699211, -4.3094365)	55.869921	-4.309437
3	Hillhead	6,275	0.96	6,536	(55.8752091, -4.293281)	55.875209	-4.293281
4	Govan	5,860	1.63	3,595	(55.860879, -4.3185273)	55.860879	-4.318527
5	Gorbals	6,030	0.83	7,265	(55.851813, -4.2531625)	55.851813	-4.253163
6	Shawlands	7,015	0.52	13,490	(55.8292301, -4.2924584)	55.829230	-4.292458

I then used the Nominatim module to find the coordinates of each district, which will be used with the FourSquare API to get information on the venues in each of the respective districts. I also added two separate columns, each containing the Latitude and Longitude of these coordinates, for ease of analysis.

This data will be used in the second part of this project, where I will employ the FourSquare API to generate information on the venues in each district, see which district prefers coffee shops over everything else and use the population density and frequency of visits to find the best spot to open a coffee shop.

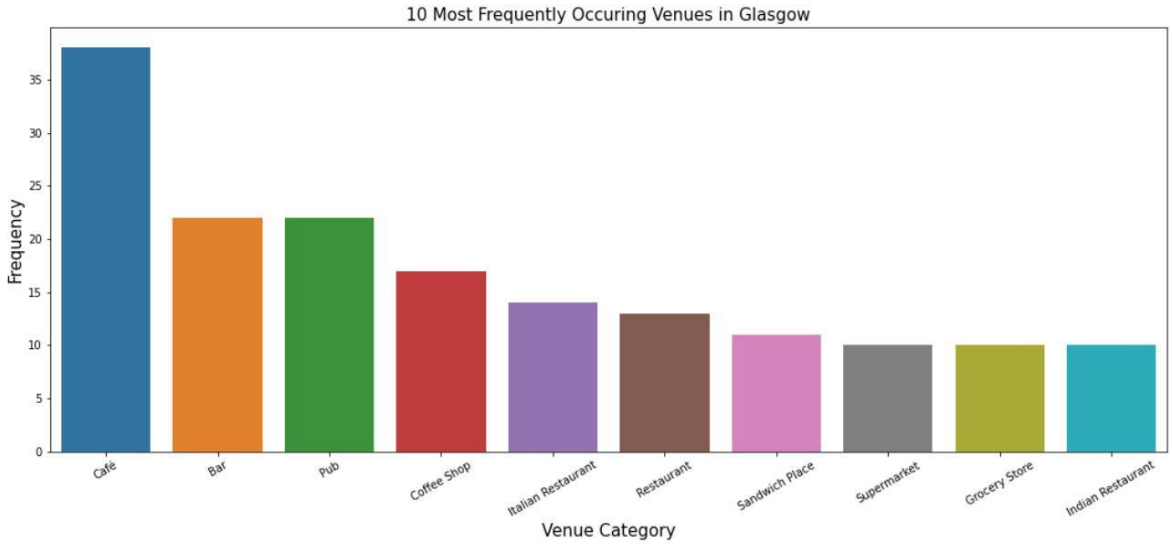
Coursera Capstone Project – Week 5

We use the FourSquare API to get the venues in all Glasgow areas. After acquiring all venues and their types from all Glasgow districts, we use the folium module to create a map of Glasgow and add the venue markers to it.



Next, we create a dataframe of all types of venues present. Having this, we index them and organize the venues to find the top 10 most frequented venue types.

	Venue_Category	Frequency
0	Café	41
1	Pub	23
2	Bar	21
3	Coffee Shop	16
4	Italian Restaurant	14
5	Restaurant	13
6	Grocery Store	13
7	Sandwich Place	11
8	Indian Restaurant	10
9	Supermarket	8



As the chart above shows, cafes are the most frequented venue type in Glasgow. This is in favor of our idea to open a coffee shop, as it shows a good client basis.

Pollokshields

	Venue	Freq
0	Park	0.14
1	Hotel Bar	0.07
2	Roller Rink	0.07
3	Café	0.07
4	Supermarket	0.07

Langside

	Venue	Freq
0	Grocery Store	0.12
1	Italian Restaurant	0.12
2	Café	0.12
3	Restaurant	0.08
4	Pizza Place	0.08

Shawlands

	Venue	Freq
0	Café	0.20
1	Pub	0.10
2	Supermarket	0.10
3	Italian Restaurant	0.07
4	Grocery Store	0.07

Partick

	Venue	Freq
0	Café	0.14
1	Pub	0.07
2	Bar	0.07
3	Coffee Shop	0.05
4	Restaurant	0.04

Gorbals

	Venue	Freq
0	Pub	0.07
1	Coffee Shop	0.07
2	Seafood Restaurant	0.05
3	Bar	0.05
4	Restaurant	0.04

Govanhill

	Venue	Freq
0	Bar	0.15
1	Indian Restaurant	0.11
2	Fast Food Restaurant	0.07
3	Train Station	0.07
4	Supermarket	0.04

Govan

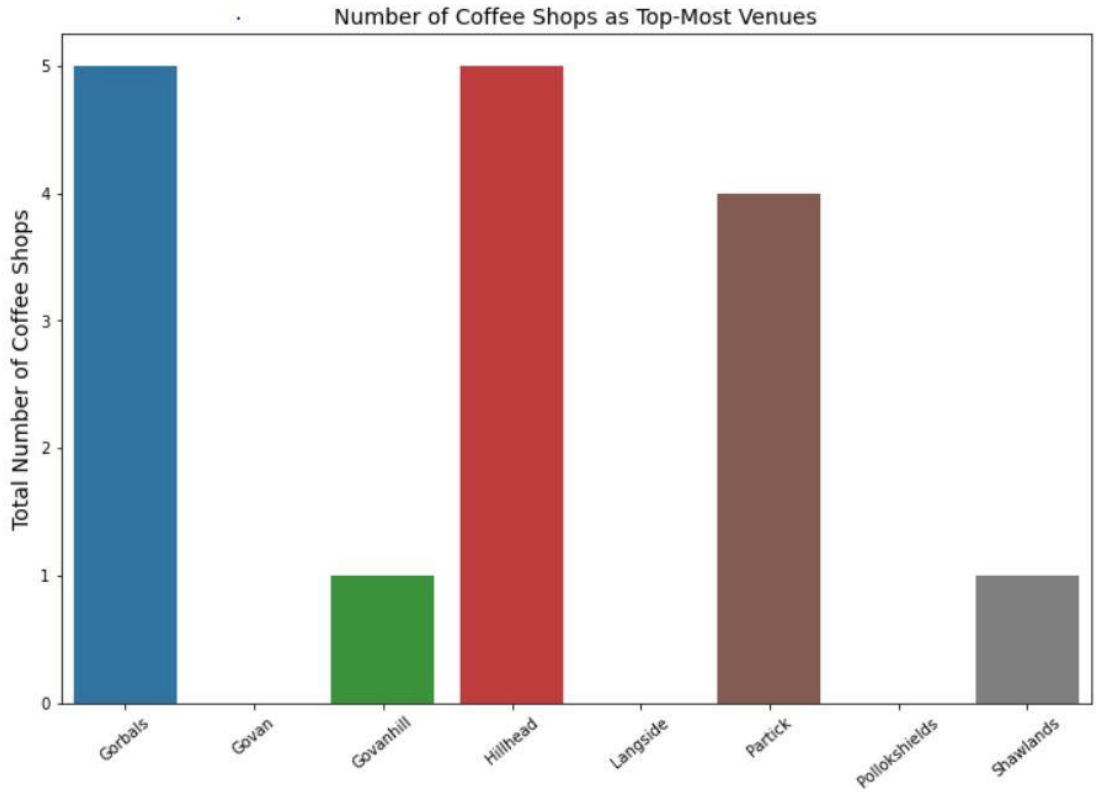
	Venue	Freq
0	Discount Store	0.15
1	Museum	0.15
2	Gas Station	0.15
3	Grocery Store	0.08
4	Metro Station	0.08

Hillhead

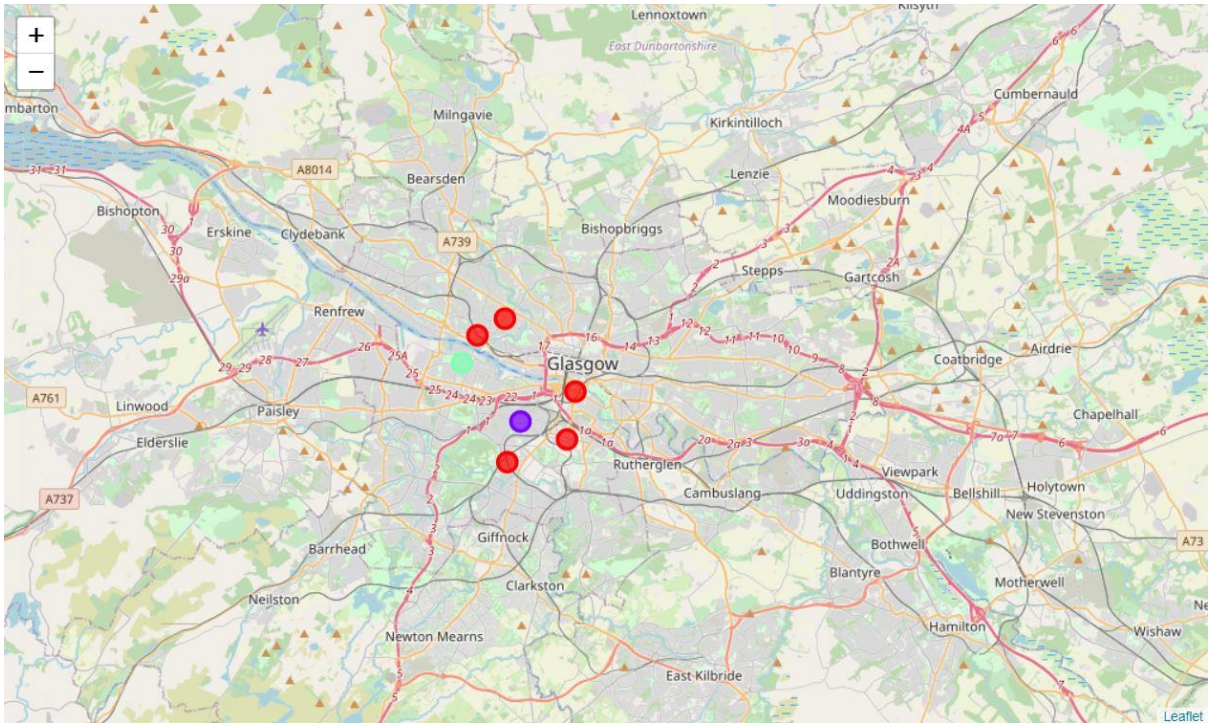
	Venue	Freq
0	Café	0.13
1	Bar	0.08
2	Pub	0.07
3	Coffee Shop	0.05
4	Indian Restaurant	0.05

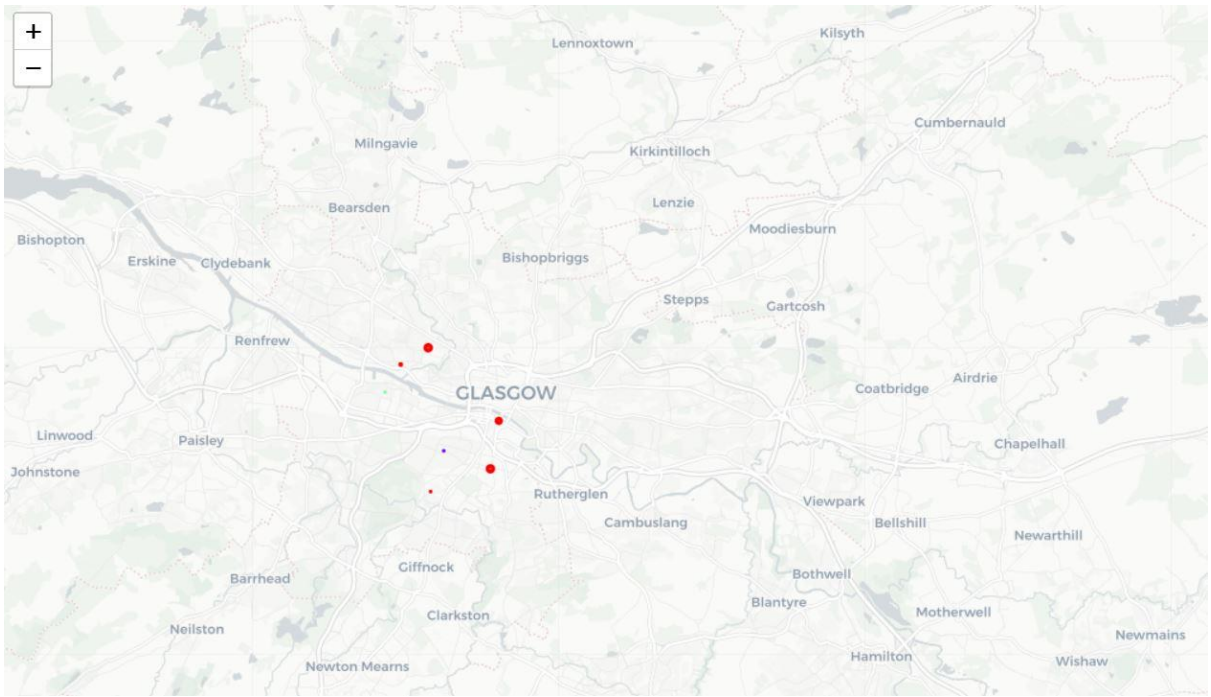
Next, we separate the districts and see which type of venue is most frequented in each of the individual districts. By inspecting this, we can get a better understand of where cafes are most visited.

We can see that Shawlands, Hillhead and Partick are the three districts of Glasgow in which cafes are most frequented. Meanwhile, Gorbals has pubs as the most popular type of venue, Govanhill prefers bars and Pollokshields parks. This reflects the social and economic status of these neighborhoods, as Govan tends to have many of the lower income class inhabitants, and as such, the most visited venues are Discount stores. Pollokshields hosts Pollok park, a popular Glasgow attraction – a large park that hosts highland cows, a widely loved species of Scottish cow.



We can now cluster the neighborhoods of Glasgow, to find the optimal one for opening a café. To do this, we use the K-Means algorithm and find the cluster points around Glasgow. We form our dataframe, add cluster labels and then use the folium module to create a map. We can then add the cluster labels as markers on this map.





Having the data available and ready to use, we can now proceed to analyze it.

V. Results

We have found that our of the 10 most popular venues in Glasgow cafes are, on average, the most frequented by people. By segmenting each district individually, we managed to find that Govan, Partick and Hillhead have cafes as the number 1 most frequented type of venue. We then added this data set in a dataframe and used this

to analyse which of these three neighborhoods (Govan, Partick, Hillhead) would be the best location for opening a new coffee shop.

VI. Discussion:

From the data analysis of the points found above, we can conclude that both Partick and Hillhead would be ideal for opening a new coffee shop. Both these neighborhoods see high traffic of students, great number of small and medium buisnesses, and the University of Glasgow is nearby, increasing the overall traffic. A quick internet search shows the rents are slightly smaller in Partick so if budget is a key factor in our decision, the odds could fall in favour of Partick.

VII. Conclusion:

Using data from Wikipedia and FourSquare, we found that Partick and Hillhead are the two best neighborhoods in Glasgow for opening a new coffee shop. Both these districts have high traffic and both students and young people which will make the majority of our customer basis.