Coursera Capstone – Battle of the Neighbourhoods

Business Problem

Business Problem Description:

Glasgow has a population of over 600,000 people [1]. This population is projected to rise in the coming years. Migration, both national and international, is attributed as the main driver of this growth [2]. This rapid population growth brings with it a lot of new opportunities within the city of Glasgow for businesses to cater to a wider variety of clientele.

Most of the Glasgow population is clustered around the city centre which is the main hub of economic activity. Glasgow also has a very reliable public transport system. According to the SPT Subway survey, 97% of subway commuters report overall satisfaction [3]. Therefore, subways are a popular mode of transport for the local population trying to commute within the city. 13.1 million subway passenger journeys were recorded in the 2018 to 2019 period [4].

So, areas close to the city centre (~ 5 km from the city centre) and close to a subway station (~ 2 km from the nearest subway station) offer businesses a great deal of opportunity and exposure to a wide customer base. So, these areas are ideal places to start a new business. In Glasgow's highly competitive business environment, identifying business opportunities and choosing the optimum location is a pivotal decision for new small businesses. Choosing the right location can have considerable impact on future profits and prospects for a business.

Therefore, this project aims to provide data driven insight to businesses by identifying possible business opportunities within Glasgow and the ideal location (postal district) for those business opportunities. The project will use venue information acquired using the FourSquareAPI to identify the most common venues within each postal code. The total population and their ethnicities are also acquired using datasets from the Glasgow City Council and the Royal mail. Insight from this data will allow postal codes to be clustered according to the most common venues and their population characteristics. This will offer insight and analysis into the type of business competition at each postal code and possible untapped business opportunities within Glasgow.

Data Description:

In order to complete this project five data sets are needed.

1) **Dataset 1:** The list of all postal codes in Glasgow

This dataset is scraped from the Wikipedia page given in [5]. It contains the postcode district, Post town, coverage and local authority area information of 56 postal codes. The dataset is processed, which reduces the postcodes to 50 valid entries. In order to find the latitude and longitude information of these 50 entries, the postcode district and the first coverage area are feed into the geolocator.geocode function that uses the FourSquare agent. This information will be concatenated with the dataset acquired from the Wikipedia webpage. The final postal code dataframe is given below.

	Postcode district	Post town	Coverage	Local authority area	District	Latitude	Longitude
0	G1	GLASGOW	Merchant City	Glasgow City	Merchant City	55.859126	-4.246316
1	G2	GLASGOW	Blythswood Hill, Anderston (part)	Glasgow City	Blythswood Hill	55.863319	-4.261671
2	G3	GLASGOW	Anderston, Finnieston, Garnethill, Park, Woodl...	Glasgow City	Anderston	55.859699	-4.271292
3	G4	GLASGOW	Calton (part), Cowcaddens (part), Drygate, Kel...	Glasgow City	Calton	55.853724	-4.232824
4	G5	GLASGOW	Gorbals	Glasgow City	Gorbals	55.851813	-4.253163
5	G11	GLASGOW	Broomhill, Partick, Partickhill	Glasgow City	Broomhill	55.877340	-4.321745
6	G12	GLASGOW	West End (part), Cleveden, Dowanhill, Hillhead...	Glasgow City	West End	55.879886	-4.292322
7	G13	GLASGOW	Anniesland, Knightswood, Yoker	Glasgow City	Anniesland	55.890019	-4.332739
8	G14	GLASGOW	Whiteinch, Scotstoun	Glasgow City	Whiteinch	55.875338	-4.337189
9	G15	GLASGOW	Drumchapel	Glasgow City	Drumchapel	55.912244	-4.369171
10	G20	GLASGOW	Maryhill, North Kelvinside, Ruchill	Glasgow City	Maryhill	55.890787	-4.290405
11	G21	GLASGOW	Balornock, Barmulloch, Cowlairs, Royston, Spri...	Glasgow City	Balornock	55.881657	-4.214442
12	G22	GLASGOW	Milton, Parkhouse, Possilpark	Glasgow City	Milton	55.897001	-4.246675
13	G23	GLASGOW	Lambhill, Summerston	Glasgow City	Lambhill	55.895932	-4.265246
14	G31	GLASGOW	Dennistoun, Haghill, Parkhead (part)	Glasgow City	Dennistoun	55.861736	-4.219010
15	G32	GLASGOW	Carmyle, Tollcross, Mount Vernon, Lightburn, S...	Glasgow City	Carmyle	55.831113	-4.155489
16	G33	GLASGOW	Cardowan, Carntyne, Craigend, Cranhill, Gartha...	North Lanarkshire, Glasgow City	Cardowan	55.860114	-4.172197
17	G34	GLASGOW	Easterhouse, Easthall, Provanhall	Glasgow City	Easterhouse	55.866744	-4.105808
18	G40	GLASGOW	Bridgeton, Calton, Dalmarnock	Glasgow City	Bridgeton	55.846267	-4.222231
19	G41	GLASGOW	Pollokshields, Shawlands	Glasgow City	Pollokshields	55.842266	-4.284997
20	G42	GLASGOW	Battlefield, Govanhill, Mount Florida, Strathb...	Glasgow City	Battlefield	55.824493	-4.266598
21	G43	GLASGOW	Mansewood, Newlands, Pollokshaws	Glasgow City	Mansewood	55.815297	-4.302073
22	G44	GLASGOW	Cathcart, Simshill, Croftfoot, King's Park, Mu...	Glasgow City, East Renfrewshire	Cathcart	55.816988	-4.262424
23	G45	GLASGOW	Castlemilk	Glasgow City	Castlemilk	55.805271	-4.230395
24	G46	GLASGOW	Arden, Carnwadric, Deaconsbank, Giffnock, Kenn...	East Renfrewshire, Glasgow City	Arden	55.806630	-4.330756
25	G51	GLASGOW	Cessnock, Plantation	Glasgow City	Cessnock	55.854152	-4.294398
26	G52	GLASGOW	Cardonald, Hillington, Penilee, Mosspark	Glasgow City	Cardonald	55.848291	-4.349974
27	G53	GLASGOW	Darnley, Pollok, Crookston, Parkhouse, Priesth...	Glasgow City	Darnley	55.808944	-4.340953
28	G60	GLASGOW	Bowling, Old Kilpatrick	West Dunbartonshire	Bowling	55.815327	-4.266320
29	G61	GLASGOW	Bearsden	East Dunbartonshire	Bearsden	55.920461	-4.333205
30	G62	GLASGOW	Baldernock, Milngavie, Mugdock	East Dunbartonshire, Stirling	Baldernock	55.947113	-4.280062
31	G63	GLASGOW	Balfron, Balmaha, Blanefield, Croftamie, Dryme...	Stirling	Balfron	56.071609	-4.333266
32	G64	GLASGOW	Bishopbriggs, Torrance	East Dunbartonshire	Bishopbriggs	55.904409	-4.226358
33	G65	GLASGOW	Croy, Kilsyth, Twechar	North Lanarkshire, East Dunbartonshire	Croy	55.889372	-4.204707
34	G66	GLASGOW	Clachan of Campsie, Haughhead, Kirkintilloch, ...	East Dunbartonshire	Clachan of Campsie	55.988662	-4.228600
35	G67	GLASGOW	Cumbernauld (south)	North Lanarkshire	Cumbernauld	55.944109	-3.994124
36	G68	GLASGOW	Cumbernauld (north), Dullatur	North Lanarkshire	Cumbernauld	55.874026	-4.184127

37	G69	GLASGOW	Baillieston, Bargeddie, Chryston, Garrowhill, ...	Glasgow City, North Lanarkshire	Baillieston	55.850211	-4.110833
38	G71	GLASGOW	Birkenshaw, Bothwell, Broomhouse, Tannochside,...	Glasgow City, North Lanarkshire, South Lanarks...	Birkenshaw	55.862058	-4.209337
39	G72	GLASGOW	Blantyre, Cambuslang	South Lanarkshire	Blantyre	55.793522	-4.092311
40	G73	GLASGOW	Rutherglen	South Lanarkshire	Rutherglen	55.828389	-4.211653
41	G74	GLASGOW	East Kilbride (north), Thorntonhall	South Lanarkshire	East Kilbride	55.760176	-4.179078
42	G75	GLASGOW	Auldhouse, East Kilbride (south)	South Lanarkshire	Auldhouse	55.818785	-4.295358
43	G76	GLASGOW	Busby, Carmunnock, Clarkston, Eaglesham, Water...	East Renfrewshire, Glasgow City	Busby	55.786185	-4.245825
44	G77	GLASGOW	Newton Mearns	East Renfrewshire	Newton Mearns	55.772673	-4.334237
45	G78	GLASGOW	Barrhead, Neilston, Uplawmoor	East Renfrewshire	Barrhead	55.801103	-4.389983
46	G81	CLYDEBANK	Dalmuir, Duntocher, Faifley, Hardgate	West Dunbartonshire	Dalmuir	55.924539	-4.402629
47	G82	DUMBARTON	Cardross, Milton, Renton	West Dunbartonshire, Argyll and Bute	Hardgate	55.964424	-4.657399
48	G83	ALEXANDRIA	Alexandria, Arrochar, Aldochlay, Ardlui, Ballo...	West Dunbartonshire, Argyll and Bute, Stirling	Alexandria	55.987874	-4.582209
49	G84	HELENSBURGH	Clynder, Cove, Garelochhead, Kilcreggan, Rhu, ...	Argyll and Bute	Clynder	55.855911	-4.305579

In order to shortlist the postal codes for clustering, the distance to the city centre is needed. There is no dataset available for the distance of a postal code to the Glasgow city centre. So, this is determined by finding the geographical coordinates of the city centre (using geolocator.geocode and FourSquare agent) and then using the haversine formula to find the distance between the latitude and longitude of the city centre to those of each of the 50 postal codes. Postal codes that are within 5 km of the city centre are considered for further analysis. Only 22 postal codes meet this requirement.

2) **Dataset 2:** The location of the Glasgow Subway station

This dataset is publicly available from the Glasgow City Council at this link [6]. The dataset is given below.

	Station Name	Address Line 1	Address Line 2	Address Line 3	Postcode	Easting	Northing	Accessible	Park and Ride	Wi-Fi	Bike Shelter	Special Car Sharing Parking Spaces	Lost Property Number	Nearby
0	Bridge Street	67	Eglinton Street	Glasgow	G5 9NR	258710	664400	Yes	Yes	Yes	Yes	No	0141 335 3276.	Citizen's Theatre, 02 Academy.
1	Buchanan Street	174	Buchanan Street	Glasgow	G1 2JZ	259073	665577	Yes	No	Yes	No	No	0141 335 3276.	Queen's Street Station, Buchanan Bus Station, ...
2	Cessnock	1	Cessnock Street	Glasgow	G51 1AS	256470	664490	Yes	No	Yes	No	No	0141 335 3276.	Glasgow Science Centre, IMAX, SECC
3	Cowcaddens	11	Dundasvale Court, Unit 1	Glasgow	G4 0SN	258679	666191	Yes	No	Yes	No	No	0141 335 3276.	Glasgow School of Art, GFT, Tenement House.
4	Govan	737	Govan Road	Glasgow	G51 2YJ	255490	665643	Yes	No	Yes	No	No	0141 335 3276.	Southern General Hospital, Govan Old Parish Ch...
5	Hillhead	248	Byres Road	Glasgow	G12 8SH	256615	667052	Yes	No	Yes	No	No	0141 335 3276.	Byres Road, Ashton Lane, Botanic Gardens, Univ...
6	Ibrox	124	Copland Road	Glasgow	G51 2UB	255830	664787	Yes	No	Yes	No	No	0141 335 3276.	Bellahouston Park
7	Kelvinhall	7	Dalcross Pass	Glasgow	G11 5RA	256184	666598	Yes	No	Yes	No	No	0141 335 3276.	Kelvingrove Museum, Western Infirmary
8	Kelvinbridge	99	South Woodside Road	Glasgow	G4 9HG	257475	666922	Yes	Yes	Yes	Yes	No	0141 335 3276.	Kelvingrove Park,
9	Kinning Park	30	Cornwall Street	Glasgow	G41 1AH	256864	664281	Yes	No	Yes	No	No	0141 335 3276.	NaN
10	Partick	15	Merkland Street	Glasgow	G11 6DB	255635	666503	Yes	No	Yes	Yes	No	0141 335 3276.	Partick Rail Station, Partick Bus Station, Riv...
11	Sheilds Road	360	Scotland Street	Glasgow	G5 8QF	257660	664205	Yes	Yes	Yes	Yes	Yes	0141 335 3276.	Scotland Street School, Museum Of Education
12	St. Enoch	10	St Enoch Square	Glasgow	G1 4DB	258946	665038	Yes	No	Yes	No	No	0141 335 3276.	Glasgow Central Station, St. Enoch Centre.
13	St Georges Cross	51	Great Western Road	Glasgow	G4 9AH	258106	666560	Yes	No	Yes	No	No	0141 335 3276.	The Mitchell Library.
14	West Street	281	West Street	Glasgow	G5 8NB	258220	664150	Yes	No	Yes	No	No	0141 335 3276.	NaN

This dataset lists the addresses of the 15 subway stations in Glasgow. In order to find the geographical coordinates, the addresses are fed to the geolocator.geocode function. The coordinates are then used to find the distance of each postal code to the nearest subway station. The postal codes that are within 2 km of the subway station are shortlisted. Only 10 postal codes meet this requirement.

3) **Dataset 3 and 4:** The Royal mail dataset and the Glasgow City Council dataset.

There is no dataset that maps postal codes directly to population and ethnicity data. Hence in order to extract this information two datasets are used. The first dataset is a Royal mail dataset [7]. This data set contains all the postal codes in Glasgow. It has almost 15,000 data entries. It maps the postal codes in Glasgow to the corresponding datazones. The Glasgow City Council dataset then maps datazones to population information [8]. The data set has almost 6200 entries. The data sets are processed to extract the information of the population at each postcode and their ethnicities. The following dataset is extracted. This information will then be inputted to the unsupervised machine learning algorithm.

	All people	White	White Percentage	Asian	Asian Percentage	African	African Percentage	Post Code	Other	Other Percentage
0	36994	26830	72.525274	8792	23.766016	639	1.727307	G41	733	1.981402
1	36269	31681	87.350079	2866	7.902065	1028	2.834376	G20	694	1.913480
2	31367	27572	87.901298	2814	8.971212	249	0.793828	G12	732	2.333663
3	26862	20312	75.616112	5083	18.922642	478	1.779465	G3	989	3.681781
4	24474	18245	74.548500	4353	17.786222	1008	4.118657	G4	868	3.546621
5	24405	20653	84.626101	2641	10.821553	740	3.032166	G51	371	1.520180
6	23157	20916	90.322581	1598	6.900721	169	0.729801	G11	474	2.046897
7	13291	10450	78.624633	2084	15.679783	306	2.302310	G1	451	3.393274
8	11781	9669	82.072829	1283	10.890417	611	5.186317	G5	218	1.850437
9	9736	7578	77.834840	1565	16.074363	218	2.239113	G2	375	3.851684

4) **Dataset 5:** The venues data

This data is extracted using the FoursquareAPI. Venues in the ten postal codes identified above are extracted. All venues within a 1.5 km radius of the postal code are imported. The dataset is cleaned to remove any repeating venues. The venue categories are one-hot encoded. The one-hot encoded venues dataframe is merged with the population data and then fed into an unsupervised learning algorithm (K-means) for clustering into 5 categories. The clusters are then examined to find possible business opportunities at each postal code.

References:

[1] https://www.citypopulation.de/en/uk/scotland/glasgow_city/S19000899__glasgow/

[2] https://www.glasgowlive.co.uk/news/glasgow-news/glasgows-population-continues-rise-the-14550988

[3] http://www.spt.co.uk/documents/latest/rtp220618_agenda6.pdf

[4] https://www.statista.com/statistics/305479/passenger-journeys-on-the-glasgow-underground-united-kingdom-uk/

[5] https://en.wikipedia.org/wiki/G_postcode_area

[6] https://data.glasgow.gov.uk/dataset/glasgow-subway-station-locations/resource/8c057567-55e7-4631-ae77-e37b4a0562e8

[7] http://www.glasgow.gov.uk/CHttpHandler.ashx?id=16002&p=0

[8] https://data.glasgow.gov.uk/dataset/national-identity-by-ethnic-group-2011-census