The Battle of Neighbourhoods (Edinburgh vs Glasgow)

Author S Cavin

1 Introduction

1.1 Background

Edinburgh and Glasgow are the two major cities in Scotland, situated one in the East side and other in the West side of the country.

There is a historical rivalry between these great cities, as one is the capital of Scotland, and the other one is the largest. Both competing in public services, entertainment, touristic attractions, and in attracting business to their constituencies.

Both cities are noted for their culture, architecture, services, tourism, and business links. Here is a more detailed summary of each city, according to Wikipedia (www.wikipedia.org):

- Edinburgh is Scotland's second most populous city and the seventh most populous in the United Kingdom. The official population estimates are 488,050 (2016) for the Locality of Edinburgh (Edinburgh pre 1975 regionalisation plus Currie and Balerno), 518,500 (2018) for the City of Edinburgh, and 1,339,380 (2014) for the city region. Edinburgh lies at the heart of the Edinburgh and South East Scotland city region comprising East Lothian, Edinburgh, Fife, Midlothian, Scottish Borders and West Lothian.
- 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; the local authority is Glasgow City Council. Glasgow is situated on the River Clyde in the country's West Central Lowlands. It is the fifth most visited city in the UK.

This report as the title suggested aims to analyse these two great neighbour cities, based on data available online regarding their business and services, and to understand the difference or commonalities of both, and their future development.

1.2 Problem

Due to a historical rivalry between Edinburgh and Glasgow, there is a perception among local people that both cities are different and offer different services to their respective population.

In recent years, however, there has been an increase in the development of both cities, in the number and variety of local businesses, as a high demand in services is required due to the continuous increase of their population, from various backgrounds and origin, and their increase popularity as a preferable place of residence and tourism destination.

This project aims to understand how wide is the difference in both cities, based on the analysis of data of each city neighbourhood, and their neighbourhood facilities/amenities provided by their local businesses.

The findings will also enable to understand, their neighbourhoods similarities and differences, the top businesses that are flourishing (in numbers) in those cities, an understanding of the population and their behaviour pattern, along with type of businesses that are popular (and the least ones) in their neighbourhoods.

1.3 Interest

The target audience for this report are, city business developers, local authorities, local people (to understand the facilities provided by their cities and their differences), people intending to re-locate to these cities, people that look for new business opportunities.

2 Data acquisition and cleaning

2.1 Data sources

Based on definition of the problem, factors that will influence the decision are:

- Type of businesses that flourish in each city's neighbourhood
- Number of business grouped in category for each city
- Clustering of neighbourhoods related to the most common businesses
- Comparison on results for each city

The data, used for this analysis, is described below for each city.

2.1.1 For Edinburgh

The datasets used to analyse the city of Edinburgh comes from the following data sources:

- City postcodes, which is provided by Wikipedia (through a web link for webpage processing) that contains the city postcodes and their catchment neighbourhoods: https://en.wikipedia.org/wiki/EH_postcode_area
- City postcodes location (latitude, longitude), which is provided by freemaptools (through a weblink to download a csv file): https://www.freemaptools.com/download/outcode-postcodes/postcode-outcodes.csv)
- Information of local businesses, types and locations for each city postcode area, obtained using Foursquare API, via a subscription to the Foursquare site.

Here are samples of the data of Edinburgh used for this analysis. See Figure 1, and 2.

	postcode	post town	neighborhood
0	EH1	EDINBURGH	Old Town, GPO, St. James Centre
1	EH2	EDINBURGH	New Town, Princes Street
2	EH3	EDINBURGH	Queen Street, Stockbridge, West End, Tollcross...
3	EH4	EDINBURGH	Dean Village, Comely Bank, A90, Barnton, Cramo...
4	EH5	EDINBURGH	Granton, Firth of Forth, Ferry Road
5	EH6	EDINBURGH	Leith, Newhaven
6	EH7	EDINBURGH	Restalrig, Craigentinny
7	EH8	EDINBURGH	Southside, Newington, Canongate, Holyrood Park...
8	EH9	EDINBURGH	Marchmont, Grange
9	EH10	EDINBURGH	A702, Bruntsfield, Morningside, Fairmilehead
10	EH11	EDINBURGH	A71, Haymarket, Gorgie, Stenhouse, Sighthill, ...
11	EH12	EDINBURGH	A8, Murrayfield, Corstorphine, Gyle
12	EH13	EDINBURGH	Colinton, Oxgangs
13	EH14	BALERNO, CURRIE, EDINBURGH, JUNIPER GREEN	Slateford, Longstone, Wester Hailes, Juniper G...
14	EH15	EDINBURGH	Portobello, Duddingston

Figure 1. Sample data of Edinburgh neighbours with their postcode

	postcode	latitude	longitude
0	EH1	55.952430	-3.188400
1	EH10	55.920770	-3.209840
2	EH11	55.933870	-3.248670
3	EH12	55.942620	-3.271370
4	EH13	55.907880	-3.241440
5	EH14	55.909250	-3.283080
6	EH15	55.946860	-3.111360
7	EH16	55.922210	-3.153870
8	EH17	55.907040	-3.142220
9	EH2	55.954170	-3.194860
10	EH28	55.930600	-3.386240
11	EH29	55.956520	-3.398380
12	EH3	55.954120	-3.199670
13	EH30	55.984550	-3.383360
14	EH4	55.962530	-3.257610

Figure 2. Sample data of Edinburgh postcodes with longitude and latitude

2.1.2 For Glasgow

The datasets used to analyse the city of Glasgow comes from the following data sources:

- City postcodes, which is provided by Wikipedia (through a web link for webpage processing) that contains the city postcodes and their catchment neighbourhoods: https://en.wikipedia.org/wiki/G_postcode_area
- City postcodes location (latitude, longitude), which is provided by freemaptools (through a web link to download a csv file): https://www.freemaptools.com/download/outcode-postcodes/postcode-outcodes.csv)
- Information of local businesses, types and locations for each city postcode area, obtained using Foursquare API, via a subscription to the Foursquare site.

Here are samples of the data of Glasgow used for this analysis. See Figure 3, and 4.

	postcode	post town	neighborhood
0	G1	GLASGOW	Merchant City
1	G2	GLASGOW	Blythswood Hill, Anderston
2	G3	GLASGOW	Anderston, Finnieston, Garnethill, Park, Woodl...
3	G4	GLASGOW	Calton, Cowcaddens, Kelvinbridge, Townhead, Wo...
4	G5	GLASGOW	Gorbals
5	G11	GLASGOW	Broomhill, Partick, Partickhill
6	G12	GLASGOW	West End, Dowanhill, Hillhead, Hyndland, Kelvi...
7	G13	GLASGOW	Anniesland, Knightswood, Yoker
8	G14	GLASGOW	Whiteinch, Scotstoun
9	G15	GLASGOW	Drumchapel
10	G20	GLASGOW	Maryhill, North Kelvinside, Ruchill
11	G21	GLASGOW	Balornock, Barmulloch, Cowlairs, Royston, Spri...
12	G22	GLASGOW	Milton, Parkhouse, Possilpark
13	G23	GLASGOW	Lambhill, Summerston
14	G31	GLASGOW	Dennistoun, Haghill, Parkhead
15	G32	GLASGOW	Carmyle, Tollcross, Mount Vernon, Lightburn, S...
16	G33	GLASGOW	Cardowan, Carntyne, Craigend, Cranhill, Gartha...
17	G34	GLASGOW	Easterhouse, Easthall, Provanhall
18	G40	GLASGOW	Bridgeton, Calton, Dalmarnock
19	G41	GLASGOW	Pollokshields, Shawlands

Figure 3. Sample data of Glasgow neighbours with their postcode

	postcode	latitude	longitude
0	G1	55.860380	-4.246710
1	G11	55.873560	-4.311420
2	G12	55.880060	-4.300610
3	G13	55.893580	-4.346200
4	G14	55.880950	-4.348640
5	G15	55.909400	-4.364760
6	G2	55.863820	-4.254900
7	G20	55.885800	-4.281760
8	G21	55.880630	-4.220690
9	G22	55.889980	-4.250020
10	G23	55.901930	-4.284310
11	G3	55.866190	-4.272620
12	G31	55.857480	-4.208190
13	G32	55.848400	-4.162930
14	G33	55.873510	-4.165930

Figure 4. Sample data of Glasgow postcodes with longitude and latitude

Note:

The data, used for each city analysis, is required to be segmented and processed, before starting the intended analysis. The dataset also contains the latitude and longitude coordinates of each neighbourhood.

The location (latitude, longitude), is important as it helps in the analysis, due to its use by the Foursquare site, who is the provider of insight information of each neighbourhood regarding top businesses, amenities provided in the neighbourhood, used by the local population.

For the Foursquare site, a subscription is used to obtain relevant data for the neighbourhood analysis.

2.2 Data cleaning

The data sets obtained for each city were of similar characteristics, as they came from the same sources. However, in processing each dataset for each city, there were some particularities found that required data cleaning processing for the information to be used for the data analysis.

The first step was to obtain the postcode data, with the longitude and latitude for both cities. This information was provided through the freemaptools website. The challenge with this dataset is that is provided a single CSV file that contains all the postcodes for the whole UK. For the purpose of this project, only the postcodes of Edinburgh and Glasgow were required. For this, it was required to understand the postcodes that belong to the cities of Edinburgh and Glasgow, as there was no description on the postcode information provided of the city or neighbourhood that they belong.

The postcode data was downloaded to a pandas dataframe, and it includes in each data row: an id, postcode description, latitude and longitude, for each postcode available. The total number of rows and columns of this dataset were 3003 rows and 4 columns.

The column id, as it was not required, was drop from the postcode data frame. The result of the postcode dataframe included 3003 rows and 3 columns.

The second step, applied to the city of Edinburgh dataset, was to download the information from the Wikipedia page for Edinburgh (as described in 2.1.1). The data provided by this Wikipedia web page include the list of postcodes that belong to the city of Edinburgh. The challenge was that the data was included as part of the content of this web page, therefore, it was required to do an information extraction process of the html page.

For the information extraction, it was used a library called BeautifulSoup. It was identified the html table that contained a set of columns that describes postcodes, post towns, coverages (textual information of neighbourhoods per postcode), and others. Coding, transformation and manipulation of data was required, and extraction of textual information from the coverage column (as it included the list of neighbourhood as part of the text available per row). A pandas dataframe with three columns (postcode, post town, neighbourhood) was the result of the clean and transformation process in this step.

The first and second steps were repeated for the datasets provided for the city of Glasgow.

After, the above processes, three datasets were resulted:

- UK Postcodes with longitude and latitude
- Postcodes with neighbourhoods for Edinburgh
- Postcodes with neighbourhoods for Glasgow

2.3 Feature Selection

From the processed datasets described in section 2.2, for each city, the resulting information should be merged to use it for the planned data analysis.

For the Edinburgh dataset, two datasets will be merged (UK postcodes with longitude and latitude, and Postcodes and neighbourhoods for Edinburgh), the pandas “merge” function is used for this purpose.

The result is a dataset containing 23 rows and 5 columns. With the following features: Postcode, Latitude, Longitude, Post Town, and Neighbourhood. Figure 5 provides a sample of this data.

	postcode	latitude	longitude	post town	neighborhood
0	EH1	55.952430	-3.188400	EDINBURGH	Old Town, GPO, St. James Centre
1	EH10	55.920770	-3.209840	EDINBURGH	A702, Bruntsfield, Morningside, Fairmilehead
2	EH11	55.933870	-3.248670	EDINBURGH	A71, Haymarket, Gorgie, Stenhouse, Sighthill, ...
3	EH12	55.942620	-3.271370	EDINBURGH	A8, Murrayfield, Corstorphine, Gyle
4	EH13	55.907880	-3.241440	EDINBURGH	Colinton, Oxgangs
5	EH14	55.909250	-3.283080	BALERNO, CURRIE, EDINBURGH, JUNIPER GREEN	Slateford, Longstone, Wester Hailes, Juniper G...
6	EH15	55.946860	-3.111360	EDINBURGH	Portobello, Duddingston
7	EH16	55.922210	-3.153870	EDINBURGH	Liberton, Cameron Toll, Craigmillar, Niddrie
8	EH17	55.907040	-3.142220	EDINBURGH	Gilmerton, Moredun, Mortonhall
9	EH2	55.954170	-3.194860	EDINBURGH	New Town, Princes Street
10	EH28	55.930600	-3.386240	NEWBRIDGE	Newbridge, Ratho
11	EH29	55.956520	-3.398380	KIRKLISTON	Kirkliston
12	EH3	55.954120	-3.199670	EDINBURGH	Queen Street, Stockbridge, West End, Tollcross...
13	EH30	55.984550	-3.383360	SOUTH QUEENSFERRY	South Queensferry
14	EH4	55.962530	-3.257610	EDINBURGH	Dean Village, Comely Bank, A90, Barnton, Cramo...

Figure 5. Sample Edinburgh neighbourhood data

For the Glasgow dataset, two datasets will be merged (UK postcodes with longitude and latitude, and Postcodes and neighbourhoods for Glasgow), the pandas “merge” function is used for this purpose.

The result is a dataset containing 53 rows and 5 columns. With the following features: Postcode, Latitude, Longitude, Post Town, and Neighbourhood. Figure 6 provides a sample of this data.

	postcode	latitude	longitude	post town	neighborhood
0	G1	55.860380	-4.246710	GLASGOW	Merchant City
1	G11	55.873560	-4.311420	GLASGOW	Broomhill, Partick, Partickhill
2	G12	55.880060	-4.300610	GLASGOW	West End, Dowanhill, Hillhead, Hyndland, Kelvi...
3	G13	55.893580	-4.346200	GLASGOW	Anniesland, Knightswood, Yoker
4	G14	55.880950	-4.348640	GLASGOW	Whiteinch, Scotstoun
5	G15	55.909400	-4.364760	GLASGOW	Drumchapel
6	G2	55.863820	-4.254900	GLASGOW	Blythswood Hill, Anderston
7	G20	55.885800	-4.281760	GLASGOW	Maryhill, North Kelvinside, Ruchill
8	G21	55.880630	-4.220690	GLASGOW	Balornock, Barmulloch, Cowlairs, Royston, Spri...
9	G22	55.889980	-4.250020	GLASGOW	Milton, Parkhouse, Possilpark
10	G23	55.901930	-4.284310	GLASGOW	Lambhill, Summerston
11	G3	55.866190	-4.272620	GLASGOW	Anderston, Finnieston, Garnethill, Park, Woodl...
12	G31	55.857480	-4.208190	GLASGOW	Dennistoun, Haghill, Parkhead
13	G32	55.848400	-4.162930	GLASGOW	Carmyle, Tollcross, Mount Vernon, Lightburn, S...
14	G33	55.873510	-4.165930	GLASGOW	Cardowan, Carntyne, Craigend, Cranhill, Gartha...

Figure 6. Sample Glasgow neighbourhood data

3 Methodology

The methodology use for this project included a set of frameworks, algorithms, and tools for the exploratory data analysis of the dataset of Edinburgh and Glasgow. This includes the use of visualisation tools like maps techniques, foursquare API to obtain information of the businesses and services provided by each city and their neighbourhoods, and machine learning techniques of segmentation and clustering of Edinburgh and Glasgow neighbourhoods, this to understand the patterns of businesses and services provided by each city, using K-nearest neighbour algorithm.

3.1 Data Visualisation (Mapping)

From the data collected and processed for Edinburgh and Glasgow, a visualisation technique is used to plot the locations of the neighbourhood of these two cities.

The library used is Folium, which allows us to plot in a geographical map by using the location (latitude and longitude) of each city's neighbourhoods.

Plotting this information, provide insight knowledge of the location of neighbourhoods, proximity, and potential data correlation.

For the Edinburgh dataset, Figure 7 is the resulting map.

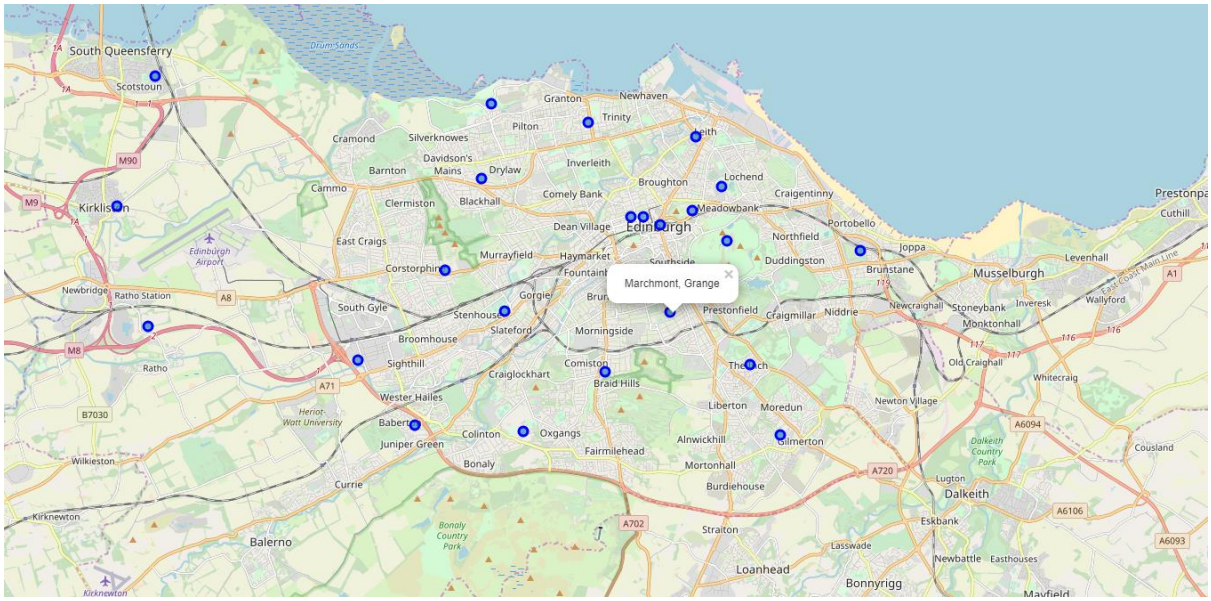

Figure 7. Edinburgh map with the plotted neighbourhoods

For the Glasgow dataset, Figure 8 is the resulting map.

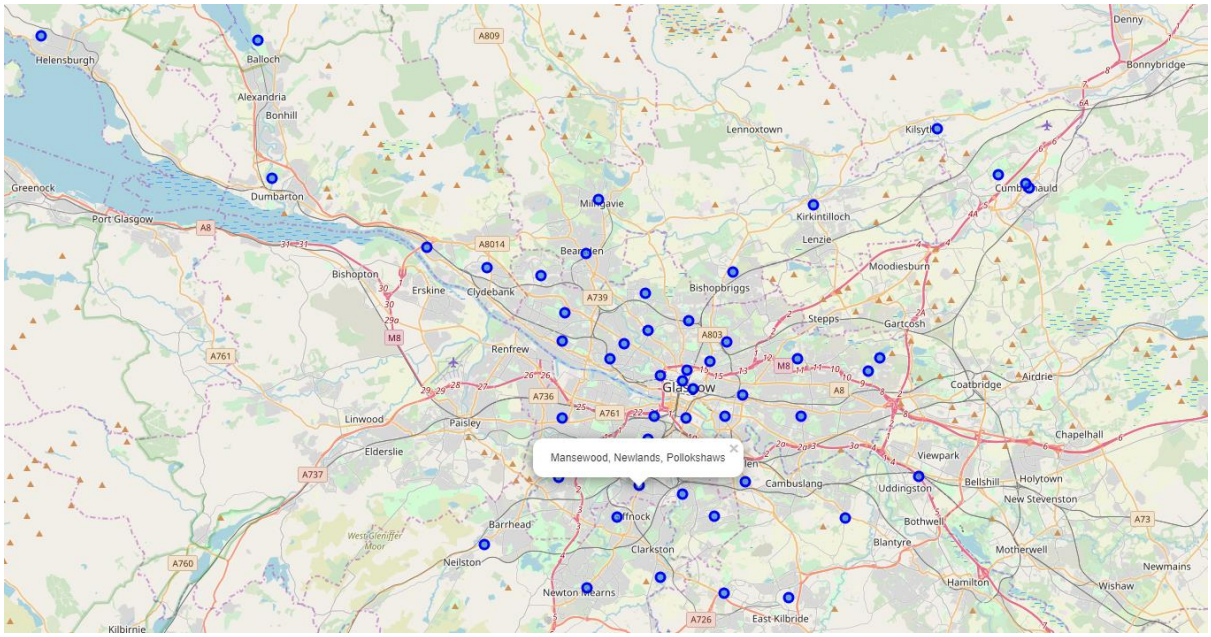

Figure 8. Glasgow map with the plotted neighbourhoods

3.2 Location-based Data Gathering and Analysis (Foursquare API)

For the data gathering and analysis of location-based information required for this project, cities of Edinburgh and Glasgow and their neighbourhoods, a set of libraries and data manipulation processes (normalisation and mean calculations) were used. The information was obtained from a subscription to the Foursquare API, which provides information for a specific location (latitude, longitude), about businesses and services provided in that area, within an assigned radius limit.

For this purpose, a set of limit of 100 business as maximum per location was set, along with a radius of 500 metres.

The information processed and obtained for each city are: list of business/services provided for each neighbourhood, number of business/services per neighbourhood, frequency of type of business/services per neighbourhood, and the 10 most common venues per neighbourhood.

For the Edinburgh dataset, a set of data set information, and data processing about businesses and services were obtained.

Figure 9, 10, 11 and 12 are samples of these datasets.

	neighborhood	neighborhood latitude	neighborhood longitude	venue	venue latitude	venue longitude	venue category
0	Old Town, GPO, St. James Centre	55.95243	-3.1884	The Balmoral Hotel	55.953113	-3.189509	Hotel
1	Old Town, GPO, St. James Centre	55.95243	-3.1884	The Guildford Arms	55.953668	-3.190052	Pub
2	Old Town, GPO, St. James Centre	55.95243	-3.1884	Apple Princes Street	55.953354	-3.189947	Electronics Store
3	Old Town, GPO, St. James Centre	55.95243	-3.1884	The Voodoo Rooms	55.953622	-3.190504	Bar
4	Old Town, GPO, St. James Centre	55.95243	-3.1884	Princes Street Suites	55.953370	-3.186934	Hotel

Figure 9. Data for one of Edinburgh neighbourhoods

neighborhood	neighborhood latitude
A702, Bruntsfield, Morningside, Fairmilehead	5
A71, Haymarket, Gorgie, Stenhouse, Sighthill, the Calders	4
A8, Murrayfield, Corstorphine, Gyle	14
Colinton, Oxgangs	4
Dean Village, Comely Bank, A90, Barnton, Cramond, Sainsbury's, Craigleith, A90	2
Gilmerton, Moredun, Mortonhall	5
Granton, Firth of Forth, Ferry Road	5
Jobcentre Plus	10
Kirkliston	4
Leith, Newhaven	27
Liberton, Cameron Toll, Craigmillar, Niddrie	4
Marchmont, Grange	4
New Town, Princes Street	100
Newbridge, Ratho	3
Old Town, GPO, St. James Centre	100
Portobello, Duddingston	1
Queen Street, Stockbridge, West End, Tollcross, Fountainbridge	100

Figure 10. Number of businesses and services per Edinburgh neighbourhoods

	neighborhood	American Restaurant	Aquarium	Argentinian Restaurant	Art Gallery	Art Museum	Arts & Crafts Store	Asian Restaurant	Auto Garage	Baby Store	...	Tram Station	Vegetarian / Vegan Restaurant	Whisky Bar	Wine Bar	Wine Shop
0	A702, Bruntsfield, Morningside, Fairmilehead	0.00	0.000000	0.00	0.000000	0.00	0.0	0.00	0.00	0.0	...	0.0	0.000000	0.000000	0.000000	0.2
1	A71, Haymarket, Gorgie, Stenhouse, Sighthill, ...	0.00	0.000000	0.00	0.000000	0.00	0.0	0.00	0.25	0.0	...	0.0	0.000000	0.000000	0.000000	0.0
2	A8, Murrayfield, Corstorphine, Gyle	0.00	0.000000	0.00	0.000000	0.00	0.0	0.00	0.00	0.0	...	0.0	0.000000	0.000000	0.000000	0.0
3	Colinton, Oxgangs	0.00	0.000000	0.00	0.000000	0.00	0.0	0.00	0.00	0.0	...	0.0	0.000000	0.000000	0.000000	0.0
4	Dean Village, Comely Bank, A90, Barnton, Cramo...	0.00	0.000000	0.00	0.000000	0.00	0.0	0.00	0.00	0.0	...	0.0	0.000000	0.000000	0.000000	0.0
5	Gilmerton, Moredun, Mortonhall	0.00	0.000000	0.00	0.000000	0.00	0.0	0.00	0.00	0.0	...	0.0	0.000000	0.000000	0.000000	0.0
6	Granton, Firth of Forth, Ferry Road	0.00	0.000000	0.00	0.000000	0.00	0.2	0.00	0.00	0.0	...	0.0	0.000000	0.000000	0.000000	0.0
7	Jobcentre Plus	0.00	0.000000	0.00	0.000000	0.00	0.0	0.00	0.00	0.1	...	0.1	0.000000	0.000000	0.000000	0.0
8	Kirkliston	0.00	0.000000	0.00	0.000000	0.00	0.0	0.00	0.00	0.0	...	0.0	0.000000	0.000000	0.000000	0.0
9	Leith, Newhaven	0.00	0.000000	0.00	0.037037	0.00	0.0	0.00	0.00	0.0	...	0.0	0.037037	0.037037	0.037037	0.0
10	Liberton, Cameron Toll, Craigmillar, Niddrie	0.00	0.000000	0.00	0.000000	0.00	0.0	0.00	0.00	0.0	...	0.0	0.000000	0.000000	0.000000	0.0
11	Marchmont, Grange	0.00	0.000000	0.00	0.000000	0.00	0.0	0.00	0.00	0.0	...	0.0	0.000000	0.000000	0.000000	0.0
12	New Town, Princes Street	0.01	0.000000	0.01	0.040000	0.01	0.0	0.01	0.00	0.0	...	0.0	0.010000	0.020000	0.000000	0.0
13	Newbridge, Ratho	0.00	0.000000	0.00	0.000000	0.00	0.0	0.00	0.00	0.0	...	0.0	0.000000	0.000000	0.000000	0.0
14	Old Town, GPO, St. James Centre	0.00	0.000000	0.01	0.030000	0.00	0.0	0.00	0.00	0.0	...	0.0	0.010000	0.030000	0.000000	0.0
15	Portobello, Duddingston	0.00	0.000000	0.00	0.000000	0.00	0.0	0.00	0.00	0.0	...	0.0	0.000000	0.000000	0.000000	0.0
16	Queen Street, Stockbridge, West End, Tollcross...	0.01	0.000000	0.01	0.020000	0.01	0.0	0.00	0.00	0.0	...	0.0	0.010000	0.010000	0.000000	0.0

Figure 11. Businesses and Services Types frequency per Edinburgh neighbourhoods

	neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue
0	A702, Bruntsfield, Morningside, Fairmilehead	Fish & Chips Shop	Hotel	Wine Shop	Bar
1	A71, Haymarket, Gorgie, Stenhouse, Sighthill, ...	Café	Discount Store	Skate Park	Auto Garage
2	A8, Murrayfield, Corstorphine, Gyle	Zoo Exhibit	Coffee Shop	Café	Grocery Store
3	Colinton, Oxgangs	Supermarket	Coffee Shop	Bowling Alley	Forest
4	Dean Village, Comely Bank, A90, Barnton, Cramo...	Indian Restaurant	Trail	Zoo Exhibit	Diner

Figure 12. 10 most common venues per Edinburgh Neighbourhood

For the Glasgow dataset, a set of data set information, and data processing about businesses and services were obtained.

Figure 13, 14, 15 and 16 are samples of these datasets.

	neighborhood	neighborhood latitude	neighborhood longitude	venue	venue latitude	venue longitude	venue category
0	Merchant City	55.86038	-4.24671	DogHouse Merchant City	55.859377	-4.247850	Beer Bar
1	Merchant City	55.86038	-4.24671	Hutchesons Glasgow	55.859800	-4.247892	Steakhouse
2	Merchant City	55.86038	-4.24671	Spitfire Espresso	55.859456	-4.245130	Coffee Shop
3	Merchant City	55.86038	-4.24671	Italian Kitchen	55.859361	-4.243795	Italian Restaurant
4	Merchant City	55.86038	-4.24671	iCafe	55.859379	-4.244422	Coffee Shop

Figure 13. Data for one of Glasgow neighbourhoods

neighborhood	neighborhood latitude
Alexandria, Arrochar, Aldochlay, Ardlui, Balloch, Bonhill, Gartocharn, Inverarnan, Jamestown, Luss, Tarbet	3
Anderston, Finnieston, Garnethill, Park, Woodlands, Yorkhill	40
Anniesland, Knightswood, Yoker	2
Arden, Carnwadric, Deaconsbank, Giffnock, Kennishead, Thornliebank	3
Auldhouse, East Kilbride	2
Baillieston, Bargeddie, Chryston, Garrowhill, Gartcosh, Gartloch, Moodiesburn, Muirhead, Springhill	1
Baldernock, Milngavie, Mugdock	11
Balfron, Balmaha, Blanefield, Croftamie, Drymen, Dumgoyne, Fintry, Killearn, Rowardennan, Strathblane	2
Balornock, Barmulloch, Cowlairs, Royston, Springburn, Sighthill	2
Barrhead, Neilston, Uplawmoor	2
Battlefield, Govanhill, Mount Florida, Strathbungo, Toryglen	5
Bearsden	8
Bishopbriggs, Torrance	4
Blantyre, Cambuslang	1
Blythswood Hill, Anderston	100
Bowling, Old Kilpatrick	8
Bridgeton, Calton, Dalmarnock	2
Broomhill, Partick, Partickhill	30
Busby, Carmunnock, Clarkston, Eaglesham, Stamperland, Waterfoot	2
Calton, Cowcaddens, Kelvinbridge, Townhead, Woodlands, Woodside	23
Cardonald, Hillington, Penilee, Mosspark	10

Figure 14. Number of businesses and services per Glasgow neighbourhoods

	neighborhood	American Restaurant	Art Gallery	Asian Restaurant	Athletics & Sports	Auto Garage	BBQ Joint	Bagel Shop	Bakery	Bar	...	Tennis Court	Thai Restaurant	Theater	Toy / Game Store	Trail	Train Station
0	Alexandria, Arrochar, Aldochlay, Ardlui, Ballo...	0.000000	0.000000	0.000000	0.000	0.000000	0.000000	0.000	0.000000	0.000000	...	0.0	0.000000	0.000000	0.000000	0.000000	0.000000
1	Anderston, Finnieston, Garnethill, Park, Woodl...	0.000000	0.000000	0.000000	0.000	0.000000	0.000000	0.025	0.000000	0.150000	...	0.0	0.000000	0.025000	0.000000	0.000000	0.000000
2	Anniesland, Knightswood, Yoker	0.000000	0.000000	0.000000	0.000	0.000000	0.000000	0.000	0.000000	0.000000	...	0.0	0.000000	0.000000	0.000000	0.000000	0.000000
3	Arden, Carnwadric, Deaconsbank, Giffnock, Kenn...	0.000000	0.000000	0.000000	0.000	0.000000	0.000000	0.000	0.000000	0.000000	...	0.0	0.000000	0.000000	0.000000	0.000000	0.000000
4	Auldhouse, East Kilbride	0.000000	0.000000	0.000000	0.000	0.000000	0.000000	0.000	0.000000	0.500000	...	0.0	0.000000	0.000000	0.000000	0.000000	0.000000
5	Baillieston, Bargeddie, Chryston, Garrowhill, ...	0.000000	0.000000	0.000000	0.000	0.000000	0.000000	0.000	0.000000	0.000000	...	0.0	0.000000	0.000000	0.000000	0.000000	0.000000
6	Baldernock, Milngavie, Mugdock	0.000000	0.000000	0.000000	0.000	0.000000	0.000000	0.000	0.090909	0.090909	...	0.0	0.000000	0.000000	0.000000	0.090909	0.000000
7	Balfron, Balmaha, Blanefield, Croftamie, Dryme...	0.000000	0.000000	0.000000	0.000	0.000000	0.000000	0.000	0.000000	0.000000	...	0.0	0.000000	0.000000	0.000000	0.000000	0.000000
8	Balornock, Barmulloch, Cowlairs, Royston, Spri...	0.000000	0.000000	0.000000	0.000	0.000000	0.000000	0.000	0.000000	0.000000	...	0.0	0.000000	0.000000	0.000000	0.000000	0.500000
9	Barrhead, Neilston, Uplawmoor	0.000000	0.000000	0.000000	0.000	0.000000	0.000000	0.000	0.000000	0.000000	...	0.0	0.000000	0.000000	0.000000	0.000000	0.000000
10	Battlefield, Govanhill, Mount Florida, Strathb...	0.000000	0.000000	0.000000	0.000	0.000000	0.000000	0.000	0.000000	0.400000	...	0.2	0.000000	0.000000	0.000000	0.000000	0.200000

Figure 15. Businesses and Services Types frequency per Glasgow neighbourhoods

	neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Alexandria, Arrochar, Aldochlay, Ardlui, Ballo...	Ice Cream Shop	Castle	Harbor / Marina	English Restaurant	French Restaurant
1	Anderston, Finnieston, Garnethill, Park, Woodl...	Bar	Indian Restaurant	Café	Nightclub	Coffee Shop
2	Anniesland, Knightswood, Yoker	Playground	Bus Stop	Whisky Bar	Furniture / Home Store	Diner
3	Arden, Carnwadric, Deaconsbank, Giffnock, Kenn...	Park	Soccer Field	Shopping Mall	Whisky Bar	Fish & Chips Shop
4	Auldhouse, East Kilbride	Bar	Scottish Restaurant	Whisky Bar	French Restaurant	Food & Drink Shop

Figure 16. 10 most common venues per Glasgow Neighbourhood

3.3 Segmentation and Clustering (K-Means)

For the machine learning and data mining stage of this project, the segmentation and clustering process has been selected, and within it, the technique used is the K-Means.

K-Means is an unsupervised learning algorithm used for clustering problems, and popular for cluster analysis in data mining. Even though there are many models for clustering, the K-means model is one of the most used and simple, even though it simplicity, k-means is vastly used for clustering in many data science applications, especially useful when needed to discover insights from unlabelled data.

K-means is used in various real-world applications, for example: customer segmentation, understand what the visitors of a website are trying to accomplish, pattern recognition, and, data compression.

In this project, it is used to understand the data gathered for each city's neighbourhood, to understand businesses and services patterns, and to provide answers to the questions present as problem in this report.

The number of clusters used in this project is 5 (K value).

The datasets for Edinburgh and Glasgow used with the K-mean algorithm, were:

- Each city dataset that was previously grouped in rows by neighbourhood, and by taking the mean of the frequency of occurrence of each business/services category.
- A new data frame created by using the K-means cluster labels results and the top 10 venues for each city's neighbourhood.

Figure 17 and 18 describe the Edinburgh segmentation and clustering using K-means.

	postcode	latitude	longitude	post town	neighborhood	Cluster Labels	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	Commo
0	EH1	55.95243	-3.18840	EDINBURGH	Old Town, GPO, St. James Centre	0	Hotel	Café	Bar	Restaurant	Pub	Coffee Shop	Art Gallery	Whisky Bar	Ste
1	EH10	55.92077	-3.20984	EDINBURGH	A702, Bruntsfield, Morningside, Fairmilehead	0	Fish & Chips Shop	Hotel	Wine Shop	Bar	Café	Diner	Electronics Store	Donut Shop	
2	EH11	55.93387	-3.24867	EDINBURGH	A71, Haymarket, Gorgie, Stenhouse, Sighthill, ...	0	Café	Discount Store	Skate Park	Auto Garage	Diner	Electronics Store	Donut Shop	Dog Run	
3	EH12	55.94262	-3.27137	EDINBURGH	A8, Murrayfield, Corstorphine, Gyle	0	Zoo Exhibit	Coffee Shop	Café	Grocery Store	Gift Shop	Chinese Restaurant	Zoo	Asian Restaurant	Food & Dr
4	EH13	55.90788	-3.24144	EDINBURGH	Colinton, Oxgangs	0	Supermarket	Coffee Shop	Bowling Alley	Forest	Dive Bar	Event Service	Electronics Store	Donut Shop	

Figure 17. Clustering each Edinburgh neighbourhood groups

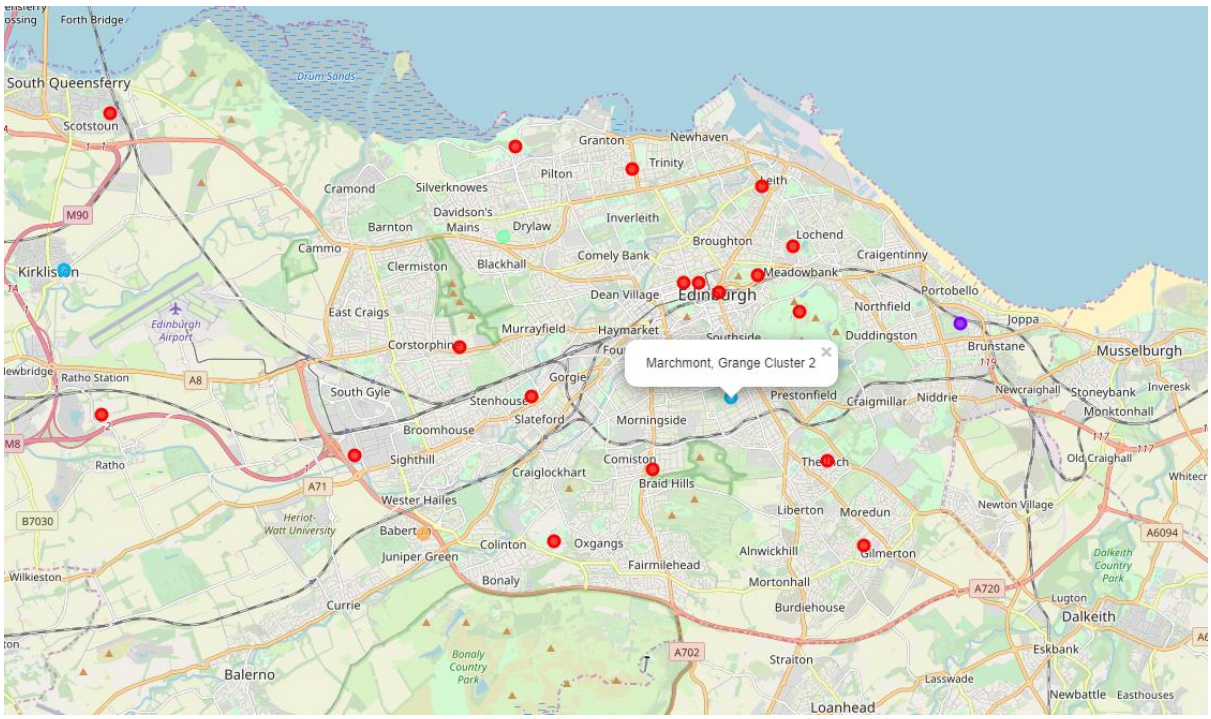

Figure 18. Mapping of the 5 clusters for the city of Edinburgh

Figure 19 and 20 describe the Glasgow segmentation and clustering using K-means.

	postcode	latitude	longitude	post town	neighborhood	Cluster Labels	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
0	G1	55.86038	-4.24671	GLASGOW	Merchant City	0	Coffee Shop	Bar	Pub	Italian Restaurant	Cocktail Bar	Seafood Restaurant	Steakhouse	Café
1	G11	55.87356	-4.31142	GLASGOW	Broomhill, Partick, Partickhill	0	Coffee Shop	Café	Deli / Bodega	Sandwich Place	Supermarket	Bus Station	Shopping Plaza	Pharmacy
2	G12	55.88006	-4.30061	GLASGOW	West End, Dowanhill, Hillhead, Hyndland, Kelvi...	0	Hotel	Convenience Store	Restaurant	Italian Restaurant	Gym	Discount Store	Diner	Dive Bar
3	G13	55.89358	-4.34620	GLASGOW	Anniesland, Knightswood, Yoker	0	Playground	Bus Stop	Whisky Bar	Furniture / Home Store	Diner	Discount Store	Dive Bar	Doner Restaurant
4	G14	55.88095	-4.34864	GLASGOW	Whiteinch, Scotstoun	0	Sports Bar	Bus Stop	Rugby Pitch	Spanish Restaurant	Whisky Bar	English Restaurant	Food & Drink Shop	Fish & Chips Shop

Figure 19. Clustering each Glasgow neighbourhood groups

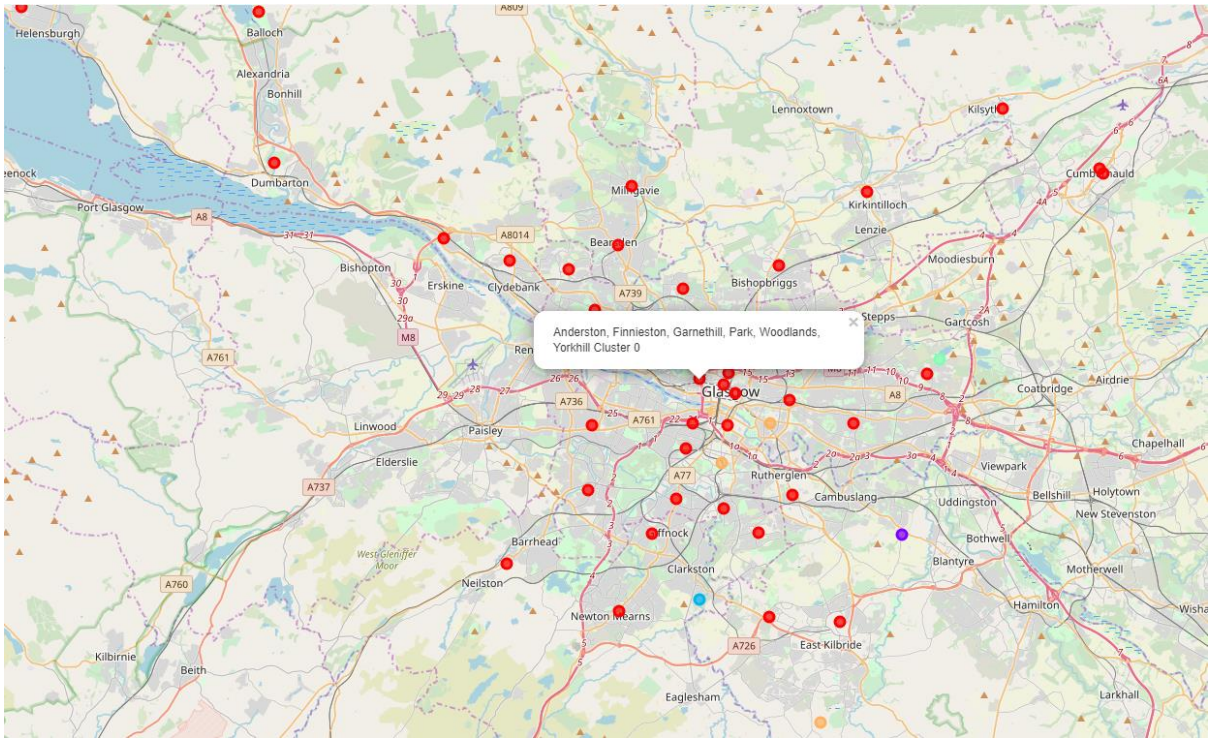

Figure 20. Mapping of the 5 clusters for the city of Glasgow

4 Results

In this section, a description is provided to explain the findings, based on the data science methods used with the available data of two major cities in Scotland, Edinburgh and Glasgow.

These results aims to provide an understanding on each cities based on their current businesses and services, and to provide the backing information to answer the questions raised as the problem section of this report.

4.1 Edinburgh

From the initial data set obtained for Edinburgh, 23 postcodes were identified that belong to the city of Edinburgh. Each postcode has a coverage of a group of neighbourhoods. This information was used as the basis for the analysis of the city of Edinburgh.

In order to obtain information about business and services of the city of Edinburgh, a data platform called Foursquare API was used. This data platform provides information of business and services (venues) for locations, access to this information require a valid subscription. A maximum of 100 venues per location is provided under this subscription from Foursquare. Additionally, it is required the longitude and latitude of each location to get information of their venues.

Due to the above requirement, it was needed the longitude and latitude of all the postcodes of Edinburgh, for this, a new dataset was obtained, this database only contained the list of postcode and their longitude and latitude.

The combined information of these two datasets provided the data required to obtain information about the venues for each postcode and neighbourhoods of Edinburgh.

From the data results and calculations, 23 neighbourhoods groups were analysed for Edinburgh, described below.

4.1.1 Business-thrived neighbourhoods in Edinburgh

Figure 21 shows the number of business and services for each neighbourhood group (based on 23 postcodes).

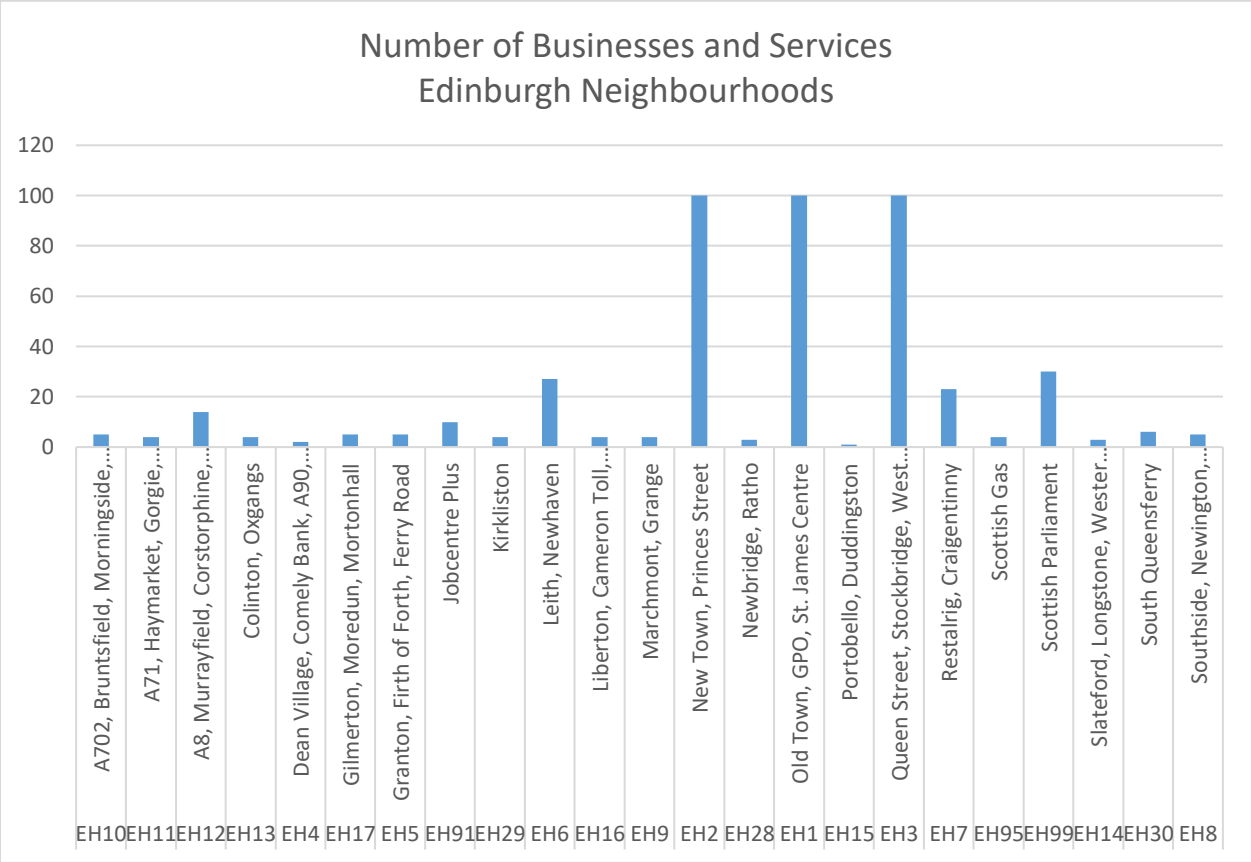

Figure 21. Number of Edinburgh Businesses and Services

From these results:

- 3 neighbourhood groups contains the largest number of business and services (max 100 venues). These are in postcodes: EH2 (New Town, Princes Street, Old Town), EH1 (GPO, St. James Centre), and EH3 (Queen Street, Stockbridge, West End, Tollcross, Fountainbridge).
- 4 neighbourhood groups contains a medium number of business and services (average 23.5 venues each). These are in postcodes: EH12 (A8, Murrayfield, Corstorphine, Gyle), EH6 (Leith, Newhaven), EH7 (Restalrig, Craigentinny) and EH99 (Scottish Parliament).
- 16 neighbourhood groups contains a small number of business and services (average 4.3 venues each). These are in postcodes: EH10, EH11, EH13, EH4, EH17, EH5, EH91, EH29, EH16, EH9, EH28, EH15, EH95, EH14, EH30, and EH8.

4.1.2 Common business types in Edinburgh's neighbourhoods

In this analysis, the findings indicated that the three most common venues in the 23 neighbourhood groups of Edinburgh are:

- Bares are the most popular and common venue type in Edinburgh, along with Hotels, which are the second most popular business type in Edinburgh, and Third is Café or Coffee Shops.
- After the three most popular businesses, other food business types are also popular in Edinburgh, such as Fish and Chips Shops, Indian and Chinese restaurants, and Food Trucks.

- Other business and services such as Supermarkets, Bus Stop, Train Station, Zoo, Parks, and other are also popular, and these are distributed evenly in all the neighbourhoods.

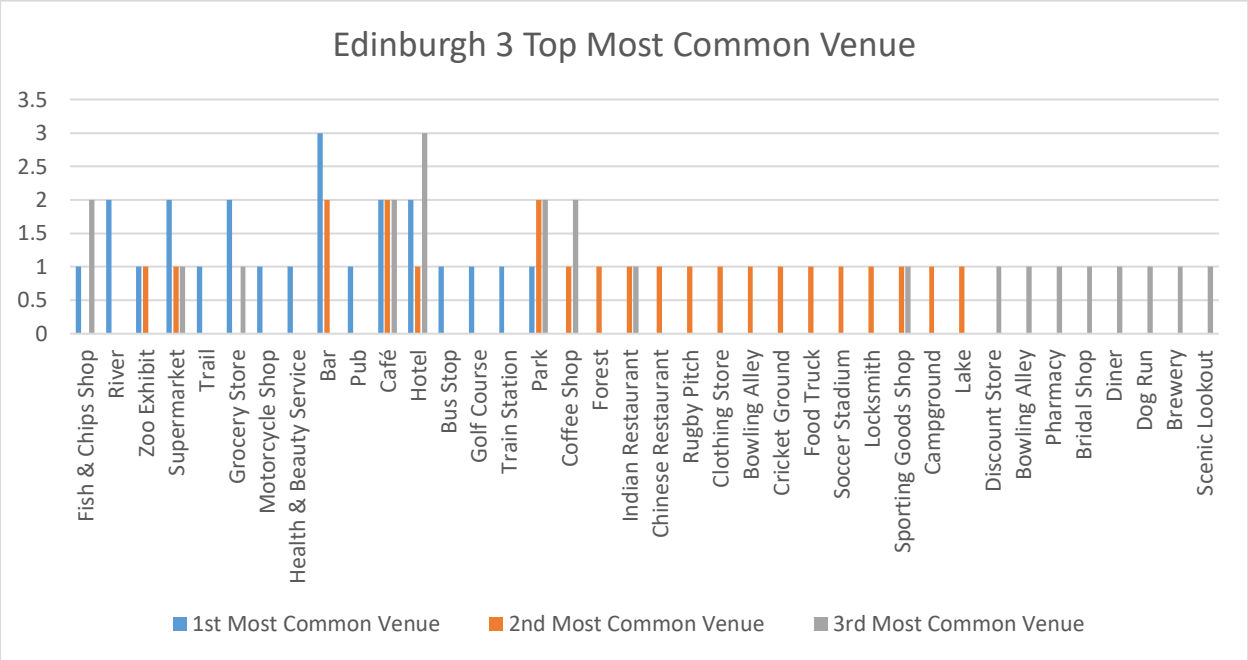

Figure 22. Edinburgh 3 top most common venue

From the analysis of grouping the results of the top most common venues (Figure 23) for all neighbourhoods. The following are the findings:

- The top venues or popular businesses are Hotels, Cafes (including Coffee Shops).
- Bares and Parks are the second most popular venues.
- Supermarkets, Grocery Shops are also popular in Edinburgh neighbourhoods.
- Businesses that are less common in Edinburgh, target niche markets such as Bridal Shop Brewery, Pharmacy, and Discount Stores.

From these results, people in Edinburgh have a variety of food-related venues, also bares, and cafes, and these are popular among the population.

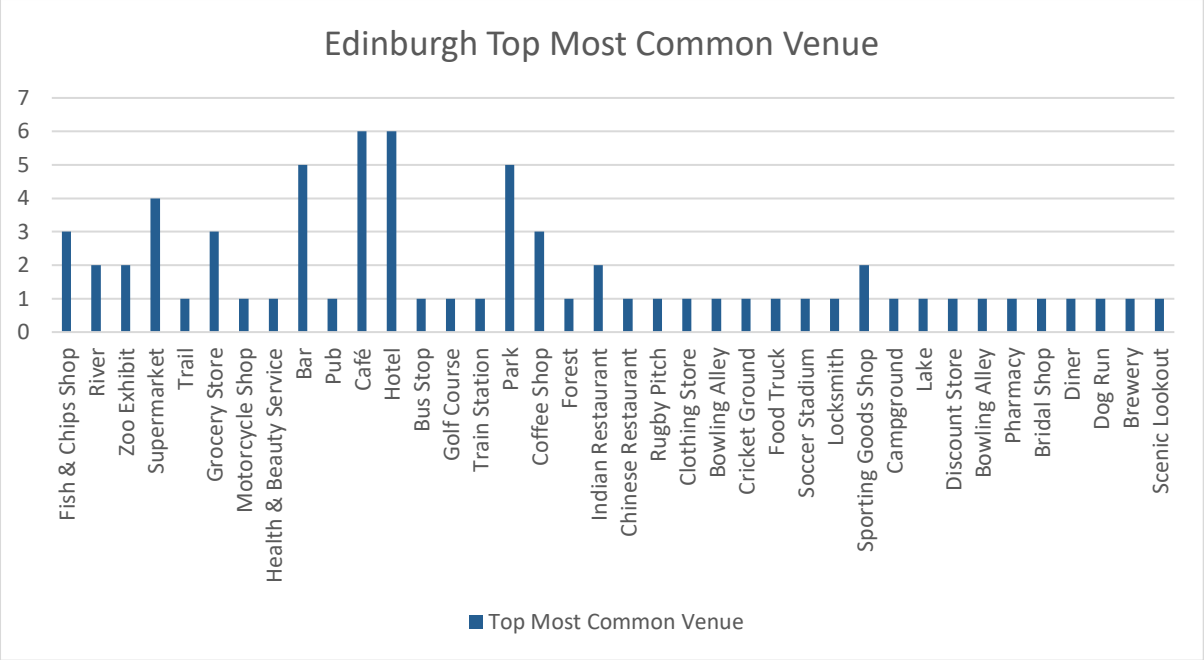

Figure 23. Edinburgh top most venues

4.1.3 Segmenting and clustering neighbourhoods

After clustering the Edinburgh neighbourhoods' data (Figure 24 and 25), it was identified the following findings:

- 5 clusters were identified from the data that describes the 10 most popular businesses per neighbourhood.
- The most popular cluster, was the group that included dining-out venues, hotels, bar, and restaurants, which means that the most popular business venue in the city of Edinburgh are social-entertainment related businesses among the local people.
- 18 neighbourhoods were clustered in the most popular type of venues (marked in red in the map Figure 24), which is spread around the city of Edinburgh neighbourhoods, and mostly in the closest the neighbourhood is to the city centre.
- 2 neighbourhoods are popular for outdoor activities, such as parks, lakes, site-seeing, these are in South Queensferry (EH30), and Southside, Newington, Canongate, Holyrood Park (EH8)
- 3 other neighbourhood, have as popular business venues bus stations, golf courses, and niche-target business such as bridal shop. These neighbourhoods are mostly located at the outskirts of the city of Edinburgh as seen in the map Figure 24, marked in green orange and light blue.

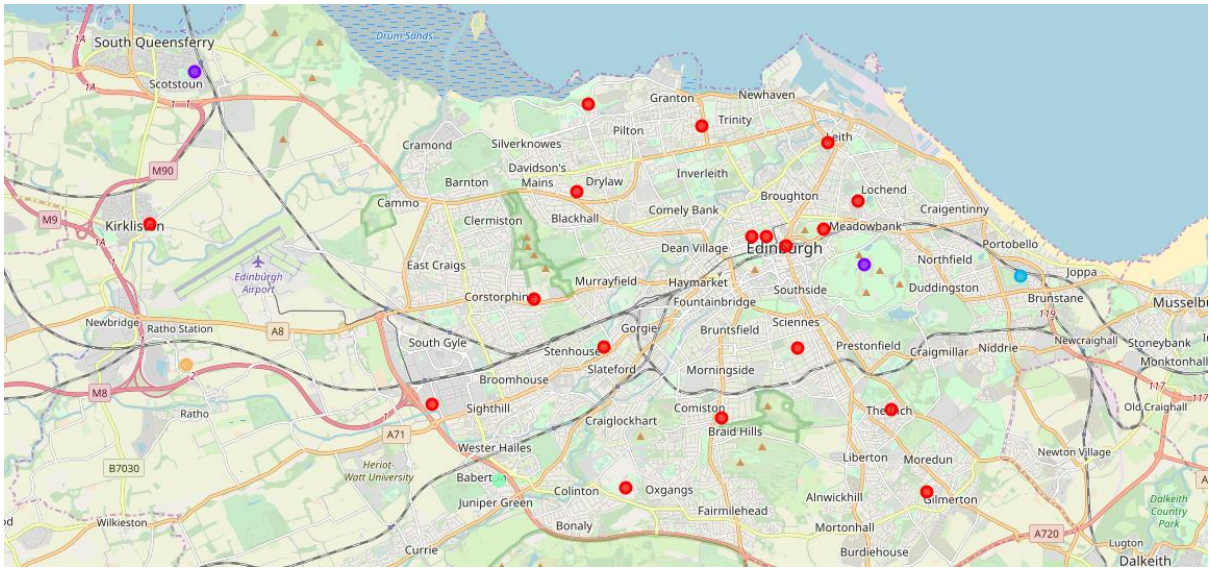

Figure 24. Map of Edinburgh Venues Clusters

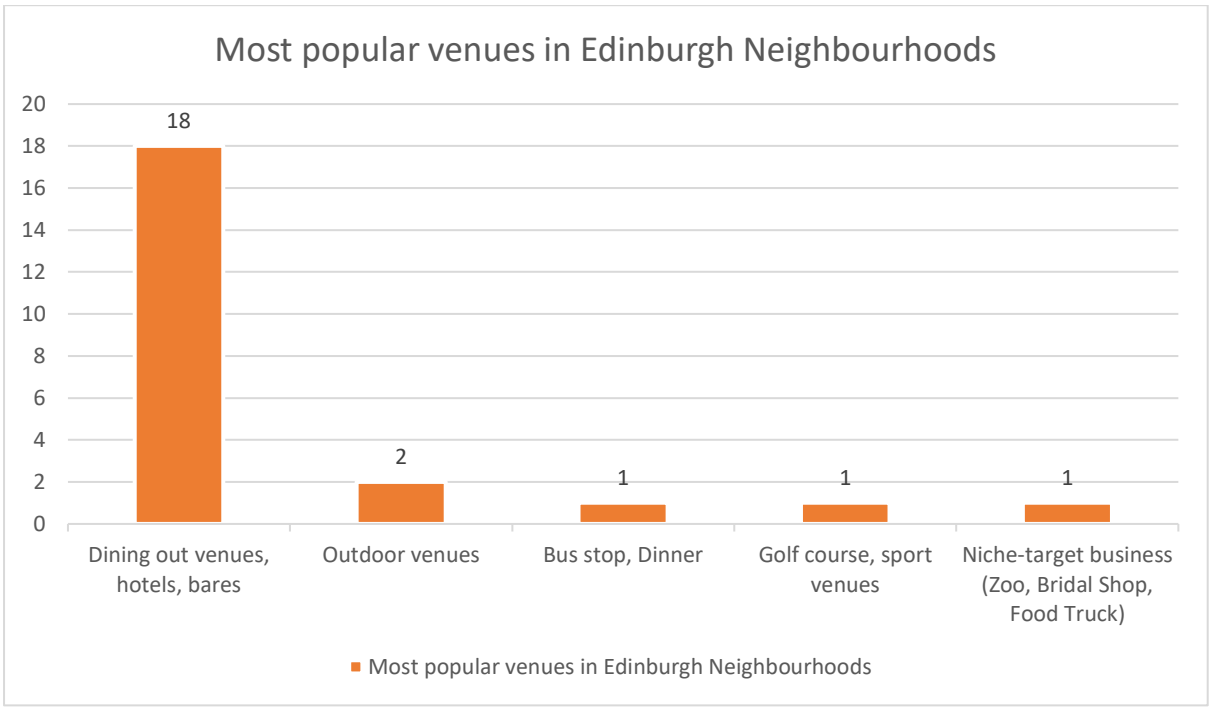

Figure 25. Most popular venues in Edinburgh clusters

4.2 Glasgow

From the initial data set obtained for Glasgow, 55 postcodes were identified that belong to the city of Glasgow. Each postcode has a coverage of a group of neighbourhoods. This information was used as the basis for the analysis of the city of Glasgow.

Similar to the Edinburgh data analysis, for the city of Glasgow, a data platform called Foursquare API was used. This data platform provides information of business and services (venues) for locations, access to this information require a valid subscription. A maximum of 100 venues per location is

provided under this subscription from Foursquare. Additionally, it is required the longitude and latitude of each location to get information of their venues.

Due to the above requirement, it was needed the longitude and latitude of all the postcodes of Glasgow, for this, a new dataset was obtained, this database only contained the list of postcode and their longitude and latitude.

The combined information of these two datasets provided the data required to obtain information about the venues for each postcode and neighbourhoods of Glasgow.

From the data results and calculations, 55 neighbourhoods groups were analysed for Glasgow, however only 50 neighbourhood groups provided available venue data, therefore these are the only ones analysed for the city of Glasgow, as described below.

4.2.1 Business-thrived neighbourhoods in Glasgow

Figure 26 shows the number of business and services for each neighbourhood group (based on 50 postcodes).

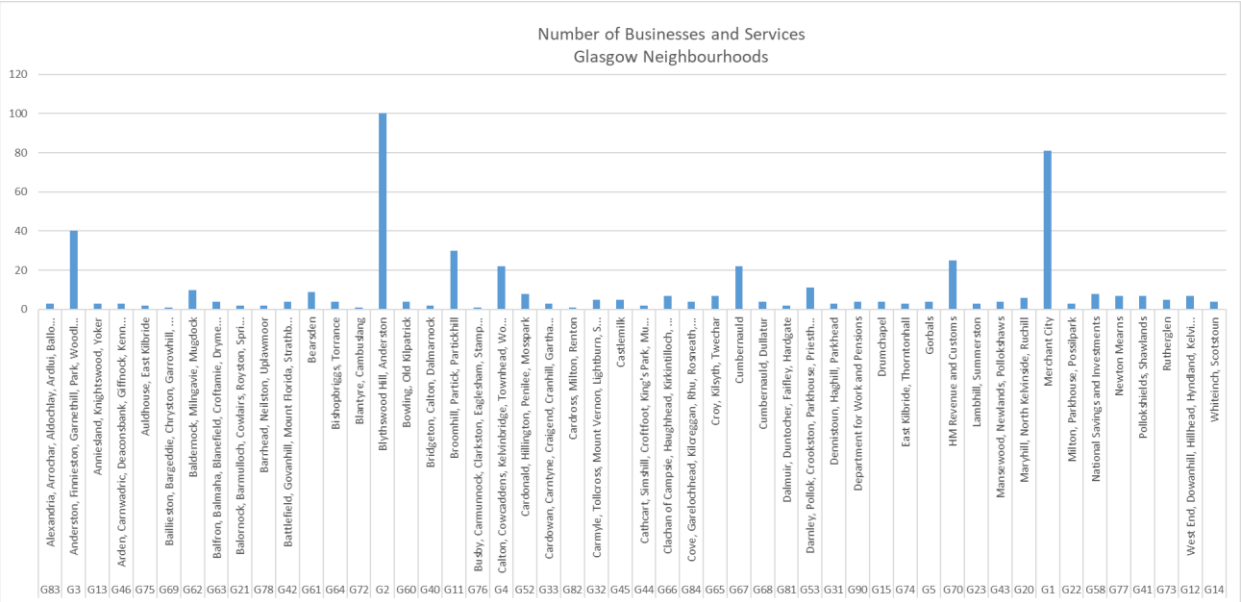

Figure 26. Number of Glasgow Businesses and Services

From these results:

- Only 1 neighbourhood group contains the largest number of business and services (max 100 venues) in postcode: G2 (Blythswood Hill, Anderston).
- The second largest neighbourhood group for business and services is located in G1 (Merchant City), with 81 venues.
- The third largest neighbourhood group for business and services I located in G3 (Anderston, Finnieston, Garnethill, Park and others) with 40 venues.
- 4 neighbourhood groups contains an average of 24.75 venues. These are in postcodes: G11 (Broomhill, Partick, Partickhill), G4 (Calton, Cowcaddens, Kelvinbridge, Townhead, Woodlands, Woodside), G67 (Cumbernauld), G70 (HM Revenue and Customs).
- 2 neighbourhood groups contains 11 venues for business and services, these are G62 (Baldernock, Milngavie, Mugdock), and G31 (Dennistoun, Haghill, Parkhead).
- 41 neighbourhood groups contains 10 or less business venues in their areas.

4.2.2 Common business types in Glasgow's neighbourhoods

In this analysis, the findings indicated that the three most common venues in the 50 neighbourhood groups of Glasgow are:

- Bares are the most popular and common venue type in Glasgow, follow by Grocery Stores, Hotels, and Fast food restaurants, and thirdly are Supermarkets and Café or Coffee Shops.
- After the three most popular businesses, other food business types are also popular in Glasgow, such as Pubs, Indian and Italian restaurants, and others type of restaurants.
- Other business and services such as Playgrounds, Gyms, Shopping malls, Golf courses, Museums, which are distributed evenly in all the neighbourhoods.

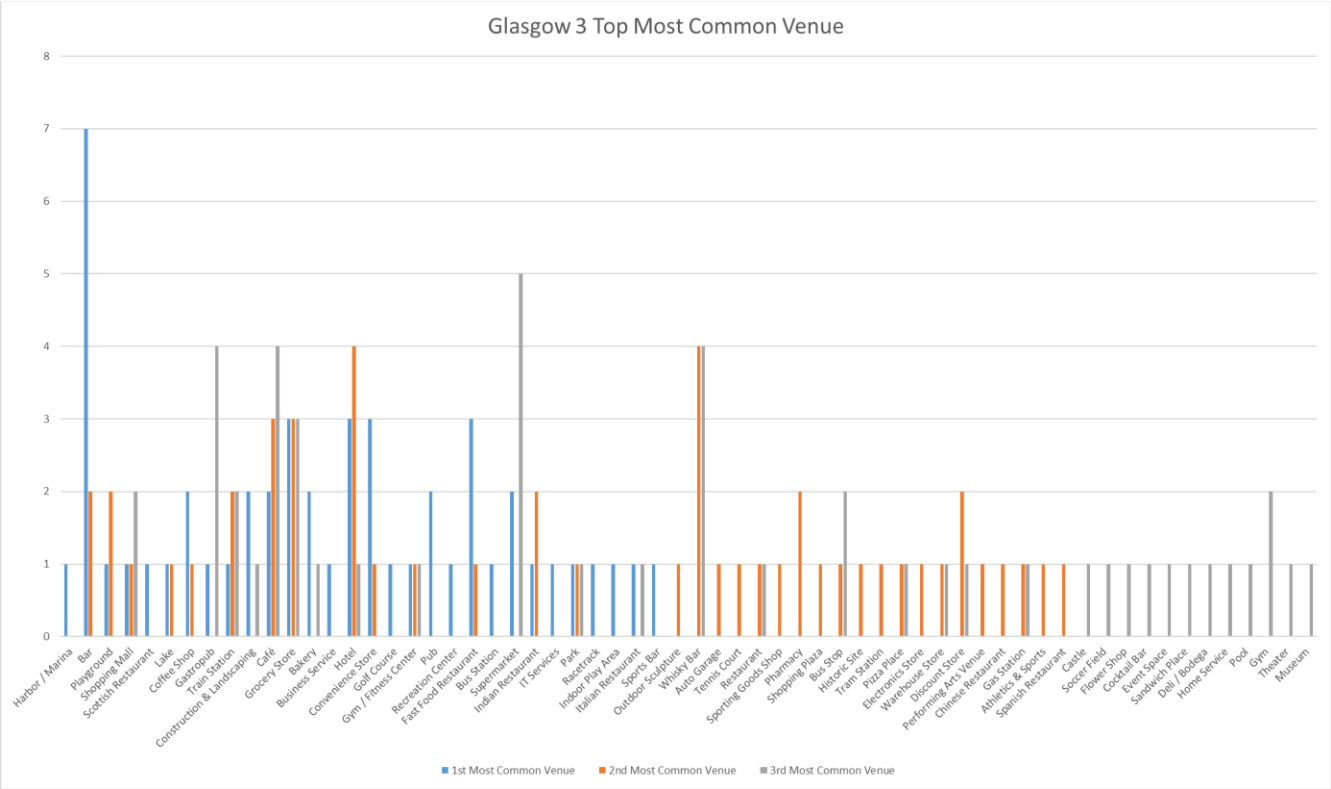

Figure 27. Glasgow 3 top most common venue

From the analysis of grouping the results of the top most common venues (Figure 28) for all neighbourhoods. The following are the findings:

- The top venues or popular businesses are Bars, Cafes (including Coffee Shops) and Grocery Stores.
- Hotels, whisky bars and supermarkets are the second most popular venues.
- Shopping Malls, Convenience Stores, are also popular in Glasgow neighbourhoods.
- Businesses that are less common in Glasgow are Gyms, Pools, Museum, Theatres, and niche-markets such as flower shops.

From these results, people in Glasgow have a variety of entertainment-related venues, including bares, and food-related venues, and these are popular among the population.

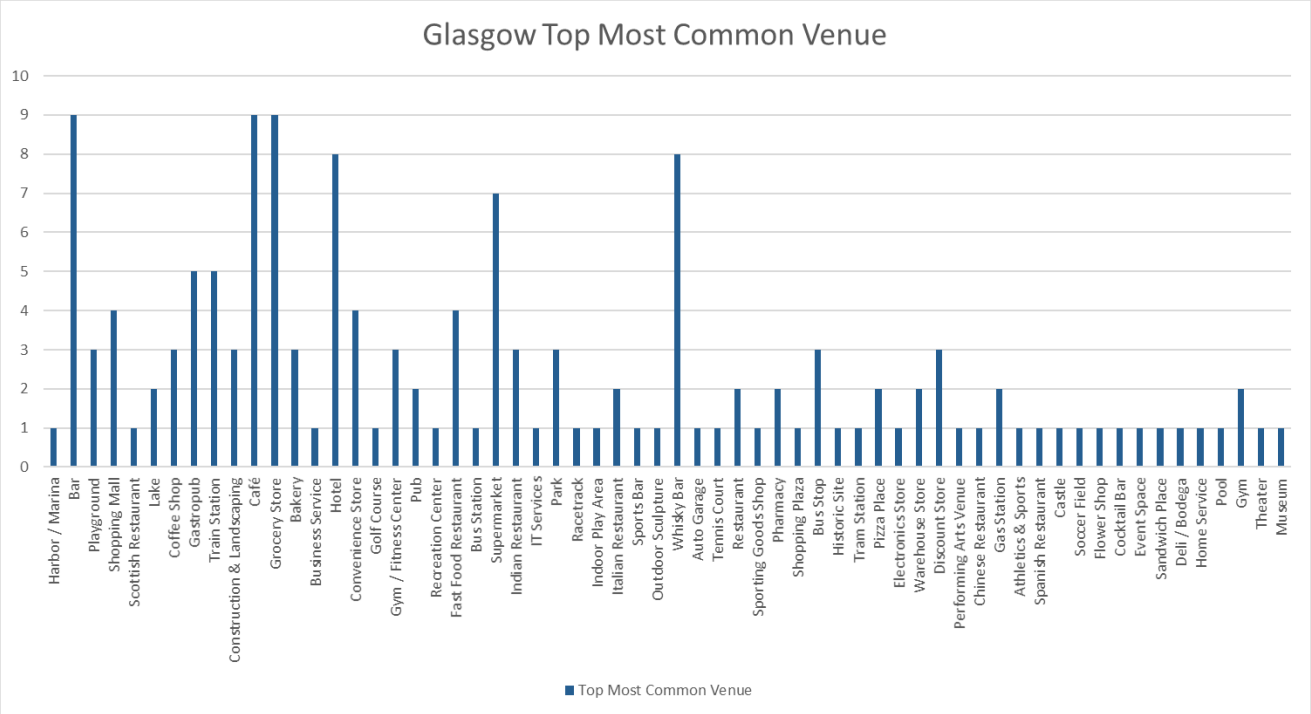

Figure 28. Glasgow top most venues

4.2.3 Segmenting and clustering neighbourhoods

After clustering the Glasgow neighbourhoods' data (Figure 29 and 30), it was identified the following findings:

- 5 clusters were identified from the data that describes the 10 most popular businesses per neighbourhood.
- The most popular cluster, was the group that included entertainment and recreational venues (e.g. bars, pubs, playgrounds, malls, gyms, golf courses), which means that the most popular business venue in the city of Edinburgh are social-entertainment related businesses among the local people.
- 27 neighbourhoods were clustered in the most popular type of venues (marked in purple in the map Figure 29), which is spread around the city of Glasgow neighbourhoods, and mostly the closest the neighbourhood is to the city centre.
- 19 neighbourhoods are popular for dining out venues, hotels, cafes (coffees shops), these are marked in red in the map Figure 29.
- 3 neighbourhood groups have as popular business venues Stores (Grocery, Convenience), bakery. These neighbourhoods are mostly located at the outskirts of the city of Glasgow as seen in the map Figure 29, marked in green.
- 1 neighbourhood group has business like venues as the most popular, located in G76 (Busby, Carmunnock, Clarkston, Eaglesham, Stamperland).
- 1 neighbourhood group has Outdoor venues as the most popular, located in G69 (Baillieston, Bargeddie, Chryston, Garrowhill, Gartcosh, Gartloch, Moodiesburn, Muirhead, Springhill).

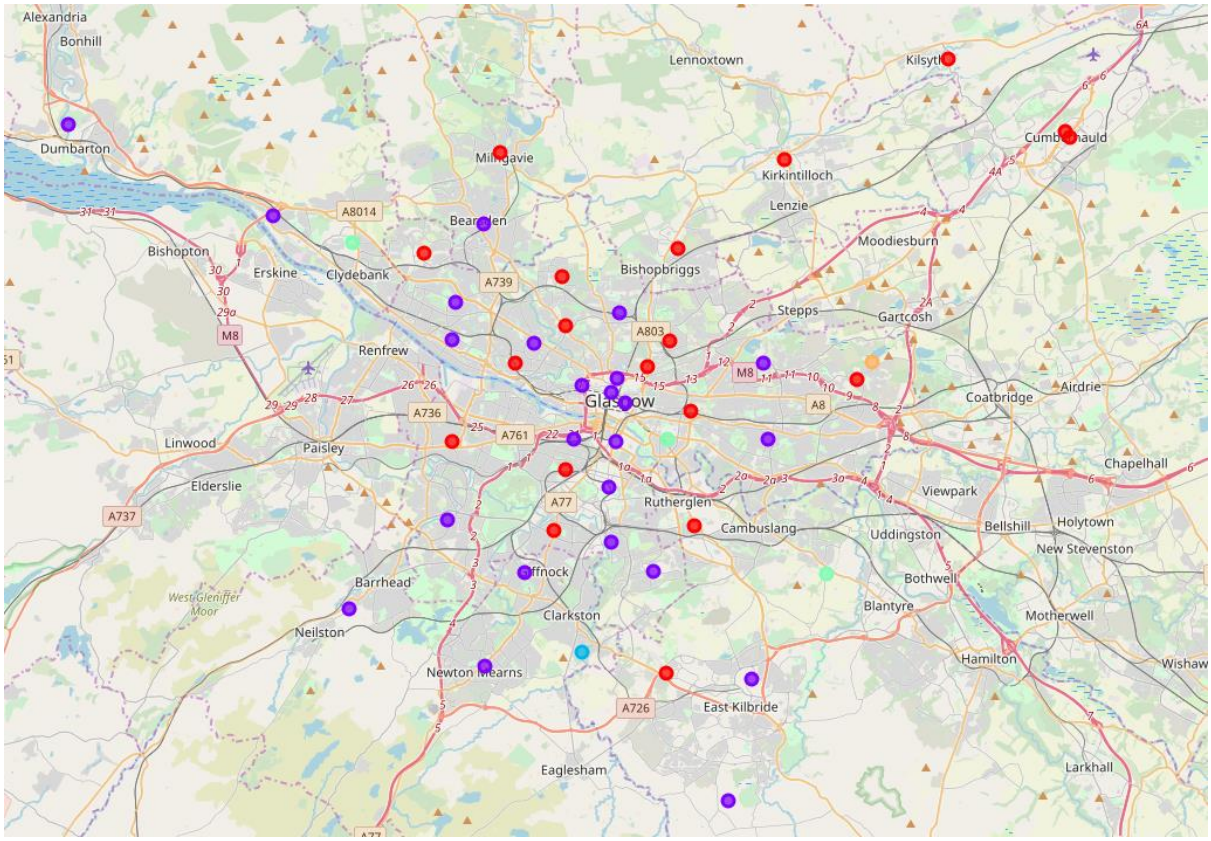

Figure 29. Map of Glasgow Venues Clusters

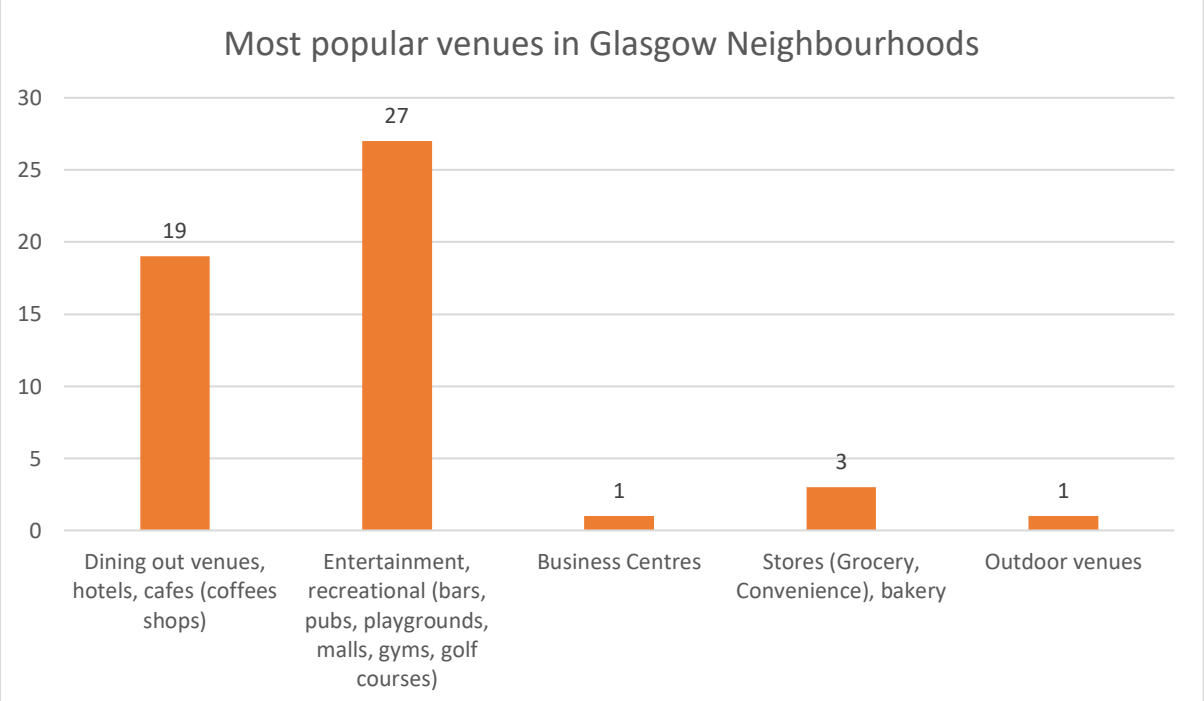

Figure 30. Most popular venues in Glasgow clusters

5 Discussion

This section aims to provide answers to the questions raised in the problem section, and from the data analysis formulate the findings to explain these results.

One of the main objectives of this project is to understand how wide is the difference of both cities (Edinburgh and Glasgow), their neighbourhoods similarities and differences, the top businesses that are flourishing (in numbers) in those cities, an understanding of the population and their behaviour pattern, along with type of businesses that are popular (and the least ones) in their neighbourhoods

Edinburgh and Glasgow, are two major cities of Scotland, it is clear that Glasgow as the largest city in Scotland provides a large data set of business types compare to Edinburgh, however based on the results obtained, it was clear that the difference between these two cities are not as wider as expected.

The most popular business and services provided by both cities are of similar nature, dining-out venues, restaurants, bars, pubs, cafes and hotels. This is clearly explain, as both are major touristic destinations of Scotland.

The difference, explained from the data, is in the scale and number of neighbourhoods and target people.

Edinburgh, as being the main historical and touristic destination of Scotland, has a preference for business related to the tourism industry, therefore the first and second most popular business types are dining-out, entertainment, outdoors venues, sightseeing, parks and lakes that boost their tourism industry, and thirdly as any big city, services and businesses that target their local population, such as supermarkets, niche-target businesses, grocery shops and others.

On the other hand, Glasgow as being a larger city, and the more residential city between them, although coincide with Edinburgh in entertainment and dining out venues as the most popular business and services, however it expands its top interest in venues like supermarkets, malls and playgrounds, therefore orientated more to entertain their local residents. The second most popular business venues includes hotels, cafes, and dining-out venues. Follow by other niche-businesses and outdoor places.

Both cities offer great business and services to their population and visitors. If a new investor is interested in setup a business in any of these cities, probably targeting tourism could be ideal in a city like Edinburgh, and targeting the entertainment of local population could be ideal for Glasgow.

Additionally, in both cities, the number of businesses increase the closer is to the city centre, therefore if new business and service is developed, it will be ideal to target areas that are further than the city centre, however for this it is necessary to understand the size of the population and age targets, which are not covered as part of this analysis. Businesses in the city centre due to the larger number will experience higher competition in both cities.

6 Conclusions

This report describes the data analysis of two major cities in Scotland, Edinburgh and Glasgow, and tried to answer the questions of their similarities and differences based on their neighbourhood groups and business-services types. This was achieved with the data that was collected, processed and analysed using the data sciences techniques and methods.

From the data analysis, it was understood the similarities of businesses between Edinburgh and Glasgow, but also their difference, and target customers for each of them. On one side, Edinburgh resulted a more tourism-focus due to its high tourism industry, while Glasgow as being a large city for Scottish residents aims to target the local population largely.

Although the questions to the problem were answered, there were other questions raised when analysing the data, such as, how the population is segmented, what is the difference in age and gender between the two cities, and others, such as size of population and incomes, those could have helped to understand the nature of existing businesses, potential new ones, their numbers and popularities. The answers to these additional questions could help to provide better insights in the understudying, and reasons behind the popularity and most common business and service types, along with the least common ones, also to provide a better insight for any new development or business growth in the area. Hence, although the first steps to analyse the difference of these two cities were provided and achieved in this report, further gathering of information, and analysis of data, will bring more answers, better understanding of patterns present in these two cities, predicting their growth, and others, these could be expanded in future work.