

# Capstone Project - The Battle of Neighbourhoods

## Introduction

In the city of Glasgow, Scotland, identify the most likely location for a new restaurant.

Choosing a location for any new restaurant is very important and, once chosen, can be expensive and time consuming to change later on.

Tips on how and where to start a new restaurant can be found on line (see example sites below), citing various factors to take into consideration when choosing the location of a new restaurant.

One of these factors is other venues in the area. The consensus being that siting a new restaurant in an untried location is a gamble but, equally, choosing a location with too many restaurants with the same concept can over saturate an area, leading to potential business failure.

- How to Open a Restaurant, Part 2: Location, Location, Location (<https://openforbusiness.opentable.com>)
- 8 Factors for Choosing a New Restaurant Location (<https://www.foodnewsfeed.com>)
- Tips on Where to Locate Your Restaurant (<https://www.thebalancesmb.com>)

It is, therefore, very important to analyse an area for the number and type of different restaurants, to assist a new restaurant owner in determining a good location for their restaurant and avoiding those areas where their restaurant would be more likely to fail.

Equally, any type of new business enterprise would benefit from analysing where to locate their premises, in order to determine the best location for their customers to find them.

## Business Problem

Using cluster analysis on Foursquare location data can we assist a new restaurant owner to identify potential 'goldilocks' neighbourhoods that are both not untried nor over-saturated and, therefore, potentially viable locations for their new venture.

For the purposes of this project, the value of oversaturation may need to be adjusted but will be initiated at 6 restaurants already in the same area.

Additional specialisation may be required to differentiate between different types of restaurant within a neighbourhood.

## Data Sources

- It will be necessary to obtain a list of neighbourhood names and associated postcodes. These can be scraped from the Wikipedia page [https://en.wikipedia.org/wiki/G\\_postcode\\_area](https://en.wikipedia.org/wiki/G_postcode_area). The data will need to be cleaned to remove un-necessary columns and extra text within the neighbourhood's field. There are also a few postcodes that are labelled 'non-geographic', which will also need to be removed from the dataset. Finally, the data will be restricted to those postcodes that are within the local authority area of 'Glasgow City' as the other local authority areas would give to great an physical area in which to locate the new restaurant.

Resulting in:

Postcode	Neighbourhood
G1	Merchant City
G2	Blythswood Hill, Anderston
G3	Anderston, Finnieston, Garnethill, Park, Woodlands, Yorkhill
G4	Calton, Cowcaddens, Drygate, Kelvinbridge, Townhead, Woodlands, Woodside
G5	Gorbals

- Due to the buggy nature of geocoder, a CSV file for the longitudes and latitudes for each postcode will be manually generated and uploaded for use in this project:

Postcode	Latitude	Longitude
G1	55.8586	4.2456
G2	55.8636	4.2612
G3	55.8644	4.2780
G4	55.8710	4.2494
G5	55.8433	4.2450

- The Foursquare location data will then be leveraged to identify and locate other venues in each of the neighbourhoods. This will be restricted to other restaurants and cafes as the objective is to recommend the location of a new restaurant.

This data can then be scraped to give the restaurant information for a particular area:

Index	name	categories	address	city	country	lat	lng	postalCode	state	id
0	Jamie's Italian	Italian Restaurant	7 George Sq	Glasgow	United Kingdom	55.860789	- 4.250206	G2 1DY	Glasgow City	4c2d1e523a4d0f47d756e386
1	ASK Italian	Italian Restaurant	3 Port Dundas Place	Glasgow	United Kingdom	55.865036	- 4.253473	G2 3LD	Glasgow City	4ef3298d77166617fd74f5ac
2	Italian Centre	Shopping Mall	7 John St.	Glasgow	United Kingdom	55.859982	- 4.247694	G1 1HP	Glasgow City	4c191f436a21c9b66f12ca97
3	ASK Italian	Italian Restaurant	31 Royal Exchange Square	Glasgow	United Kingdom	55.860021	- 4.252045	G1 3AJ	Glasgow City	53727f5111d211359f739058
4	Caffe Italiano	Café	Buchanan Galleries	Glasgow	United Kingdom	55.864611	- 4.254989	NaN	Glasgow City	52f39962498e1c70065f7fac
5	Centro Italiano	Italian Restaurant	NaN	NaN	United Kingdom	55.859824	- 4.247916	NaN	NaN	4c2cea6bb34ad13a4514ecc

- Using the venue ID, a second call to Foursquare can retrieve the rating for that venue, for example, Jamie's Italian is 7.3 and The ASK Italian in Port Dundas Place is 6.6. This could be useful additional information for deciding the suitability of a location for a new restaurant, as if there is a highly popular venue of the same category in an area, their customer's may not be tempted away to a new restaurant. However, these are premium calls to Foursquare, which are limited per day. Once a suitable 'goldilocks' neighbourhood has been identified, this additional step may be useful to confirm its viability.

## Methodology

The Methodology is separated into 3 main sections:

1. Retrieving and Cleaning Neighbourhood data for Glasgow
2. Retrieving Venue Details for Glasgow
3. Distribution Analysis of Glasgow Venues

## Retrieving and Cleaning Neighbourhood data for Glasgow

Postcode data for the Glasgow area was retrieved from Wikipedia<sup>1</sup> and scraped into a Pandas dataframe using the BeautifulSoup<sup>2</sup> Python library.

Postcode district	Post town	Coverage	Local authority area
G1	GLASGOW	Former C1 district: Merchant City	Glasgow City
G2	GLASGOW	Former C2 district: Blythswood Hill, Anderston (part)	Glasgow City
G3	GLASGOW	Former C3 district: Anderston, Finnieston, Garnethill, Park, Woodlands (part), Yorkhill	Glasgow City
G4	GLASGOW	Former C4 district: Calton (part), Cowcaddens (part), Drygate, Kelvinbridge, Townhead, Woodlands (part), Woodside (part)	Glasgow City
G5	GLASGOW	Former C5 district: Gorbals	Glasgow City
G9	GLASGOW	Non-geographic - Newspaper Competitions	Glasgow City
...			

The list included postcodes that were non-geographical, such as G9 above, would not contain any venues and, therefore, not be of any relevance to this project. These could easily be removed by dropping any rows where the entry in the Coverage column started with “Non-geographic”. Also, a lot of the Coverage entries contained extra text in the form “Former XX district:”. This text was removed by checking each row in the Coverage column and, if ‘:’ was present (position i), slicing the string in that cell from i+2 to the end.

Former C1 district: Merchant City

0 i i+2

Finally, the columns “Post town” and “Local authority area” were removed as unnecessary and the remaining two columns were relabelled “Postcode” and “Neighbourhood”:

Postcode	Neighbourhood
G1	Merchant City
G2	Blythswood Hill, Anderston (part)
G3	Anderston, Finnieston, Garnethill, Park, Woodlands (part), Yorkhill
G4	Calton (part), Cowcaddens (part), Drygate, Kelvinbridge, Townhead, Woodlands (part), Woodside (part)
G5	Gorbals
...	

<sup>1</sup> [https://en.wikipedia.org/wiki/G\\_postcode\\_area](https://en.wikipedia.org/wiki/G_postcode_area)

<sup>2</sup> <https://www.crummy.com/software/BeautifulSoup/>

However, the 'G' postcode covers a much wider area than the city of Glasgow, as can be seen below. Therefore, only G1 to G58 and G73 were included as being relevant for the purposes of this project; suggesting a person open a new restaurant in Dumbarton would not be beneficial as it is 22 miles from Glasgow.



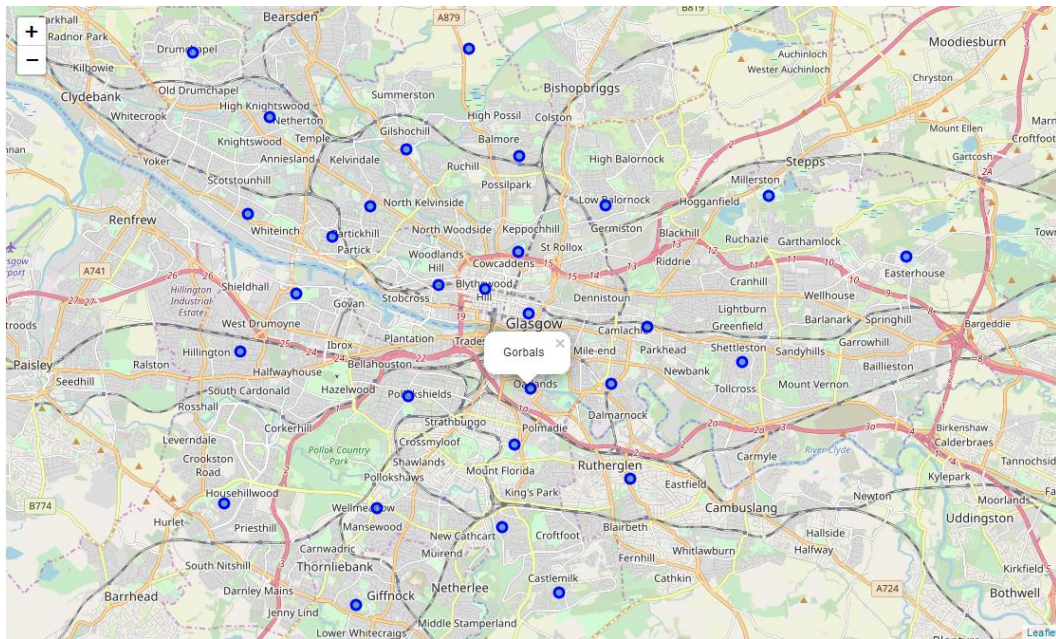
**Fig 1.** G postcode area map, taken from the Wikipedia page

The scraped data having been cleaned, the next step was to retrieve latitudes and longitudes for each postcode. This was attempted using Geocoder<sup>3</sup> but it was found to be buggy and, as the dataset in this project was deemed small enough, a .csv file was manually created. This was then merged with the scraped data to form the Neighbourhood dataframe.

Postcode	Neighbourhood	Latitude	Longitude
G1	Merchant City	55.8586	-4.2456
G2	Blythswood Hill, Anderston (part)	55.8636	-4.2612
G3	Anderston, Finnieston, Garnethill, Park, Woodlands (part), Yorkhill	55.8644	-4.2780
G4	Calton (part), Cowcaddens (part), Drygate, Kelvinbridge, Townhead, Woodlands (part), Woodside (part)	55.8710	-4.2494
G5	Gorbals	55.8433	-4.2450
...			

<sup>3</sup> <https://geocoder.readthedocs.io/>

And could be plotted on a map using the Folium Python library<sup>4</sup> to confirm the distribution was correct



**Fig 2.** Distribution of G postcodes within Glasgow

#### Retrieving Venue Details for Glasgow

To retrieve restaurant information within each neighbourhood, the Foursquare<sup>5</sup> search API was employed. The radius was determined by measuring the distance between adjacent postcodes<sup>6</sup> and taking half of the average. This was found to be 1300m, rounded to 1500m. Also, a limit of 10,000 venues was set to ensure the maximum number of venues was captured. For each venue, its Neighbourhood name, latitude and longitude and the Venue name, category, latitude and longitude were recorded. Some neighbourhoods did not return any results, but an if statement was used to check if the 'name' feature was present in the Foursquare results. In the event that it was not the Neighbourhood name, latitude and longitude were recorded a before but a value of 'none' was recorded for Venue name and category and 0 for Venue latitude and longitude.

Neighbourhood	Latitude	Longitude	Venue_Name	Venue_Cat	Venue_Lat	Venue_Lng
Anderston...	55.8644	-4.278	Gesso Tea and Coffee Lounge	Coffee Shop	55.866776	-4.271015
Anderston...	55.8644	-4.278	Marcellos Restaurant	Restaurant	55.865394	-4.268078
Anderston...	55.8644	-4.278	Loon Fung Cantonese Restaurant	Chinese Restaurant	55.865716	-4.268138
Anderston...	55.8644	-4.278	U Jarka Restaurant	Eastern European Restaurant	55.867373	-4.284590
Anderston...	55.8644	-4.278	Heera Indian Restaurant	Indian Restaurant	55.864776	-4.272287
...						

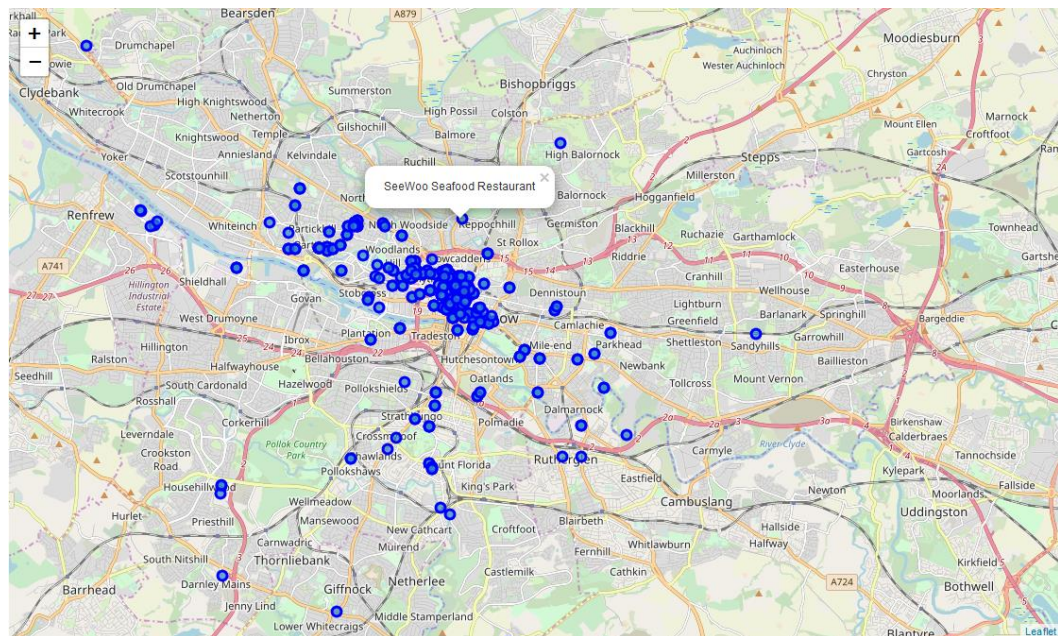
<sup>4</sup> <https://pypi.org/project/folium/>

<sup>5</sup> <https://foursquare.com/>

<sup>6</sup> <https://andrew.hedges.name/experiments/haversine/>



Again, Folium was used to plot the points on a map:



**Fig3.** Distribution of venues with Glasgow

### Distribution Analysis of Glasgow Venue

The first analysis done was a simple count of the number of venues within each neighbourhood, which was presented on a Choropleth map<sup>7</sup>. For the purposes of this, a geoJSON file was generated using data from [www.opendoorlogistics.com/data/](http://www.opendoorlogistics.com/data/).

DBSCAN<sup>8</sup> was then utilised to perform cluster analysis on the venue data, giving a more precise visual representation of the venue distribution.

Finally, the count and cluster analysis results were combined into a single map for interpretation and neighbourhood recommendation

## Results and Discussion

Overall, the Foursquare search found 409 venues in Glasgow. This, in itself seemed low considering Glasgow is the largest city in Scotland with a population in mid-2017 of 621,020<sup>9</sup>. For example, The Foursquare search found no venues near the Ibrox football stadium, however, a quick search on Google found at least 6, as can be seen below in Fig 4.

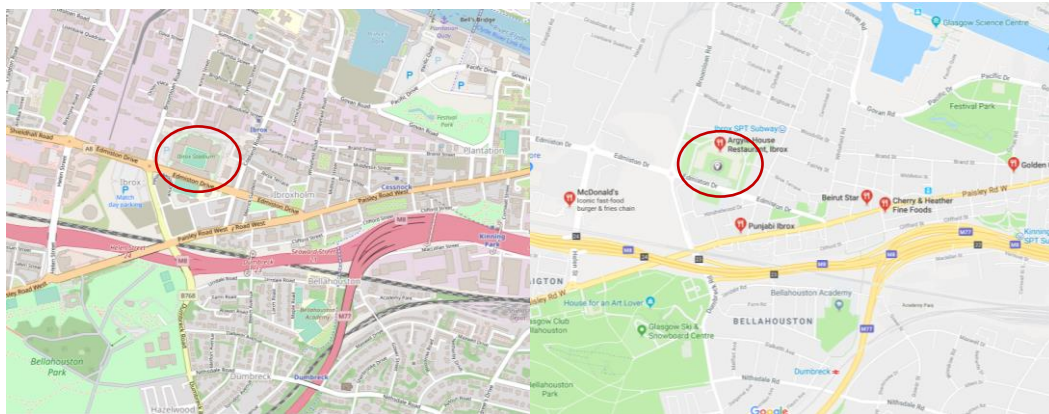
It was not possible to identify the source of this discrepancy during the course of this project, but would certainly need to be investigated further if this work was to be taken forward.

For the purposes of this project it was deemed the available data would be adequate to demonstrate the usefulness in this type of data analysis for the business problem at hand; identifying a location for a new venue within Glasgow.

<sup>7</sup> <https://plot.ly/python/choropleth-maps/>

<sup>8</sup> <https://scikit-learn.org/stable/modules/generated/sklearn.cluster.DBSCAN.html>

<sup>9</sup> <https://www.glasgow.gov.uk/index.aspx?articleid=22481>



**Fig 4.** Search result from Foursquare (left) and Google (right) for venues near Ibrox stadium

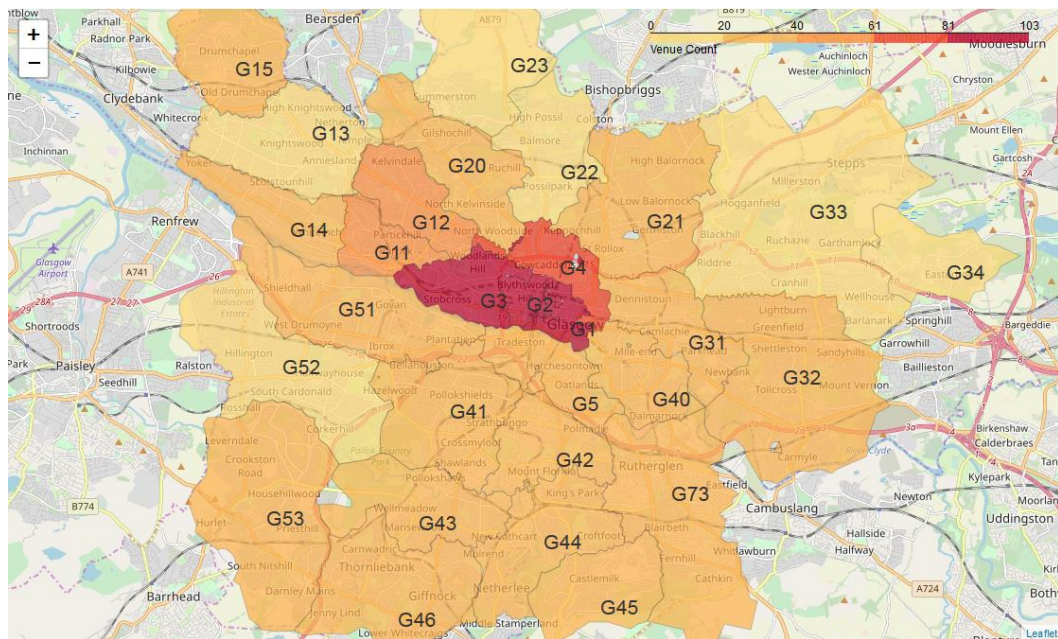
Counts of the number of Venues per Neighbourhood

Postcode	Neighbourhood	Latitude	Longitude	Counts
G2	Blythswood Hill, Anderston (part)	55.8636	-4.2612	82.0
G1	Merchant City	55.8586	-4.2456	73.0
G3	Anderston, Finnieston, Garnethill, Park, Woodl...	55.8644	-4.2780	68.0
G4	Calton (part), Cowcaddens (part), Drygate, Kel...	55.8710	-4.2494	56.0
G11	Broomhill, Partick, Partickhill	55.8742	-4.3166	24.0
G12	West End (part), Clevedon, Dowanhill, Hillhead...	55.8803	-4.3028	23.0
G5	Gorbals	55.8433	-4.2450	15.0
G41	Pollokshields, Shawlands	55.8417	-4.2891	10.0
G42	Battlefield, Govanhill, Mount Florida, Strathb...	55.8319	-4.2506	9.0
G40	Bridgeton, Calton, Dalmarnock	55.8443	-4.2156	9.0
G51	Govan, Ibrox, Drumoyne, Kinning Park	55.8626	-4.3295	7.0
G31	Dennistoun, Haghill, Parkhead (part)	55.8559	-4.2027	7.0
G14	Whiteinch, Scotstoun	55.8787	-4.3470	6.0
G44	Cathcart, Simshill, Croftfoot, King's Park, Mu...	55.8151	-4.2553	5.0
G53	Darnley, Pollok, Crookston, Parkhouse, Priesth...	55.8199	-4.3559	4.0
G73	Rutherglen	55.8250	-4.2089	4.0
G20	Maryhill, North Kelvinside, Ruchill	55.8919	-4.2898	1.0
G15	Drumchapel	55.9115	-4.3672	1.0
G43	Mansewood, Newlands, Pollokshaws	55.8191	-4.3003	1.0
G32	Carmyle, Tollcross, Mount Vernon, Lightburn, S...	55.8487	-4.1683	1.0
G45	Castlemilk	55.8018	-4.2347	1.0
G46	Arden, Carnwadric, Deaconsbank, Giffnock, Kenn...	55.7993	-4.3081	1.0
G21	Balornock, Barmulloch, Cowlares, Royston, Spri...	55.8805	-4.2177	1.0
G33	Cardowan, Carntyne, Craigend, Cranhill, Gartha...	55.8825	-4.1588	0.0
G34	Easterhouse, Easthall, Provanhall	55.8702	-4.1089	0.0
G23	Lambhill, Summerston	55.9122	-4.2672	0.0
G13	Anniesland, Knightswood, Yoker	55.8985	-4.3391	0.0
G52	Cardonald, Hillington, Penilee, Mossbank	55.8508	-4.3498	0.0
G22	Milton, Parkhouse, Possilpark	55.8905	-4.2489	0.0

Of the 409 venues found, 279 or 68%, are in the postcodes areas G1-G4, as can be seen in Fig 5. This, in itself, is not actually surprising as G1 and G2 comprise the main city centre and G3 and G4 several university campus areas.

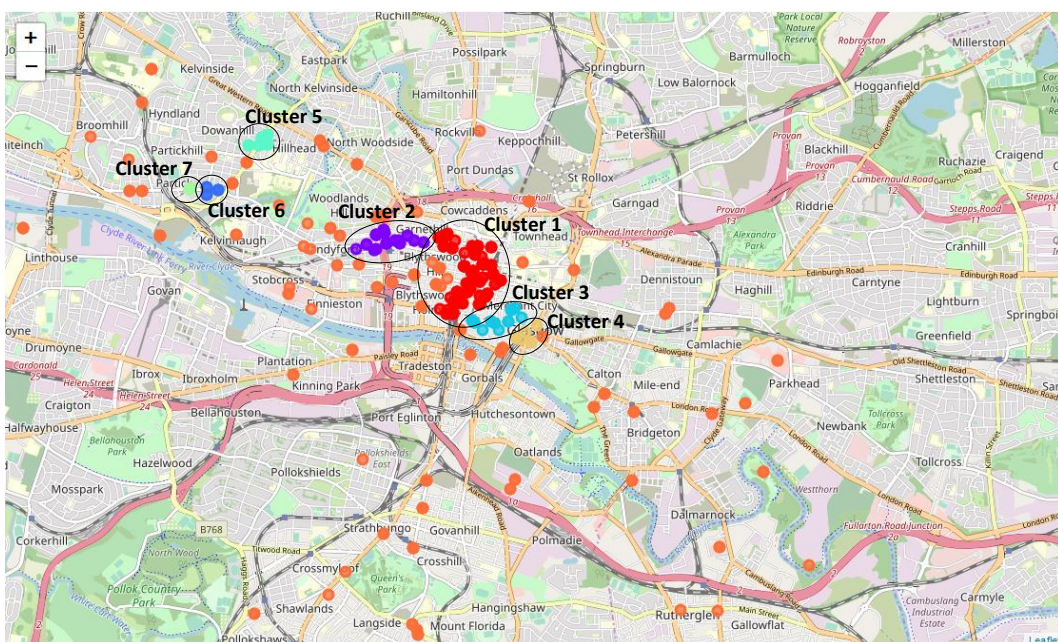


It should be noted that the scale in Fig 5. is slightly skewed, because '20' was added to the counts so they would be differentiated more clearly in the figure from those postcode areas that returned no venues at all (lightest orange in the figure).



**Fig5.** Choropleth map showing counts distribution of venues within Glasgow

The DBSCAN cluster analysis of the data identified 5 clusters of venues, Fig 6. Again, there is the expected clusters within the city centre, especially Cluster 1, which runs along major retail streets, theatres and also includes the two main train stations for Glasgow. Clusters 2, 3 and 4 also span high foot traffic retail streets with nearby theatres. Cluster 5 is just west of a large campus area, which may account for its presence and Clusters 6 and 7 are also on a retail street, though not as busy as those in the city centre.

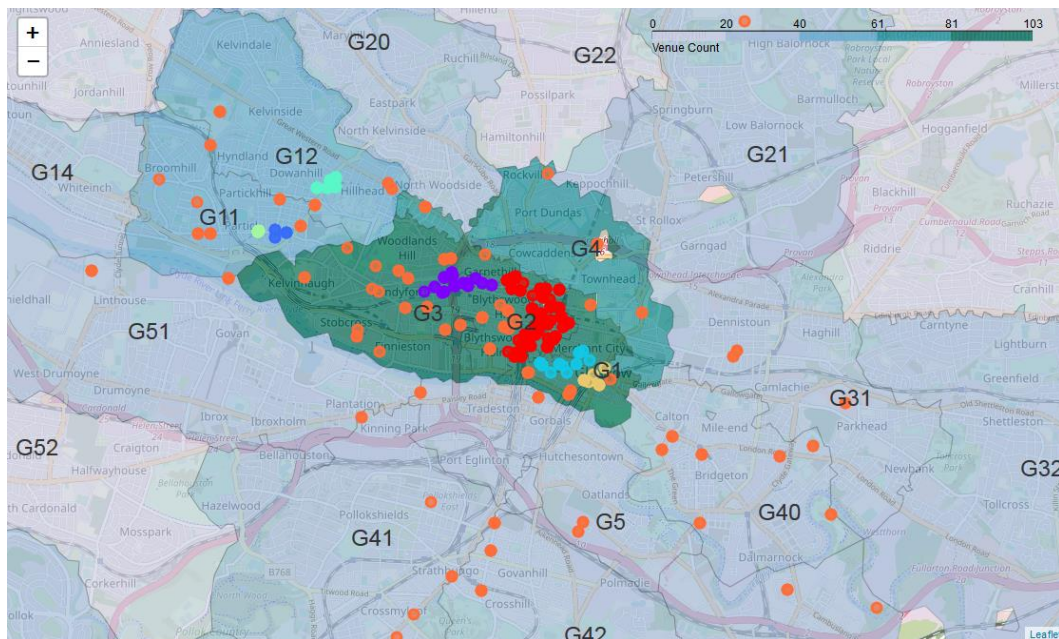


**Fig 6.** DBSCAN cluster analysis of venues within Glasgow. Showing 5 clusters, circled, and outliers in orange.

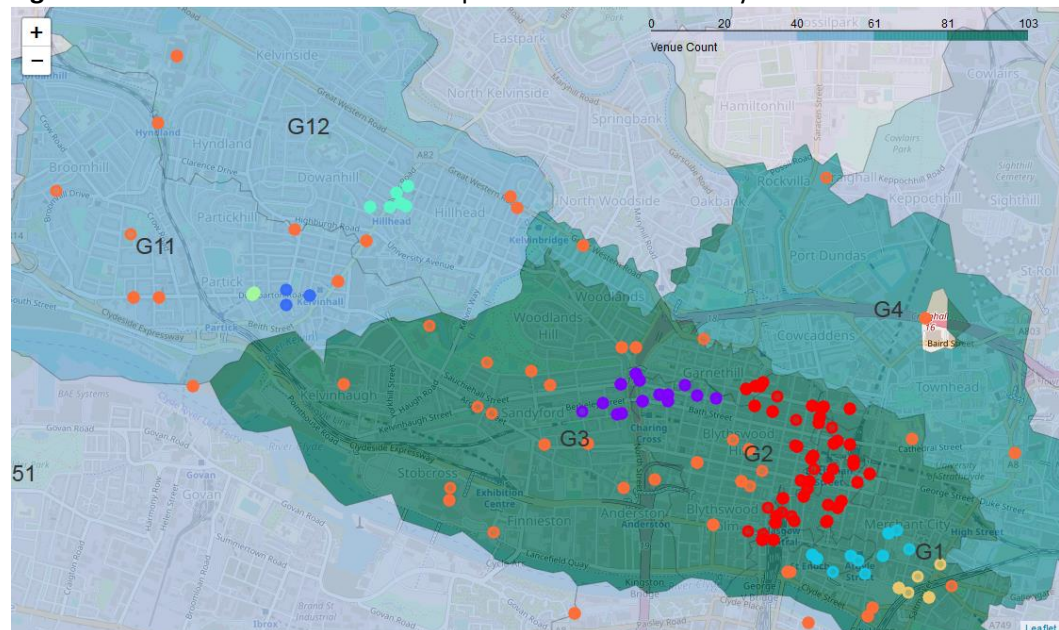


The cluster analysis brings up a very important point with regards to locating a new restaurant. All the clusters are in high foot traffic areas. Unless there is space for parking, which is limited in Glasgow, a restaurant needs to be sited where people are going to be anyway or more likely to pass; it needs to be visible. This is a factor sited in the links at the start of this report and would certainly need to be taken into consideration if this project were used to seriously consider siting a new restaurant.

Finally, the Choropleth and cluster analysis data were combined, Fig 8 and 9, to see the clustering in relation to the Neighbourhoods. As the Choropleth suggested, G1-G3 are pretty saturated with Venues though there are less data points in G4 than it's colour would suggest.



**Fig 7.** Combination of the count choropleth and DBSCAN analysis



**Fig 8.** City centre detail for the combination of the count choropleth and DBSCAN analysis

The initial premise to view a Neighbourhood as over saturated with 6 or more venues did underestimate the size of the Neighbourhoods. It may be beneficial to re-evaluate the clusters based on postcode sectors, which are subsections of each postcode district, 'G1 1' as opposed to just 'G1'. This would allow for further analysis of the G1-G3 neighbourhoods and may identify a more lucrative site for a new venue. For example, Cluster 1 spans the edge of neighbourhood G2, a deeper analysis may indicate a good spot within a sector of that neighbourhood.

Within the criteria of the project to find a 'goldielocks' Neighbourhood, it would likely be beneficial to site the new venue in G11 or G12. Other neighbourhoods are also viable as having at least one venue, however, the presence of clusters within these neighbourhoods indicates the viability that they would be able to sustain a new venture.

## Conclusion

The purpose of this project was to suggest a neighbourhood within Glasgow where a new restaurant owner could potentially site a new venue. A location that was neither untried nor oversaturated with other venues. To this end, Glasgow was divided into its district postcodes and cluster analysis was applied to lists of current venues obtained from Foursquare.

Through analysis of the results it was suggested that the most viable neighbourhoods for sustaining and new venture would likely be G11 or G12.

However, it was also noted that

1. There was some concern that Foursquare was not returning a full list of venues
2. Location is only one factor to consider when siting a new business
3. The district postcode areas are quite large and it may be beneficial to fine tune the analysis by reapplying it to the sector postcodes.

That being said, this is a good first step for any budding restaurant owner to take in setting up their new business.