



IDENTIFYING NEW BUSINESS OPPORTUNITIES IN GLASGOW

For IBM Data Science Professional Certification

OVERVIEW

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 entire project is divided into four stages.

- **Stage 1** : Post Codes are shortlisted based on whether they are within 5 km of the City Centre and whether they are within 2 km of the nearest Subway station
- **Stage 2** : The population statistics are extracted from 2 datasets for the shortlisted postal codes
- **Stage 3** : The venues at each of the shortlisted postal code are found
- **Stage 4** : An supervised learning algorithm is used to cluster the postal codes based on their similarities to each other.

DATA USED FOR THIS PROJECT

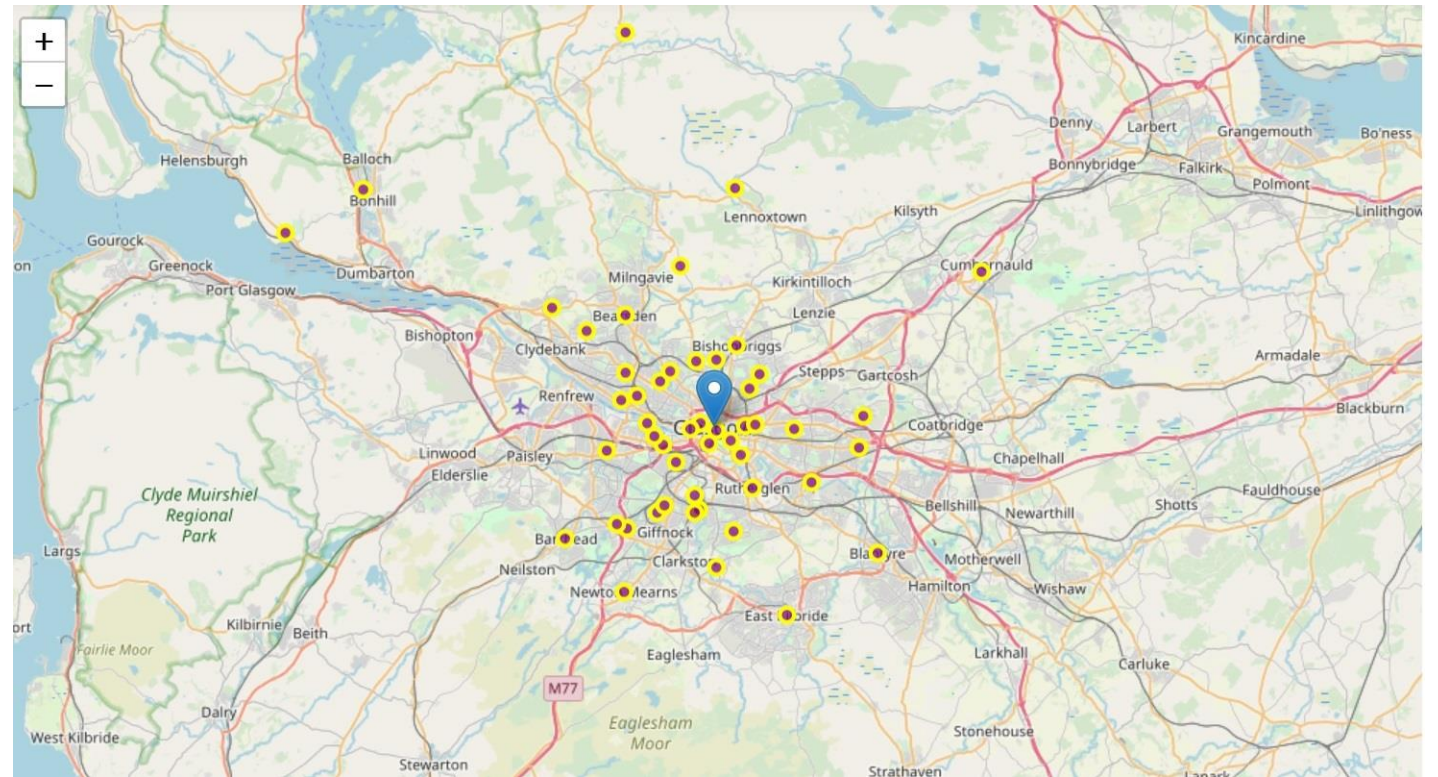
Five datasets are used for this project. The datasets come from three sources, namely, webpages, publicly available datasets and the FourSquareAPI.

- **Dataset 1:** All the postal codes in Glasgow with their latitude and longitude information. This dataset comes from webpage [1] and the FourSquareAPI.
- **Dataset 2:** The addresses of the Glasgow Subway stations and the latitude and longitude information. This data comes from a publicly available dataset [2] and the FourSquareAPI.
- **Dataset 3 and 4:** The Royal mail dataset [3] and the Glasgow City Council dataset [4] to extract total population and ethnicity Information. These are publicly available datasets
- **Dataset 5:** The business venues in each postal code This data set is acquired from the FourSquareAPI.

[1] https://en.wikipedia.org/wiki/G_postcode_area [2] <https://data.glasgow.gov.uk/dataset/glasgow-subway-station-locations/resource/8c057567-55e7-4631-ae77-e37b4a0562e8>
[3] <http://www.glasgow.gov.uk/CHttpHandler.ashx?id=16002&p=0> [4] <https://data.glasgow.gov.uk/dataset/national-identity-by-ethnic-group-2011-census>

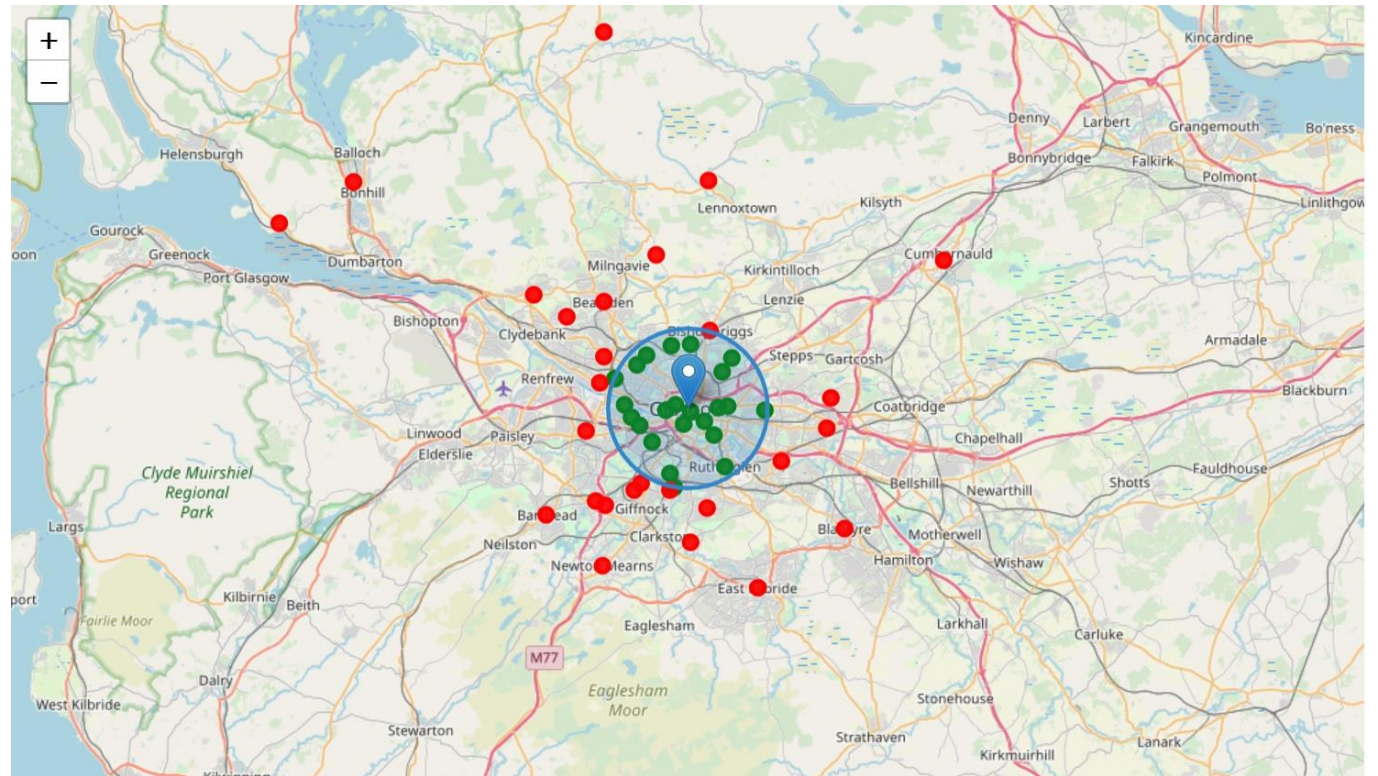
STAGE I – SHORTLISTING POSTCODES

- Glasgow has a total of **50** non-geographical postal codes that are spread over a very large area
- Postal codes closer to the city centre offer more business opportunities



STAGE I – SHORTLISTING POSTCODES

- Postal codes within 5 km of the city centre are considered suitable for new businesses.
- There are only 22 postal codes within 5 km of the city centre.



STAGE I – SHORTLISTING POSTCODES

- Glasgow has 15 subway stations
- In the 2018-19 period, 13.1 million subway journeys were recorded
- Postal codes within 2 km of the subway stations offer businesses a greater exposure to a wider client base

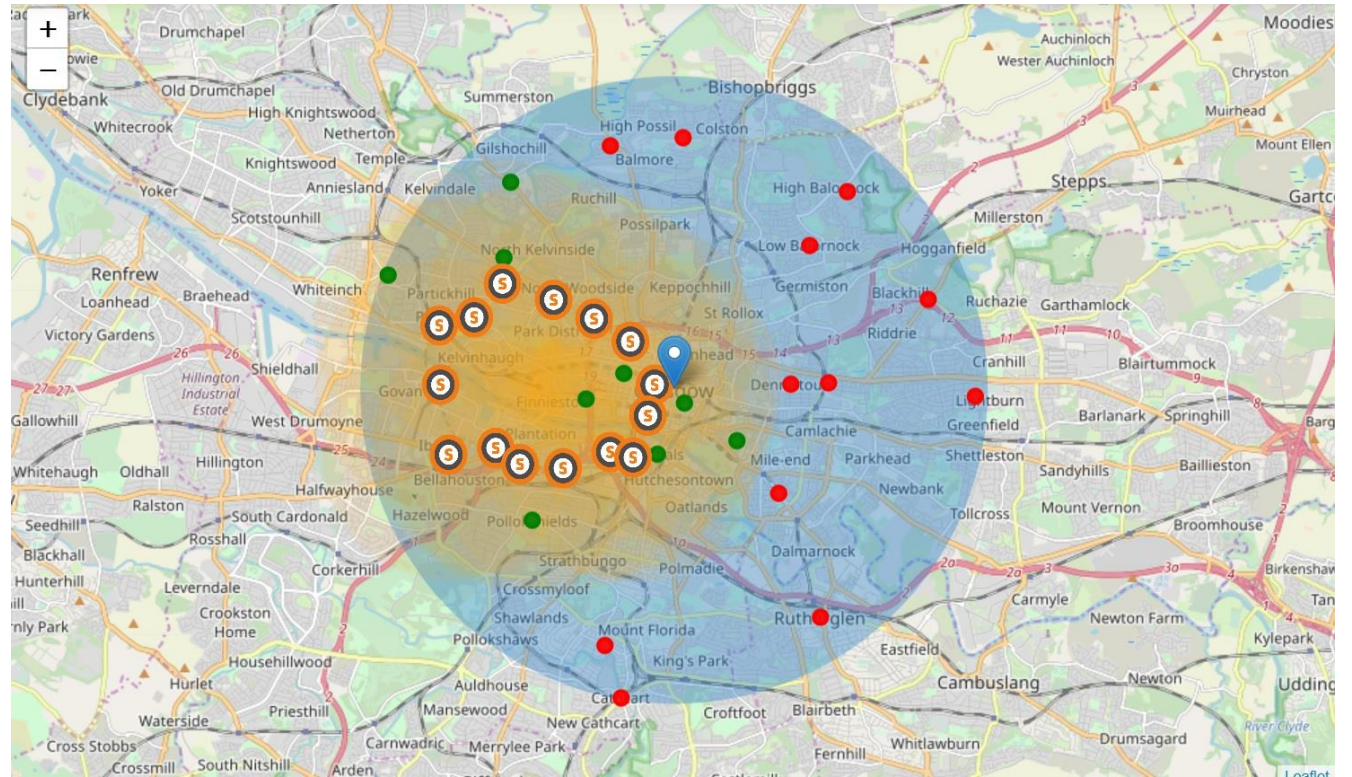


STAGE I – SHORTLISTING POSTCODES

- There are only ten postal codes that are both within 5 km of the city centre and within 2 km of a subway station

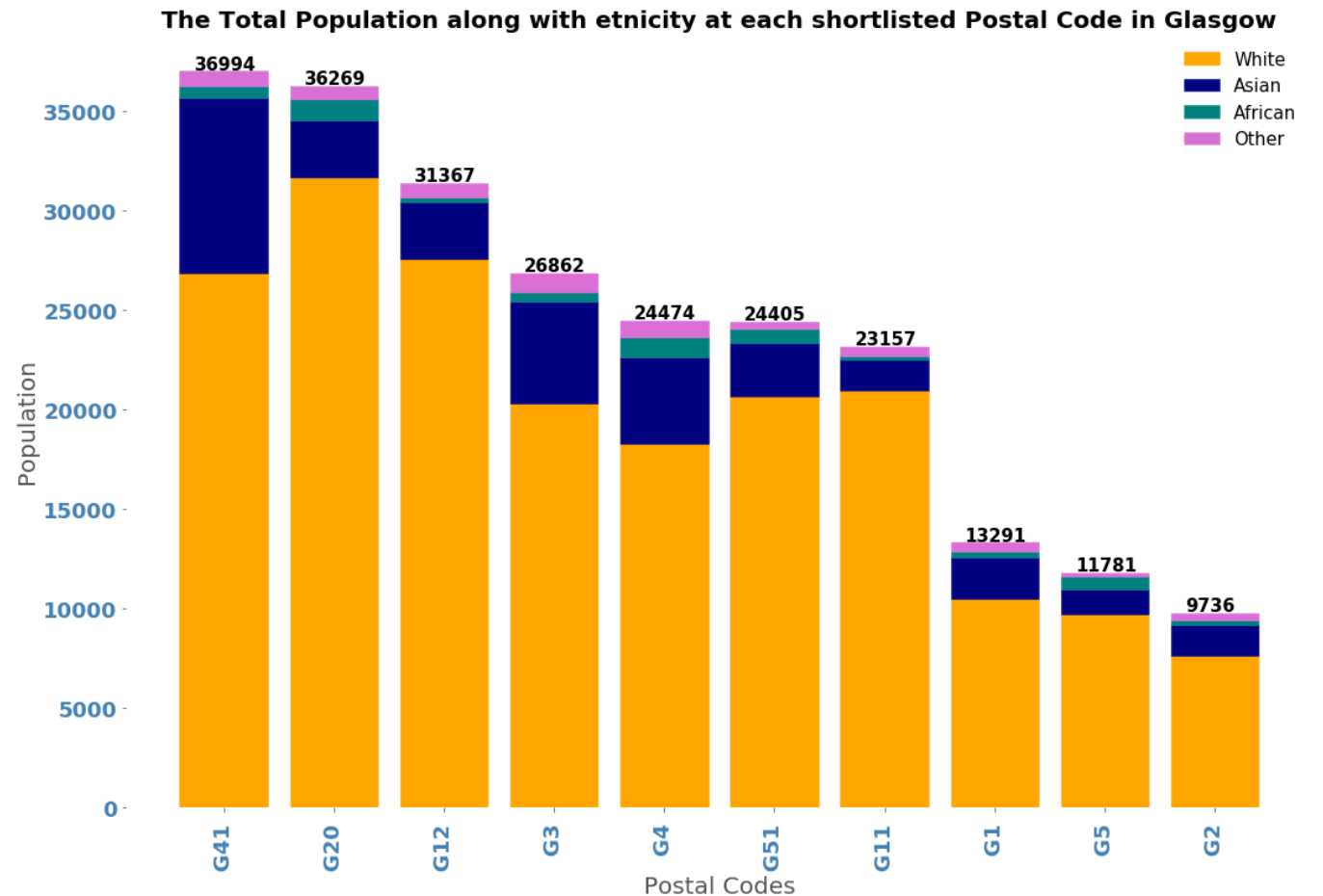
Shortlisted Post codes

G1 G2 G3 G4 G5 G11 G12 G20 G41 G51



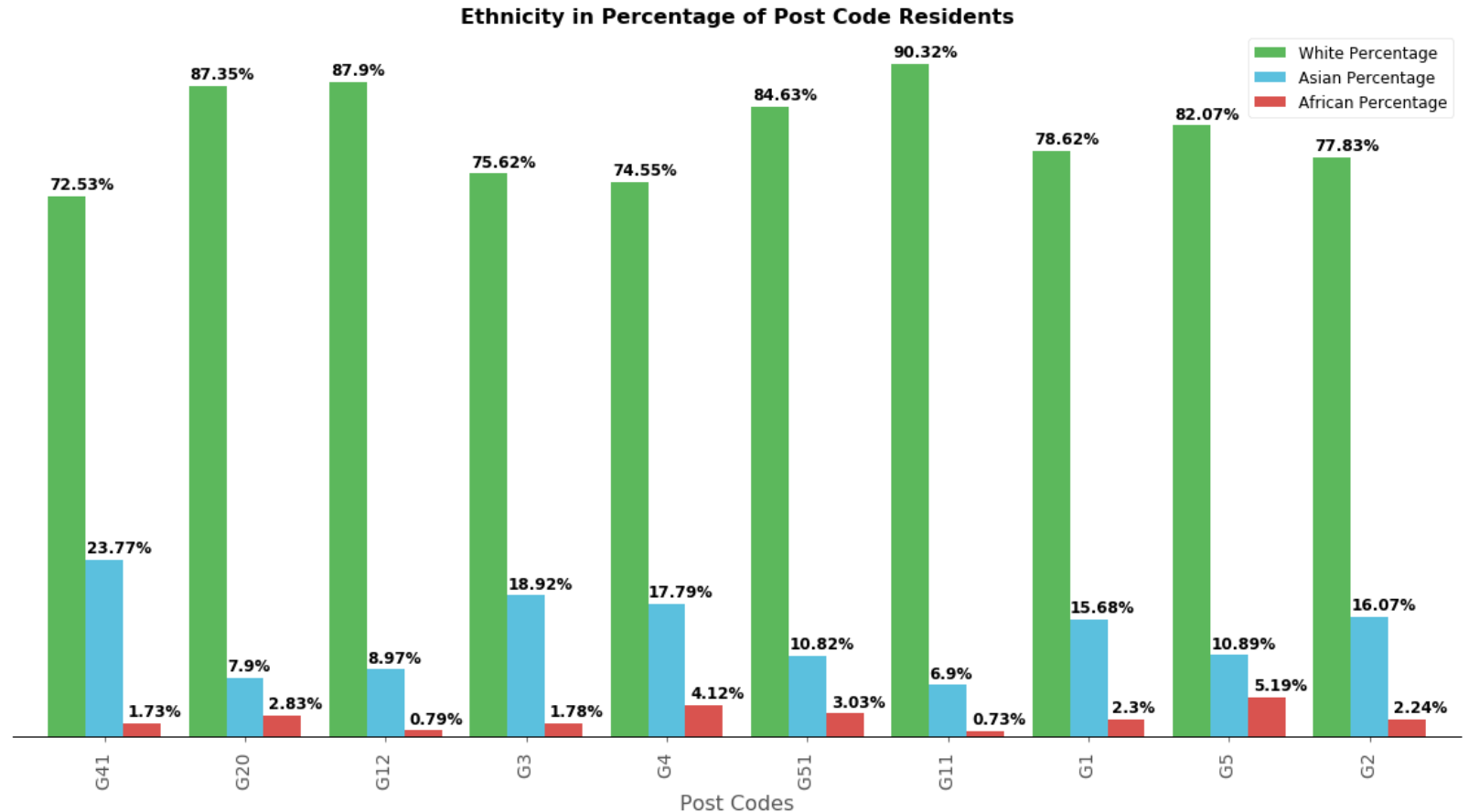
STAGE 2 – EXTRACTING POPULATION DATA

- A direct dataset that maps postal codes to population is not available. So two datasets are used for this purpose. Dataset 3 is used to map postal codes to data zones. Dataset 4 is used to map data zones to population data.
- G41, G20 and G12 have the highest populations.



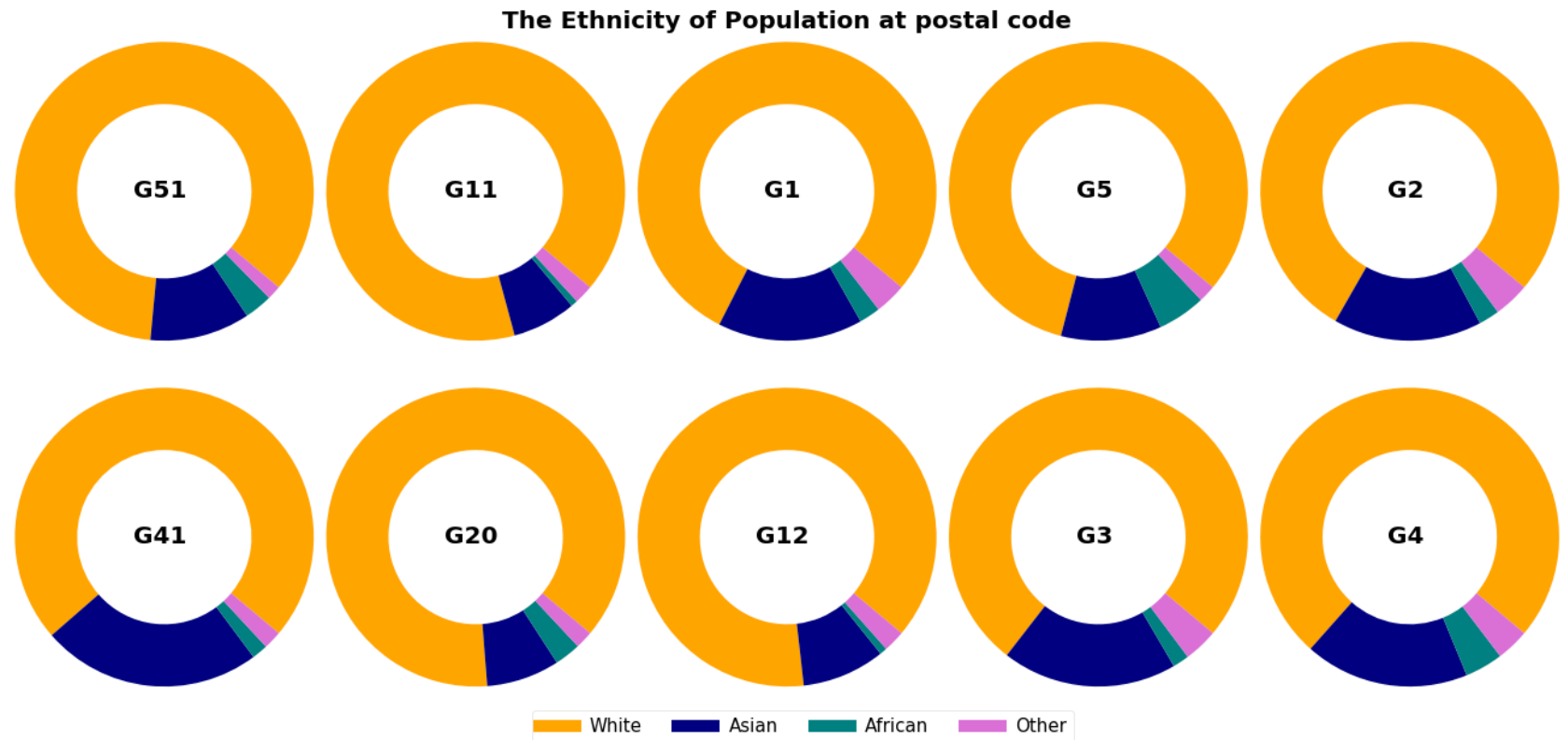
STAGE 2 – EXTRACTING POPULATION DATA

- G11 has the highest percentage of population with White Ethnicity
- G41 has the highest percentage of population with Asian Ethnicity
- G5 has the highest percentage of population with African Ethnicity



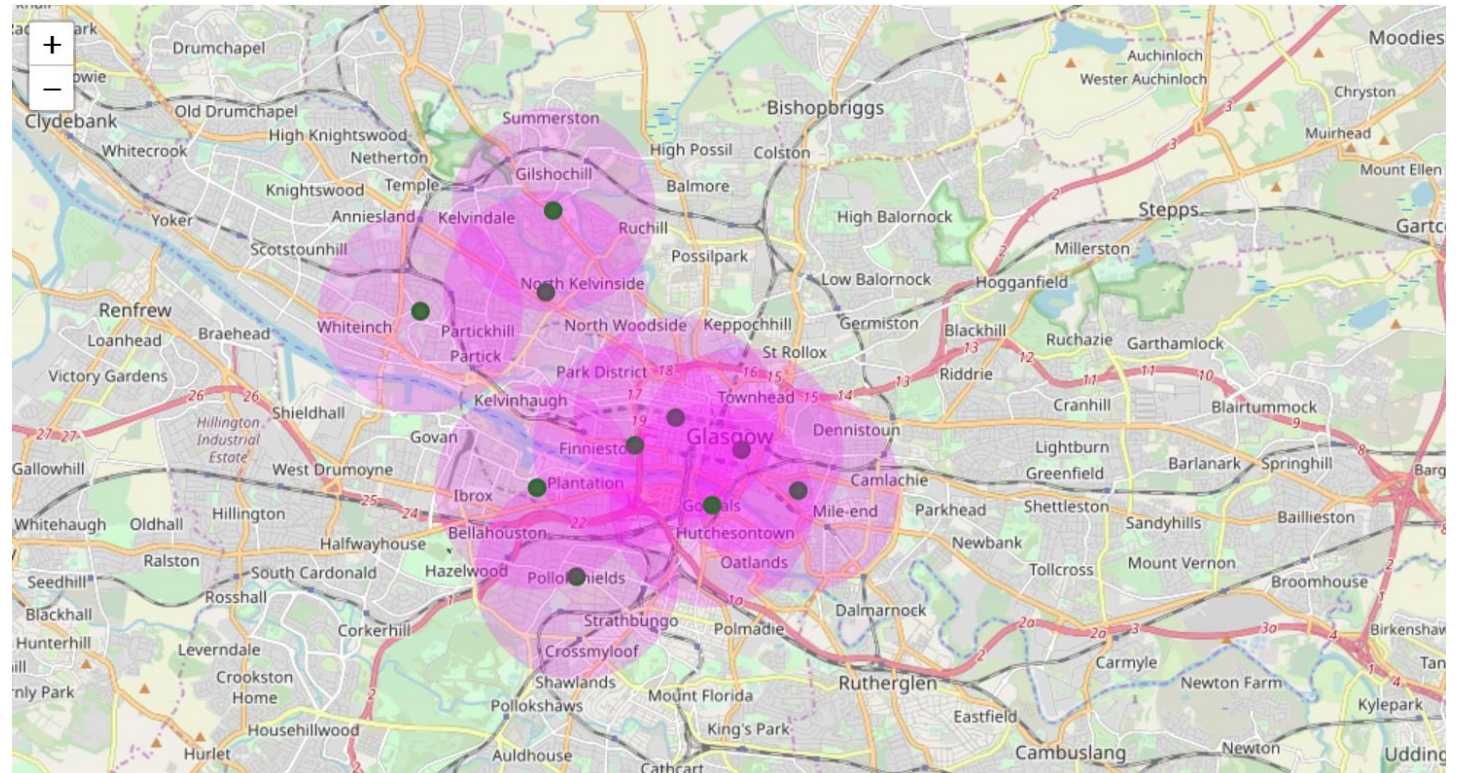
STAGE 2 – EXTRACTING POPULATION DATA

- G1 and G2 have similar population make-up
- G11 and G12 also have similar population make-up



STAGE 3- EXTRACTING VENUES

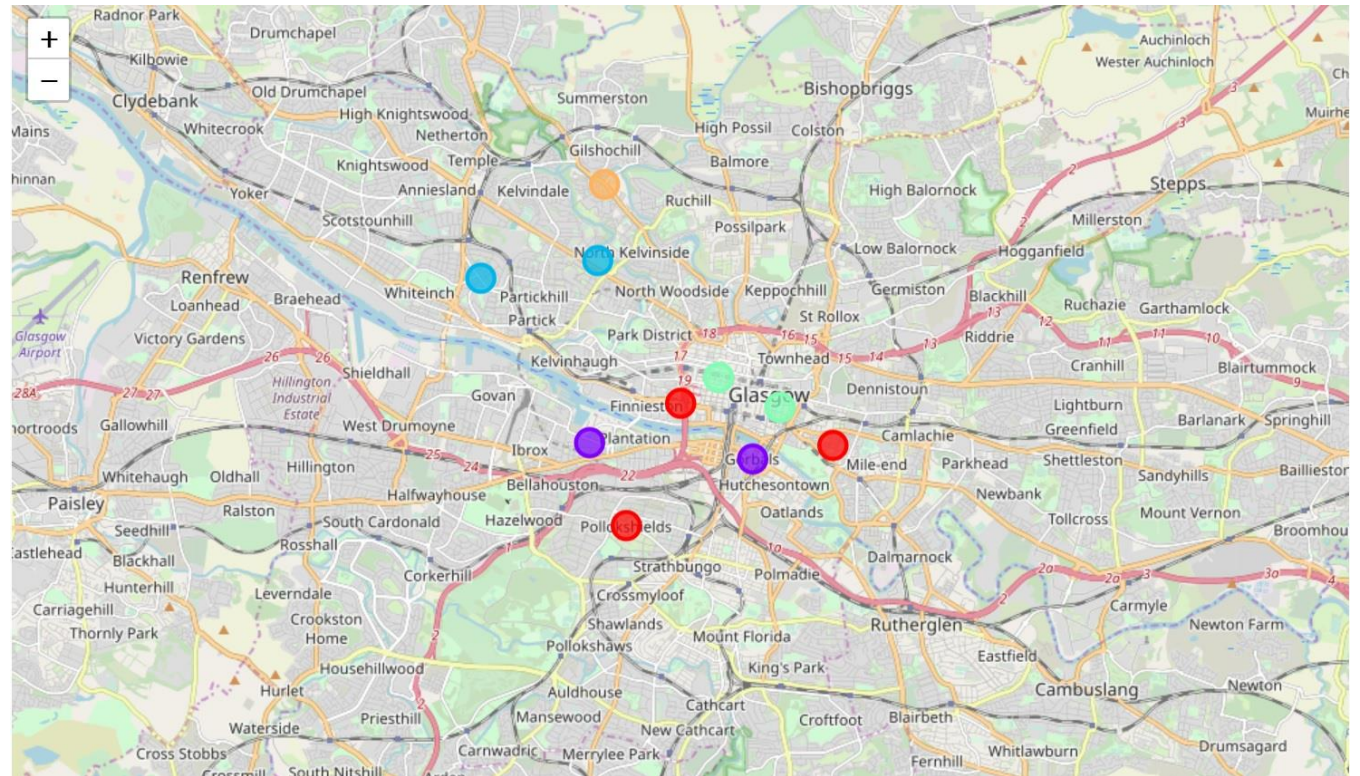
- Venues within 1.5 km of each postal codes are retrieved from the FourSquareAPI.
- The API returns 786 venues
- Venues belong to neighbouring postal codes and repeating venues are removed
- There are 420 unique business venues within the 10 postal codes



STAGE 4- CLUSTERING OF POSTAL CODES

- The venues data and the population data are fed into K-means algorithm.
- The 10 postal codes are clustered into 5 categories.

Postal Code	G3	G4	G41	G5	G51	G11	G12	G1	G2	G20
Cluster Number	1	1	1	2	2	3	3	4	4	5
Colour on map	Red	Red	Red	Purple	Purple	Blue	Blue	Green	Green	Orange



RESULTS OF CLUSTERING AND POSSIBLE BUSINESS OPPORTUNITIES SUMMARY

K-means Cluster label	0	1	2	3	4
Postal codes in the cluster	G3, G4 and G41	G5 and G51	G11 and G12	G1 and G2	G20
Description of the clusters	Cluster members have the highest population diversity and food related venues are the commonest in this cluster.	This cluster has a lot of hospitality and entertainment venues in the postcode	This cluster has medium population and low diversity. Food related venues are common in this cluster	These post codes have low population and medium diversity. Food related venues are common in this cluster. These postcodes also have the lowest distance from the city centre.	This cluster is furthest away from the city centre and has a high population with no entertainment venues
Possible Business Opportunities in this Cluster	These Postal codes have among the highest Asian population density in Glasgow. Non-food Asian themed businesses are likely to face little competition in these three postal codes. So, G3, G4 and G41 are recommended as venues for new Asian businesses.	In this cluster, there are a lot of hotels, bowling alleys, music venues and not a lot of food venues like restaurants, bars/pubs etc. So, new restaurants, bars and pubs will face little competition. So, they are identified as possible business opportunities in this cluster.	The Postal codes in this cluster are close to one of the largest universities in Glasgow. So, this cluster offers opportunities for student service businesses like books, stationary, printing shops etc. These businesses have the potential to succeed in this area and have little competition at these postcodes.	This cluster has postal codes that are near the city centre. So, at these postal codes there are opportunities for businesses that want to provide services to tourists like tour operators, souvenir shops, hotels etc.	This post code lacks entertainment venues. So, in G20 there are opportunities for entertainment & leisure businesses like movie cinemas, music clubs etc that want to target a new sector of the Glasgow population.