IBM DATA SCIENCE PROFESSIONAL CERTIFICATE

FEBRUARY 2021

Spilling the Beans on the Coffee Game: The Battle of Neighbourhoods

---

Akhilesh MISTRY

Contents

1 Background

In a world where 80% of the UK population visits a coffee shop at least once a week, 16% visit daily and we consume 95 million cups of coffee a day [1], it suffices to say that the coffee industry isn't going anywhere soon. As Britain's first 'European Green Capital' [2], Bristol would welcome a coffee shop using sustainably sourced products with a future-friendly business plan.

With a large student and millennial population [3] [4] , the demand for that daily cup of joe is only rising, along with its prevalence on social media. Be it for the '#latteart' post on Instagram or simply to get their day started, as The Bristol Mag so simply put it "we're bound to contract a slight case of FOMO" [5].

So, you want to open up a coffee shop in Bristol, where do you start? The answer is with location. It has been shown that the number one contributing factor to the success of a coffee shop is its location [6] and so, here lies our problem. Based on the location of existing coffee shops, population density of area in Bristol, proximity to the city centre, train stations and common cycle routes; where is the optimal location to open a coffee shop in Bristol?

Budding entrepreneurs keen on opening successful yet sustainable businesses or even large franchises would be interested in the benefits of this study by way of boosting sales, and so profits, as a result of carefully considering the location of their new coffee shop.

2 Data

2.1 Using data to solve the problem

In order to determine the optimal location, we will explore the relationships between the location data of existing coffee shops in Bristol and the following:

- Bristol population data
- Train station locations
- Proximity to the city centre

2.2 Data Sources

Data on existing coffee shops was obtained using the Foursquare API [7]. Population data was obtained through Nomis, a tool offered by the Office for National Statistics [8]. Location data for the city centre was obtained using geopy's geocode and postal code area centres were obtained using pgeocode. Train station data was web scraped from Wikipedia [9] [10] and location data obtained using geocode also. Finally, the postal code geometry data for map overlays was obtained via GitHub [11].

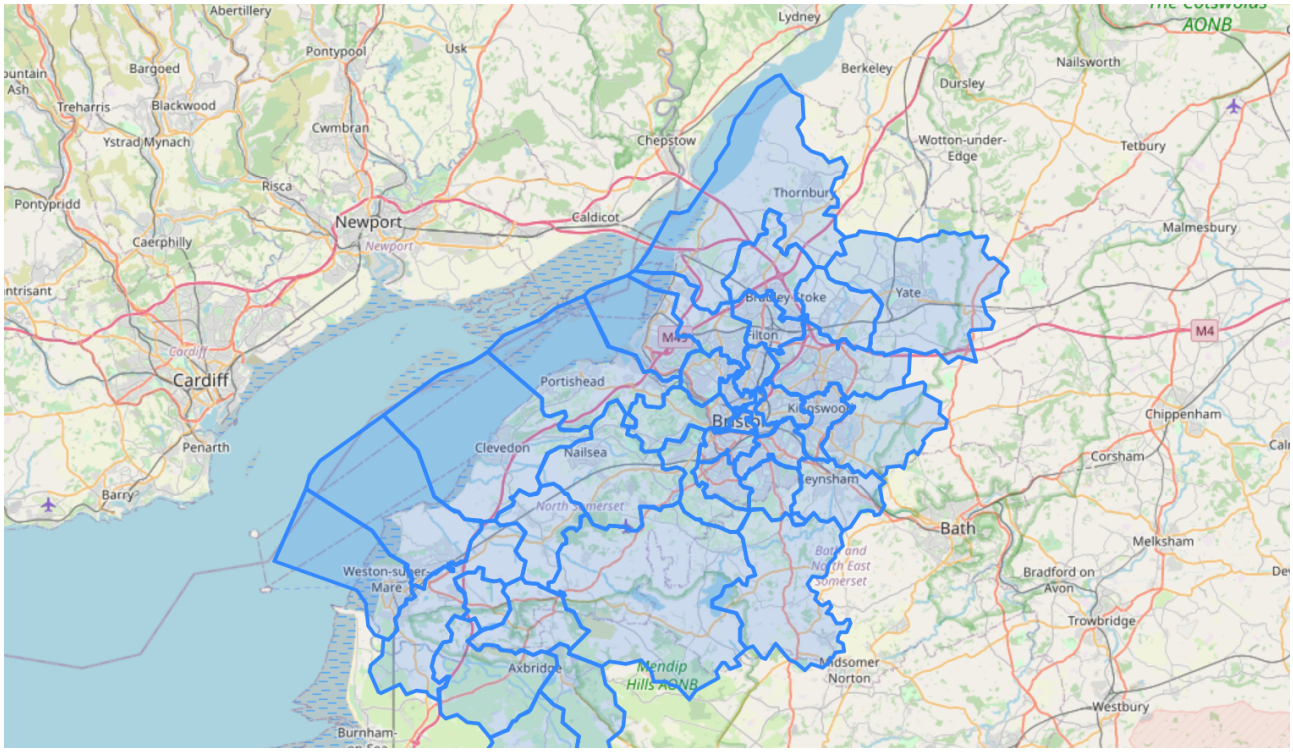

Figure 1: A map of the postal code zones in Bristol from BS1 - BS49, omitting discontinued zones.

2.3 Data Cleaning

As is common with data science projects, the data cleaning process was the most time-consuming phase of the project. For ease of understanding, the data cleaning process will be split into sections with each section assigned to a particular data set.

2.3.1 Foursquare Coffee Shop Data

The Foursquare search API gives access to vast amounts of data, most of which is not necessary for the data analysis process. Therefore firstly, of the data obtained, a filter was applied to remove the prefix 'location' from column headings to make the data frame more readable. Subsequently, a number of columns were dropped as they contained either redundant or irrelevant data. Following this, all duplicate rows were removed and any missing 'post code' data from rows was replaced with the mode 'BS1'. The post code data, as is the usual format in the UK, contained a combination of 6 or 7 alphanumeric characters. Only the first 3 or 4 characters are required, therefore the data frame was formatted to remove the trailing characters.

	Name	Categories	Latitude	Longitude	PostCode
27	Small St Coffee	Café	51.454937	-2.594604	BS8
35	Kiosk Coffee	Coffee Shop	51.452770	-2.592462	BS1
42	Brew Coffee Co	Coffee Shop	51.461548	-2.608434	BS1
46	Cotham Coffee House	Café	51.462591	-2.597850	BS1
1	Coffee #1	Coffee Shop	51.485549	-2.611794	BS1

Figure 2: This table shows the format of the data on the location of each coffee shop in Bristol, obtained from Foursquare.

From this data frame, the relevant maps could be plotted.

2.3.2 Population Data

The population data obtained from ONS had data from the 2011 UK census pertaining to 'male', 'female' and 'occupied household' data for postcodes beginning with letters A-F. Firstly, the total of the male and female population was taken along with the post codes, and the remaining columns were dropped. Then, as only the first 3/4 characters of the post codes were required, the trailing characters were removed in order to match the Foursquare coffee shop data. The coordinates of each postal code centre were also obtained and added to the data frame. There have been a few postcodes which have been discontinued over the years and as a result have no population data, these rows were dropped.

	Postcode	Population Count	Latitude	Longitude
0	BS1	238211	51.4552	-2.59660
1	BS2	184134	51.4552	-2.59660
2	BS3	231989	51.4552	-2.59660
3	BS4	101631	51.4335	-2.55525
4	BS5	52461	51.4620	-2.55190

Figure 3: This table shows the format of the data on the density of the population in Bristol, obtained from the Office for National Statistics.

The resulting data frame contained all the necessary data to plot informative maps and carry out the subsequent analysis.

2.3.3 Train Station Data

The clean-up for the train station data was fairly simple. After web-scraping the html data from Wikipedia, only underscores and a '/wiki/' prefix had to be removed from each station name before being added to a data frame. Simultaneously, the coordinates of each train station were obtained and added to the date frame.

	Station Name	Latitude	Longitude
0	Avonmouth railway station	51.499722	-2.698862
1	Bedminster railway station	51.440789	-2.593497
2	Bristol Temple Meads railway station	51.449099	-2.580403
3	Clifton Down railway station	51.464453	-2.611157
4	Lawrence Hill railway station	51.458580	-2.564170

Figure 4: This table shows the format of the data on the location of train stations in Bristol, obtained from Wikipedia and using Nominatim.

2.4 Exploratory Data Analysis

The following plots and tables illustrate the coffee shop distribution, population density and various location data points around Bristol.

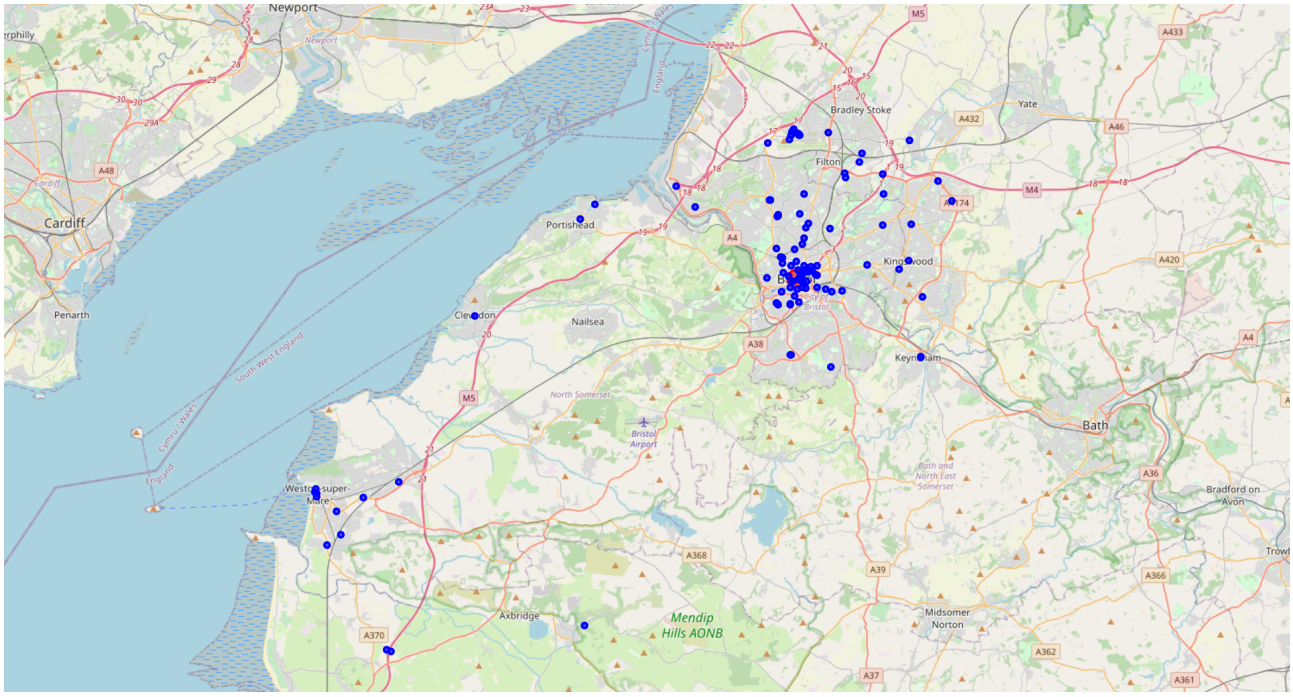

Figure 5: A map of Bristol showing the distribution of shops that primarily serve coffee.

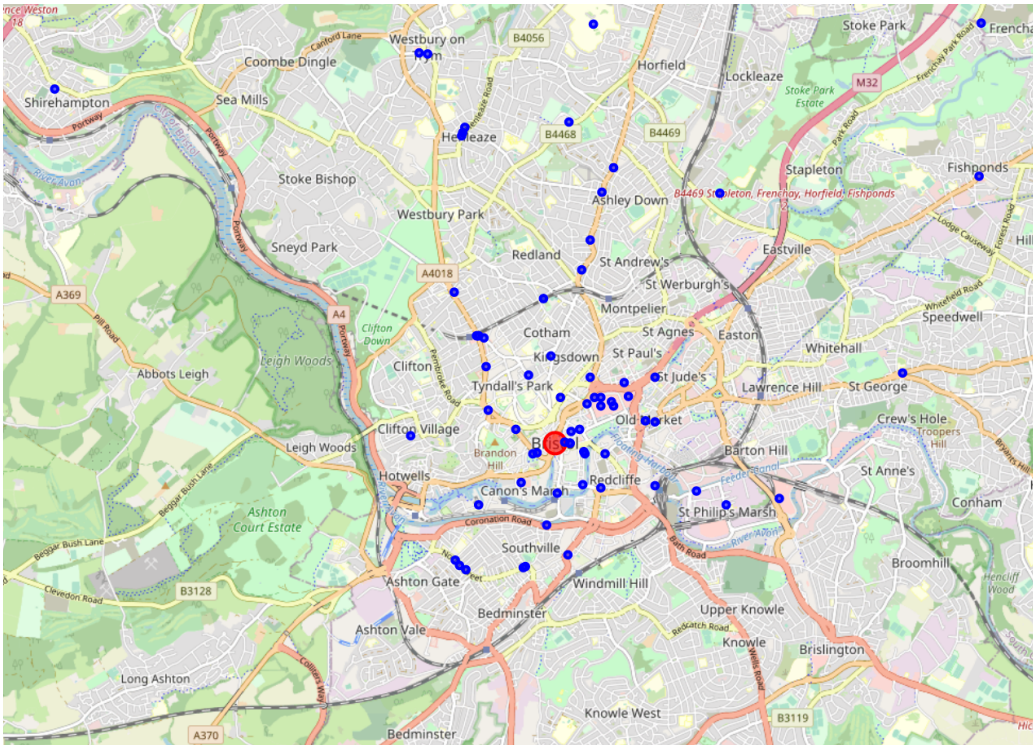

Figure 6: A map of Bristol showing the distribution of shops that primarily serve coffee with a closer zoom.

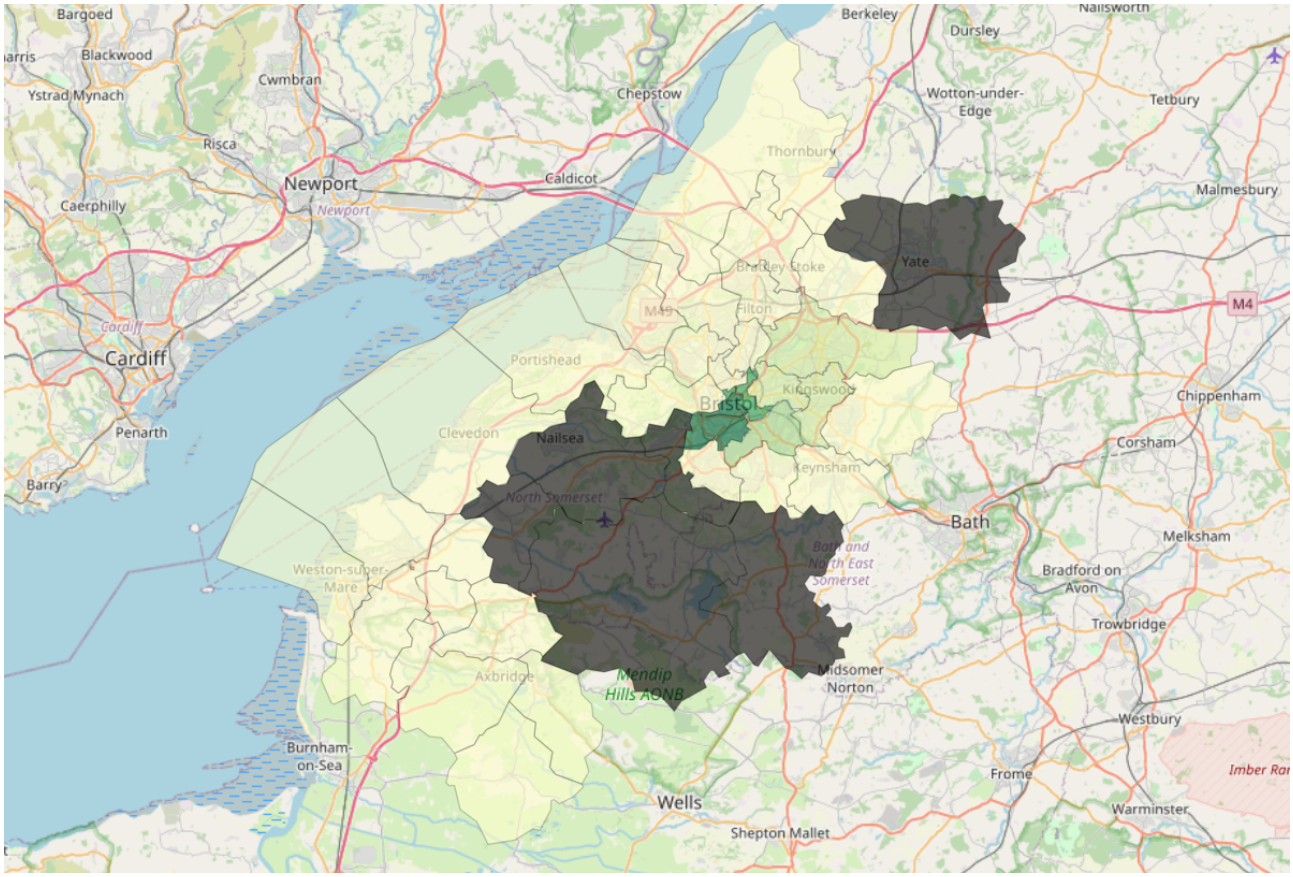

Figure 7: A choropleth map of Bristol showing the population density.

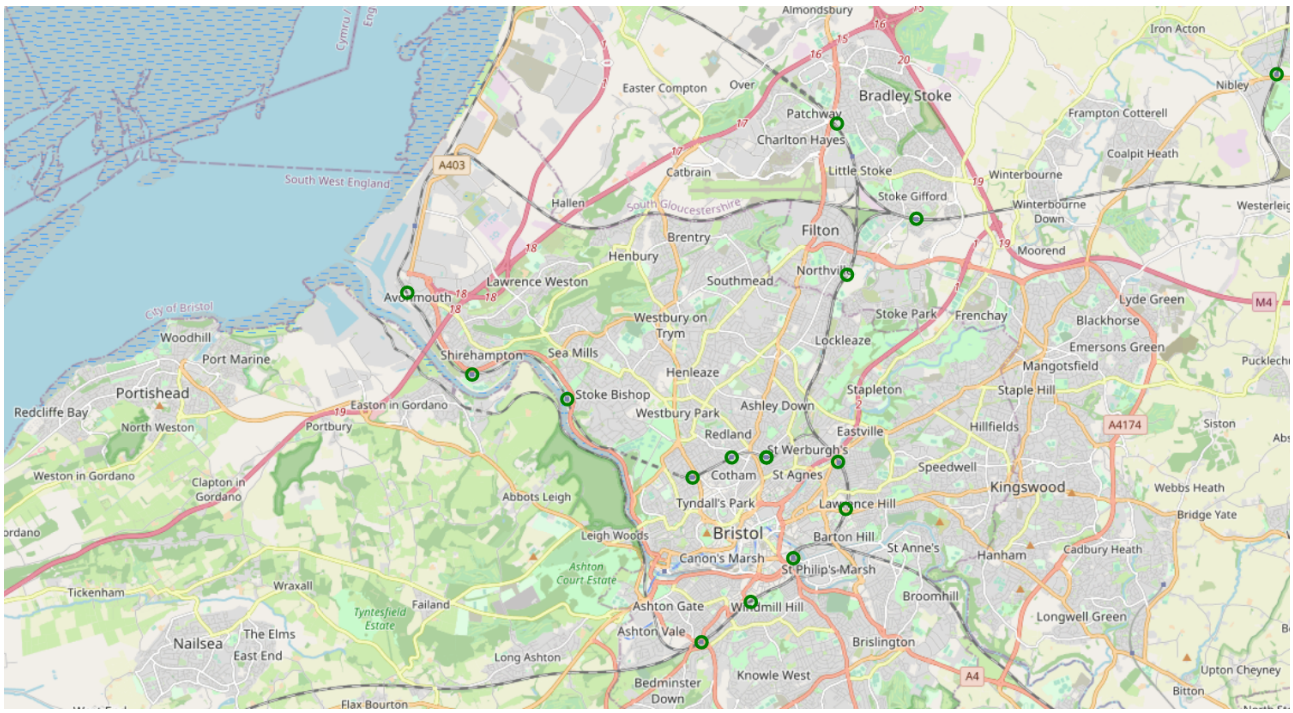

Figure 8: A map showing the location of train stations situated about Bristol.

3 Methodology

In order to segment the coffee shops in Bristol into groups with similar characteristics, k-means clustering was employed. This unsupervised method uses the distance of datapoints from self-generated centroids (based on the data) to cluster datapoints. This is extremely useful to us, as it can identify the cluster of coffee shops that would provide the best location for a new coffee shop.

The k-means clustering required the coffee shop location and venue category data as well as the population data of each postcode.

4 Results

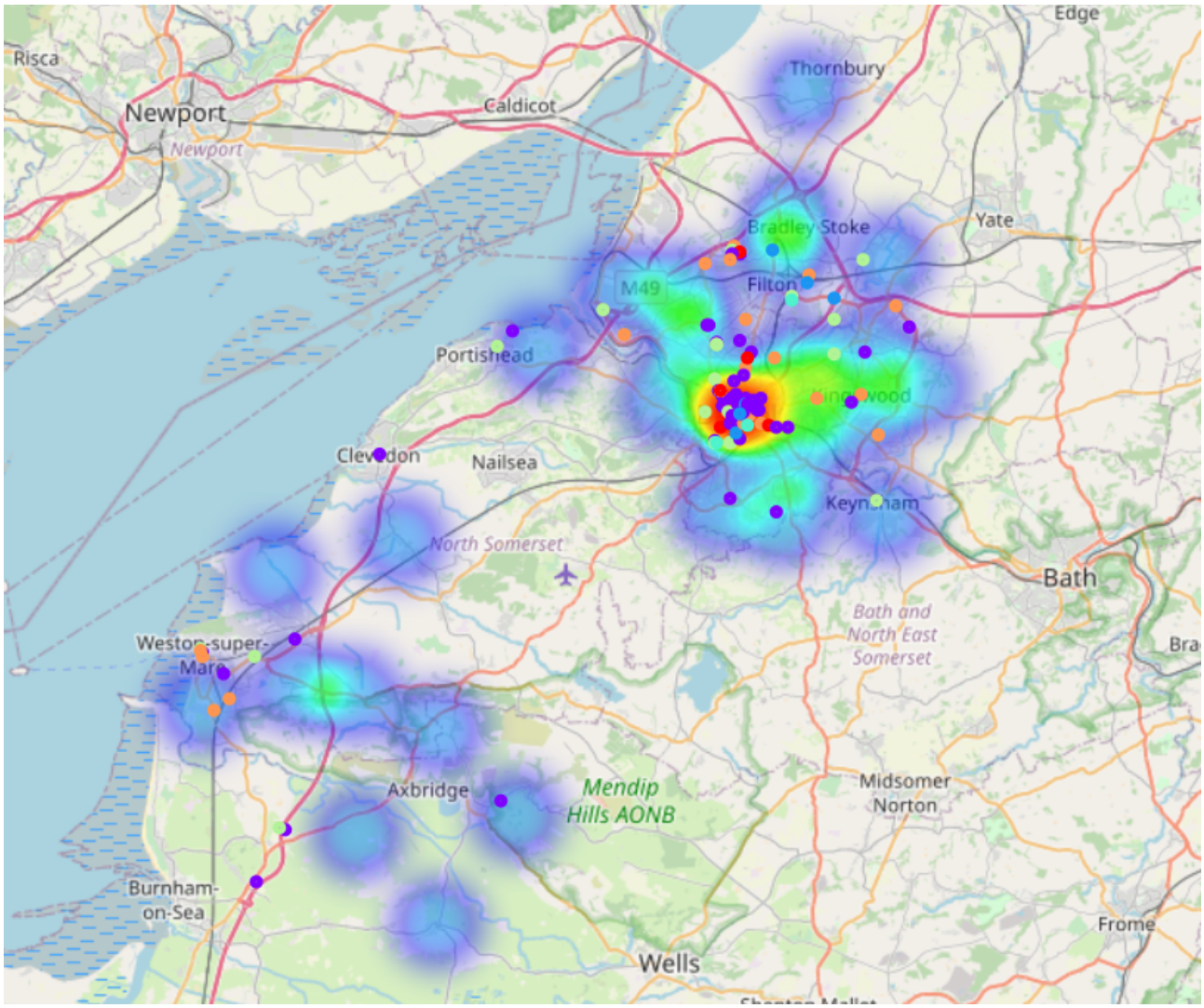

Figure 9: This map shows the k-means clusters on top of the population density of each postcode, indicating a possible correlation.

5 Discussion

The results display a promising correlation between population density and coffee shop location. However, I believe that by including venue category within the clustering analysis the correlation between location and population density has been negatively impacted. Despite this, a relationship between location and proximity to the city centre can also be inferred.

6 Conclusion

In this report, the endeavour of finding the optimal location for a new coffee shop in Bristol was explored. Primarily using the data on exiting coffee shops and population density statistics, a k-means clustering analysis was conducted, resulting in a promising correlation in the data being identified, however, with room for improvement. As far as recommending a location, this study can be improved and refined though ultimately many other factors weigh in on the business and logistical decisions required when starting a new business.

7 Future Work

A few simple measures that can be taken to improve the analysis would be to find the optimal number of clusters using a loop and trials from 1 to 10 for example. Also, omitting venue category from the clustering would produce less nuanced though possibly more useful information to put toward making a decision as to where a new coffee shop should be located.

A References

[1] Coffee Facts | British Coffee Association, date retrieved: 25/02/21
https://www.britishcoffeeassociation.org/coffee-in-the-uk/coffee-facts

[2] Sustainability | The city of Bristol | University of Bristol, date retrieved: 25/02/21
http://www.bristol.ac.uk/city/bristol-is-famous-for/sustainability/

[3] The Population of Bristol - November 2019, date retrieved: date retrieved: 25/02/21
https://democracy.bristol.gov.uk/documents/s45306/App%20A%20Population%20Bristol%20Nov%2019%20Extract.pdf

[4] The Bristol neighbourhood that's one of the most 'millennial' in the country - Bristol Live: date retrieved: 25/02/21
https://www.bristolpost.co.uk/news/bristol-news/bristol-neighbourhood-thats-one-most-2806440

[5] Bristol's Caffiene Scene - The Bristol Magazine Online, date retrieved: 25/02/21
https://thebristolmag.co.uk/the-caffeine-scene/

[6] Trends of the UK Coffee Shop Industry | Esquires Coffee UK, date retrieved: 25/02/21
https://esquirescoffee.co.uk/news/trends-uk-coffee-shop-industry/

[7] Foursquare Developer Console, date retrieved: 25/02/21
https://foursquare.com/developers/apps

[8] Postcode Headcounts and Household Estimates - 2011 Census - Nomis - Official Labour Market Statistics, date retrieved: 25/02/21
http://www.nomisweb.co.uk/census/2011/postcode$_h$eadcounts$_a$nd$_h$ousehold$_e$stimates

[9] Category:Railway stations in Bristol - Wikipedia, date retrieved: 25/02/21
https://en.wikipedia.org/wiki/Category:Railway$_s$tations$_i$n$_B$ristol

[10] Category:Railway stations in Bristol, Bath and South Gloucestershire - Wikipedia, date retrieved: 25/02/21
https://en.wikipedia.org/wiki/Category:Railway$_s$tations$_i$n$_B$ristol,$_B$ath$_a$nd$_S$outh$_G$loucestershire

[11] uk-postcode-polygons/BS.geojson - at master, date retrieved: 25/02/21
https://raw.githubusercontent.com/missinglink/uk-postcode-polygons/master/geojson/BS.geojson