

The Battle of Neighbourhoods: Bristol, UK

(Applied Data Science Capstone by IBM/Coursera)

Jens Sauther

March 4th, 2021

1. Introduction

1.1. Background

Deyan Sudjic's deeply original book "The Language of Cities" begins with the statement, that *City* is a word used to describe almost anything, from tiny settlements in the Midwestern United States to urban mega-structures like Tokyo with populations of tens of millions people. Eventually, according to Sudjic, "a city is made by its people, within the bounds of the possibilities that it can offer them".

Individual humans are more mobile than ever, millions of us moving homes, cities, and countries every year. We find ourselves in strange new environments, often not knowing much about our new surroundings, and exploring them is naturally part of the enjoyment. However, the initial question is obviously: where to start?

1.2. Business Problem

Different areas in the same city are usually quite different from each other, and - especially as newcomers - we are usually not aware of the possibilities each neighbourhood offers. That is true no matter if it is in terms of housing, transportation, entertainment & culture, tourism, quality of life, opportunities to open a new business, etc. Especially when looking for a place to live in a new city, knowledge about the different parts of town can be quite helpful to decide in which neighbourhood to live in for the next few years.

A careful analysis of the single areas of your new home can offer deep insights about their differences and commonalities, what is available here, and what is lacking there. While this might of course also - for example - help business owners considering opening a new restaurant and looking for a suitable location to ensure long-term success, the main focus here will be on individuals and families looking for suitable neighbourhoods (e.g. offering lots of entertainment possibilities or being family-friendly, respectively).

The city to be analysed will be Bristol in the Southwest of the United Kingdom. The intention of the project is to use openly accessible data to determine the different types of wards (the local term for neighbourhoods) and their possibilities, resulting in a list of the most suitable areas.

NOTE: This of course is rather subjective and a matter of taste, since some people might prefer different venues over others, so we will stick with or client's personal preference in that case.

Here, the focus is clearly on individuals and families moving to Bristol that would like to know which ward would suit them most and thus look for a new home there, the results could also provide information that will be of interest to the following groups:

- Property management companies that would like to estimate the influence of the location of the property as additional factor in the final price (a "good neighborhood" allows for a higher overall price).
- Business owners that would like to know about up-and-coming, popular neighborhoods to open their new business (e.g. restaurant, hair dresser, bicycle shop, etc.).
- Tourists who would like to enjoy certain types of entertainment, restaurants, etc.

2. Data acquisition and cleaning

Based on the definition of the above problem, factors that will determine the type of ward (local term for neighbourhood) in Bristol are:

- number of venues in neighbourhood
- types of venues in neighbourhood
- quality of life in neighbourhood

2.1. Data sources

The following data sources will be used to extract the required information:

- Foursquare API: Using the Foursquare API allows for getting information on all sorts of venues in the different wards, including the type of venue, nearby venues, etc. That way the types and numbers of venues can be acquired for each ward.
- OpenData Bristol: The webpage <https://opendata.bristol.gov.uk/> offers both a list of postcodes for each ward as well as large datasets on quality of life, air pollution, employment rates, safety, etc. for each ward. Additionally, the coordinates and shapes (GeoJSON file) of each ward can be found and scraped here.

NOTE: These numbers were acquired through surveys of Bristol inhabitants. They, as all qualitative data, are naturally susceptible to all sorts of errors (e.g. coverage error, sample size error, etc.).

2.2. Data preparation and cleaning

The data gained through Opendata Bristol had to be cleaned by dropping unused columns

and data such as the number of councillors, the Ward ID etc. for each ward. Headers had to be added, the wards alphabetically ordered and the index reset. A dataframe with the name of the wards on their coordinates was created.

After survey questions and answers (percentage of agreement) were added, the one-hot-encoding method had to be applied to get a consistent dataframe. It showed the wards, their coordinates, and the percentage of agreement of their inhabitants with the 17 survey questions (e.g. "Do you feel safe after dark?").

The Foursquare data provided information on the number and types of venues in each ward. All in all 818 venues were found in Bristol, consisting of 140 unique categories.

2.3. Feature selection

Since some of the 17 survey questions regarding the quality of life were very similar, features had to be reduced significantly. The 17 features were combined into the following five new features by calculating their mean value):

- Clean streets and air
- Accessible public transport
- Safety
- Children activities
- Feeling Home

For the Foursquare data we focused on the distribution of interesting types of venues in each ward. For families for example Parks and Playgrounds are more of interest than Bus Stops and Coffeeshops.

3. Methodology

The goal of this project was to use both geographical information from Foursquare as well as data on quality of life (from Opendata Bristol) to determine the popularity of the single wards in Bristol. The focus was clearly on individuals and families moving to Bristol and how suitable each ward would be for them.

Naturally, the analysis was limited, in this case to 17 factors describing the quality of life features (taken from questions of a survey) and the venue information provided by the Foursquare database.

In the first step, we collected data on the 34 wards of Bristol, namely their names, coordinates, as well as 17 features each (percentages of YES responses to respective questions). Furthermore, with the help of Foursquare we created a list of 818 venues in Bristol, their categories, names, and locations (wards they belong to).

The second step in our analysis was the reduction of our 17 features to a smaller number of five. Choropleth maps gave us a first idea of suitable areas in Bristol. With the Foursquare data we determined the five most common types of venue for each ward, which allowed us to understand the type of neighbourhood we were looking at.

In the third step we focused on most promising wards and applied some in-depth analysis. We applied k-means clustering to identify clusters of areas most suitable for families with children as well as individuals. We then focused our attention on these clusters of areas and calculated the overall quality of life for each ward in them. This allowed us to rank the wards and decide which wards were the most suited for families and individuals, respectively.

Finally, we visualized our acquired data by presenting a map with the results of our efforts, easily displaying which wards are the ideal ones for individuals or families moving to Bristol (e.g. offering lots of entertainment possibilities or being family-friendly, respectively).

4. Analysis

The five features mentioned in chapter 2.3 gave us a rough idea about the different wards of Bristol according to the 2019/2020 survey. The information acquired via Opendata Bristol was stored in a dataframe containing all 34 wards, see figure 1 for a snippet showing the first 11 wards.

	Ward	Latitude	Longitude	Clean Streets and Air	Accessible Public Transport	Safety	Children activities	Feeling Home
0	Ashley	51.468309	-2.582675	6.960	83.65	73.6425	44.990	69.2425
1	Avonmouth & Lawrence Weston	51.507333	-2.676407	20.495	88.52	77.3025	40.620	54.7025
2	Bedminster	51.437159	-2.621875	17.445	93.18	80.1250	51.630	75.7650
3	Bishopston & Ashley Down	51.481452	-2.586465	13.015	90.61	89.0475	74.840	86.4825
4	Bishopsworth	51.421787	-2.616458	32.635	92.55	83.0650	36.740	66.5850
5	Brislington East	51.440282	-2.538654	24.435	89.38	82.6150	21.965	55.2875
6	Brislington West	51.434392	-2.549071	15.235	89.65	87.6125	42.640	66.5525
7	Central	51.454413	-2.594338	13.460	92.29	76.1850	31.620	60.9825
8	Clifton	51.456514	-2.619966	29.435	89.56	89.5600	52.190	80.6250
9	Clifton Down	51.464968	-2.612659	31.960	86.64	89.1500	49.010	85.5825
10	Cotham	51.465962	-2.599913	30.850	91.52	88.1925	52.325	81.3500

Fig. 1: Snippet of the used dataframe.

This data can be displayed in several ways. A very intuitive way for example are choropleth maps. In figure 2 you can see a choropleth map showing the percentage of people feeling safe for each ward in Bristol.

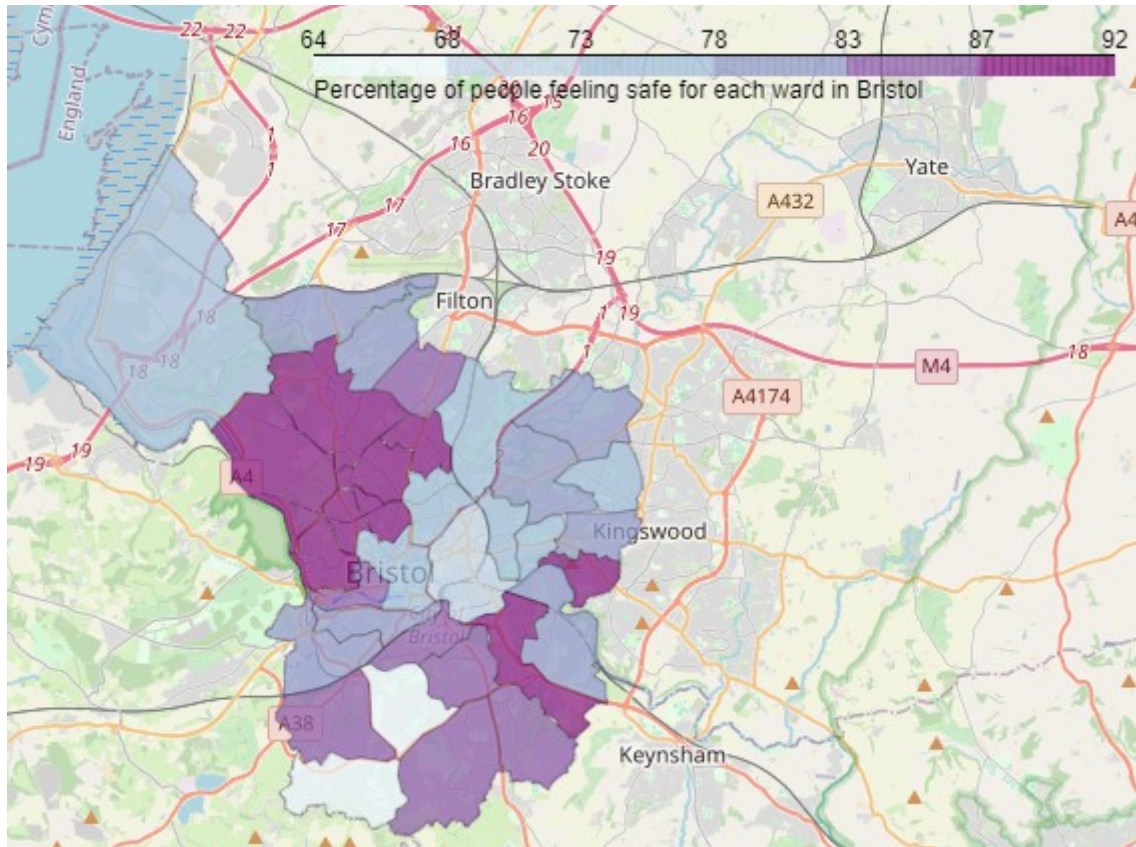


Fig. 2: Percentage of people feeling safe for each ward.

Similar maps can be acquired for the other features as well of course. Even a short overview of the data (see figure 3) yields some insights. For example, even in the cleanest area of Bristol, still only 44% would consider it to be clean. The maximum values for the other four features are much higher, and so is their mean value.

	Latitude	Longitude	Clean Streets and Air	Accessible Public Transport	Safety	Children activities	Feeling Home
count	34.000000	34.000000	34.000000	34.000000	34.000000	34.000000	34.000000
mean	51.459534	-2.584273	21.331912	89.586176	81.521029	47.668382	69.582794
std	0.027113	0.035898	9.048538	3.266449	6.882918	13.198921	11.654197
min	51.407702	-2.676407	6.960000	80.680000	63.655000	10.010000	33.897500
25%	51.439392	-2.611560	14.447500	87.572500	77.310000	40.671250	62.394375
50%	51.461127	-2.586423	19.117500	89.990000	82.357500	48.047500	70.606250
75%	51.476879	-2.551928	25.895000	92.097500	87.295000	54.301250	78.838125
max	51.508972	-2.522576	43.845000	94.040000	92.250000	74.840000	87.662500

Fig. 3: Overview of value distribution of the five features.

With the Foursquare dataset all sorts of venues for each ward were investigated. All in all 818 venues of 141 unique categories were found. Via one-hot-encoding and further

analysis the top 5 most common venues per ward were identified. Figure 4 shows the first five wards and their top 5 most common venues.

	Neighbourhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Ashley	Pub	Café	Bar	Gastropub	Coffee Shop
1	Avonmouth & Lawrence Weston	Go Kart Track	Rental Car Location	Kitchen Supply Store	Bakery	River
2	Bedminster	Pub	Burger Joint	Gym / Fitness Center	Coffee Shop	Supermarket
3	Bishopston & Ashley Down	Café	Pub	Gastropub	Grocery Store	Coffee Shop
4	Bishopsworth	Supermarket	Discount Store	Grocery Store	Bus Stop	Zoo Exhibit

Fig. 4: The top 5 most common venues for each ward. Only the first five wards are shown here.

Combining the data from Opendata Bristol and Foursquare allowed us to apply k-means clustering to further gain information. First, the number of clusters was set to five and k-means clustering applied to the data of venues for the wards of Bristol. The map in figure 5 shows the different clusters on a choropleth map.

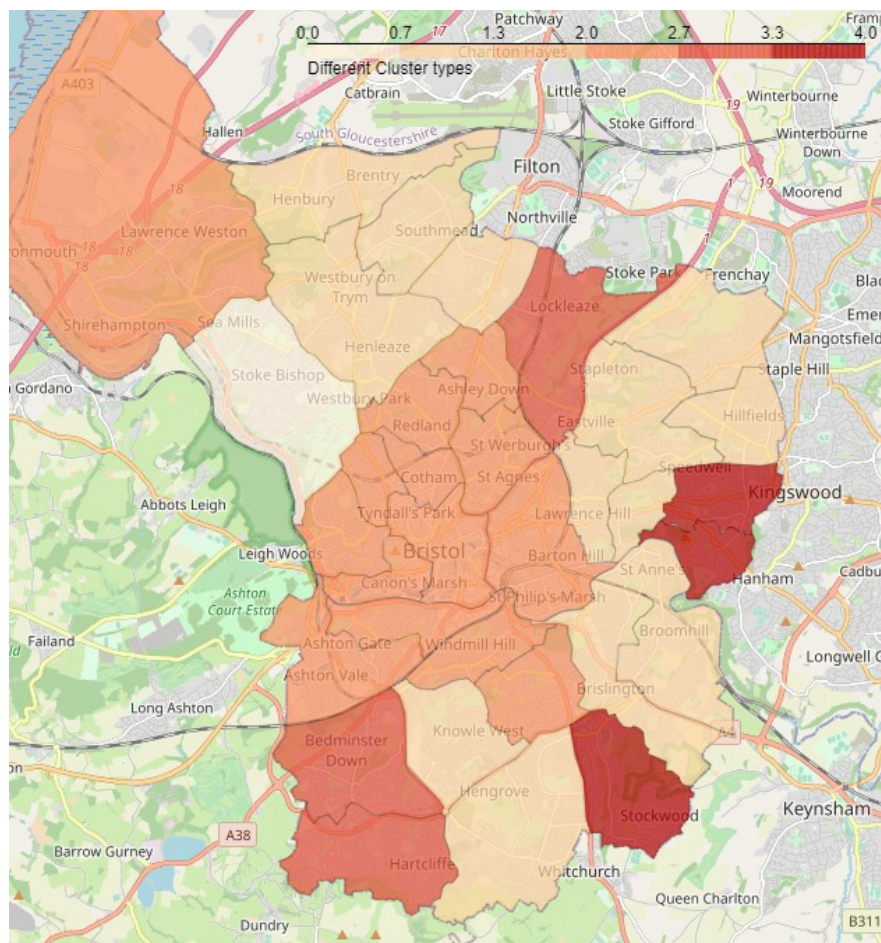


Fig. 5: Choropleth map of the five different cluster types for the venues in wards.

It was interesting to see that there seems to be a clear and detectable difference between the central wards of Bristol (cluster no. 2) and the rest. Most surrounding wards are also similar to each other (cluster no. 1), with some exceptions.

Further examination of each cluster further allowed for determining the discriminating venue categories that distinguish each cluster. The above-mentioned clusters no. 1 and no. 2 turned out to be very interesting.

Cluster no. 1 is made up by 13 wards surrounding central Bristol. According to the top 5 venues, these are wards with supermarkets and grocery stores, pubs and restaurants, so rather mixed areas where people live in relative quiet.

Cluster no. 2 is made up by 13 wards in the centre of Bristol (plus "Avonmouth & Lawrence Weston" as a literal outlier). According to the top 5 venues, this is a typical inner city area, full of pubs, cafés/coffeeshops, and fast food places such as pizza and burger joints (see figure 6.)

	Ward	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	Cluster Labels
0	Ashley	Pub	Café	Bar	Gastropub	Coffee Shop	2
1	Avonmouth & Lawrence Weston	Go Kart Track	Rental Car Location	Kitchen Supply Store	Bakery	River	2
2	Bedminster	Pub	Burger Joint	Gym / Fitness Center	Coffee Shop	Supermarket	2
3	Bishopston & Ashley Down	Café	Pub	Gastropub	Grocery Store	Coffee Shop	2
7	Central	Bar	Café	Coffee Shop	Pub	Asian Restaurant	2
8	Clifton	Pub	Café	Park	Italian Restaurant	Bar	2
9	Clifton Down	Pub	Bar	Café	Coffee Shop	Italian Restaurant	2
10	Cotham	Pub	Coffee Shop	Bar	Deli / Bodega	Burger Joint	2
20	Hotwells & Harbourside	Café	Pub	Bakery	Coffee Shop	Cocktail Bar	2
21	Knowle	Café	Convenience Store	Grocery Store	Fast Food Restaurant	Discount Store	2
22	Lawrence Hill	Pub	Café	Gym / Fitness Center	Coffee Shop	Brewery	2
24	Redland	Pub	Bar	Coffee Shop	Café	Pizza Place	2
26	Southville	Pub	Coffee Shop	Café	Restaurant	Pizza Place	2
33	Windmill Hill	Pub	Park	Fast Food Restaurant	Bar	Café	2

Fig. 6: Dataframe of cluster no. 2, the typical inner city area with many entertainment options.

So by combining all this data it seemed that cluster no. 1 (areas surrounding the central part of town) seems suitable for families with children moving to Bristol, whereas cluster no. 2 (centre of Bristol) seems to be more interesting to individuals because of its nightlife options.

Our further focus was on these two clusters. Both families with children and single people or couples without children prefer areas with high quality of life, obviously. This means we focused on our five features ("Clean Streets and Air", "Safety", etc.) - resp. four features for individuals, since they are not interested in "Children Activities" - that indicate a high quality of life next.

Since we were looking at a selection of the 34 wards of Bristol only, we normalised the feature data first. This should not be a problem, since we were interested in the optimum value of a limited data set (the respective cluster), so we weren't interested in absolute values, but in relative ones (meaning we can use maximum values for normalisation). Simple feature scaling was therefore sufficient here.

After normalising the feature data, we calculated a weighted mean of the features (the weights were of course rather subjective). See figure 7 for the selected weights.

Feature	Weights_Families	Weights_Individuals
Clean Streets and Air	0.2	0.2
Accessible Public Transport	0.1	0.2
Safety	0.2	0.3
Children activities	0.3	0
Feeling Home	0.2	0.3

Fig. 7: Weights for calculating the feature means. The weight for "Children activities" is obviously set to zero for individuals.

That allowed us to calculate the top 5 wards in terms of quality of life for both the family-friendly cluster (figure 8) as well as for the one most suited for individuals (figure 9).

	Ward	Longitude	Latitude	Clean Streets and Air	Accessible Public Transport	Safety	Children activities	Feeling Home	Quality of Life
32	Westbury-on-Trym & Henleaze	-2.617485	51.489794	0.200000	0.098947	0.200000	0.269832	0.200000	0.968779
29	St George West	-2.546940	51.460116	0.061391	0.100000	0.173513	0.300000	0.175475	0.810379
16	Henbury & Brentry	-2.621037	51.508972	0.136848	0.092822	0.176592	0.213347	0.171781	0.791391
19	Horfield	-2.586381	51.495083	0.085211	0.094715	0.187003	0.231411	0.187687	0.786027
14	Frome Vale	-2.531200	51.487089	0.110398	0.099766	0.176093	0.213368	0.160902	0.760528

Fig. 8: Top 5 wards in terms of "Quality of Life" for families with children.

"Westbury-on-Trym & Henleaze" is the clear winner here. "St George West" might have also been considered if only the streets and air were cleaner there.

	Ward	Longitude	Latitude	Clean Streets and Air	Accessible Public Transport	Safety	Feeling Home	Quality of Life
9	Clifton Down	-2.612659	51.464968	0.200000	0.185963	0.290659	0.292882	0.969504
24	Redland	-2.601160	51.475244	0.159606	0.196179	0.300000	0.300000	0.955785
10	Cotham	-2.599913	51.465962	0.193054	0.196437	0.287537	0.278397	0.955425
8	Clifton	-2.619966	51.456514	0.184199	0.192230	0.291996	0.275916	0.944341
3	Bishopston & Ashley Down	-2.586465	51.481452	0.081446	0.194484	0.290325	0.295962	0.862216

Fig. 9: Top 5 wards in terms of "Quality of Life" for individuals.

The top 4 wards for individuals are very close to each other, both in terms of the feature "Quality of Life" as well as geographically. It seems that the differences are only minute, and that individuals might be as happy in "Clifton Down" as they'd be in "Redland", "Cotham" or "Clifton".

Finally, two more choropleth maps displaying that data, see figure 10 and figure 11.

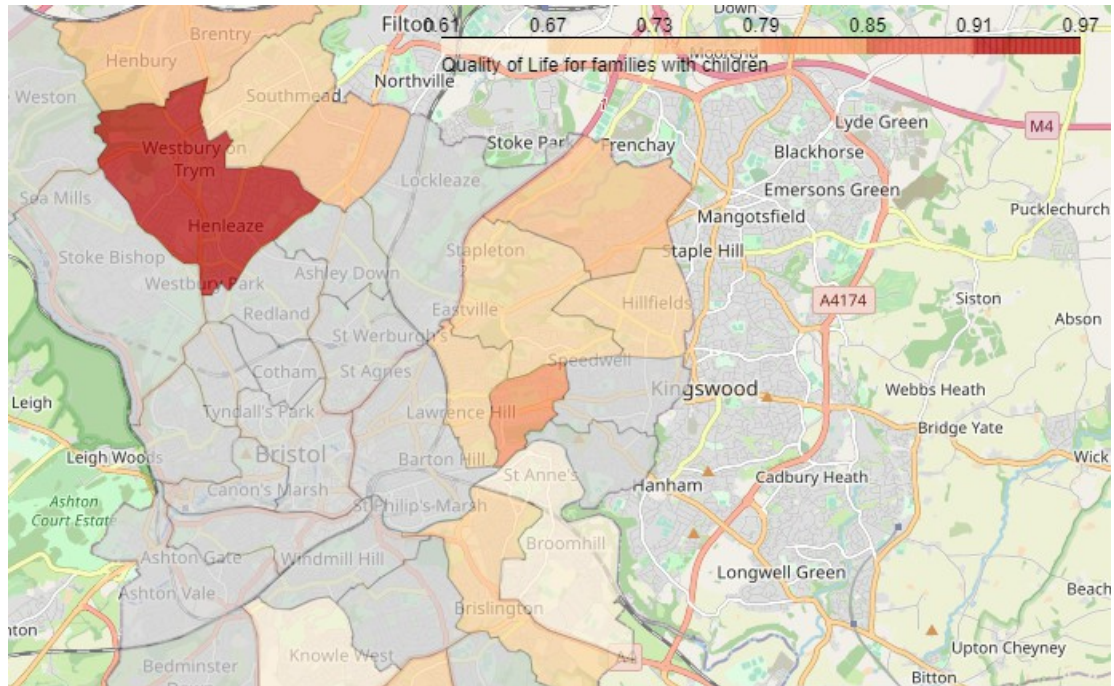


Fig. 10: Map of Bristol showing the "Quality of Life" for families with children for cluster no. 1. "Westbury-on-Trym & Henleaze" is the clear winner here.

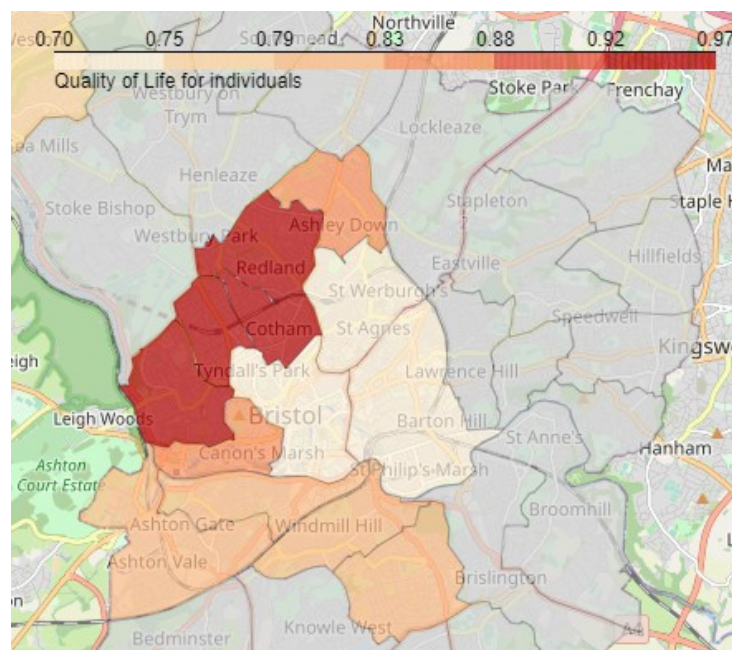


Fig. 11: Map of Bristol showing the "Quality of Life" for individuals for cluster no. 2. The four dark red wards are very close both in "Quality of Life" and geographically.

5. Results and Discussion

The final results show that some areas and wards are clearly more suitable for families resp. individuals than others. Large differences can be seen both in terms of the number and types of venues in each ward (data acquired through Foursquare) as well as when it comes to the perceived quality of life there (data acquired through surveys and published on opendata.bristol.gov.uk).

Looking at the venues and their distribution, it soon showed that areas with a high entertainment value (high density of pubs, restaurants, cafés, etc.) are very central in Bristol. Just northwest of the centre of town are the areas most suitable for individuals due to a very high rating in terms of quality of life. They are "Clifton Down", "Redland", "Cotham", and "Clifton".

Even more, the ward with the highest quality of life for families with children - "Westbury-on-Trym & Henleaze" - is directly neighbouring that area (in general it seems that family-friendly wards are just a little bit outside the town centre). It seems that this whole part of town offers great possibilities and a high quality of life for its inhabitants, no matter if they are singles or families.

Our results clearly suggest ideal wards for both individuals and families when moving to Bristol. It should be mentioned, however, that our analysis is only scratching the surface.

There are several issues to consider. First off, the Foursquare venue data might not be complete and the results therefore somewhat inaccurate. Also, the data from opendata.bristol.gov.uk was gained by the Bristol Council through surveys on their population. These numbers, as all qualitative data, are naturally susceptible to all sorts of errors (e.g. coverage error, sample size error, etc.). Finally, the weight of the quality of life calculation and other preferences are rather subjective and a matter of taste (e.g. since some people might prefer different venues over others or have different priorities when it comes to a "suitable neighbourhood").

So in the end this analysis does not imply that these wards are by far the best places to live in Bristol. And even if this was true, factors like rental prices - which can make quite a difference for a family looking for a new home - weren't even considered here. Still, our results should work at least as a rough guide for newcomers to Bristol that would like to know what certain wards are like, and where to start exploring Bristol and hopefully enjoying it the most.

6. Conclusion

The purpose of this project was to identify the most suitable neighbourhoods of Bristol (UK) for both individuals and families with children, and especially for newcomers to the city that ask themselves: where to start?

Since knowledge about the different parts of town can be quite helpful to decide in which neighbourhood to live in for the next few years, a careful analysis of the single areas of Bristol was conducted. Foursquare data allowed to estimate how suitable a ward is for individuals resp. families, whereas data acquired through opendata.bristol.gov.uk helped to find out about the areas with the highest perceived quality of life. The optimum wards for both target audiences were determined and shown in both rankings (tables) as well as on

choropleth maps of Bristol.

In this analysis, my own preferences shaped the outcome, but the final decision on where to move in Bristol of course depends on the individuals or families themselves, their priorities and preferences in terms of their environment, their financial position, and of course their idea of what "Quality of Life" means in the end.