Scenario Week, November
2020
Word account: 2000
Data Analytics II

DATA ANALYTICS REPORT

SCENARIO WEEK 2

Hans-Alexander Miguel Viessmann
Francois Barone - 19113943
Oliver Haulund - 19002407
Marek Istok - 19025338
Gabrielius Valiunas - 19020857
Ninon Lavolle - 19098458
Bhavya Gupta - 19073155

THE BOROUGH THAT FITS YOU

DATA ANALYTICS REPORT

2020

Table of Contents

Business Model

Our business

Borough that fits you recommends tailor fit areas to our users. London is one of the cities with the lowest life satisfaction in the UK (Dourou, 2019). At the same time, it is argued that residential environment is an important factor to take into account when it comes to assessing our well-being (Lee, 2020). Thus, we considered using open data from London boroughs and wards to develop a business with the aim of helping people who are new to London find the best place to live according to their personal preferences.

First, we evaluate each area according to different categories corresponding to different characteristics of an area such as safety, entertainment or facilities. Through data visualisation and analysis, we evaluate each area's performance in every category. In order to provide a complete and personalized service to customers, they have the option of weighting the categories by rating their importance on a scale of 1 to 10.

Our website

The website gives a personalised, interactive and intuitive experience to the customer[1]. In our final personalised map, the colour of each ward corresponds to the compatibility rate. The higher the match, more intense is the colour. By clicking on an area, the customer gets the compatibility rate, information about this area and our inferences. To ensure customer retention and to help newcomers explore the new area, we provide newsletters about activities and new businesses in the chosen area.

The use and visualisation of open data is the key asset of our recommendations model and interactive maps, which ultimately provide added value to the customer in terms of variety (we consider 7 categories and 25 sub-categories), accuracy (most of our datasets include 1000+ observations across 380+ local authorities) and precision

[1] See it visualised in Appendix 5 – Website mock-up

(data granularity spans from boroughs to wards and output areas [2]). For our business model, the free access and use of open data allows us to reduce costs significantly, while the accuracy of our recommendations model and suggestions is leveraged to show targeted advertisements of local businesses according to the areas the customer is interested in. This gives a break-even at a fairly low revenue generation, which is expected to be around just £38,000 the first year and £68,000 the second year[3].

Data's role in our business

Initially, we considered running our business in the United Kingdom as a whole, thinking that the larger the market, the higher the revenue. Upon searching for datasets, we observed that for some metrics of assessing an area, the geographical scale was mostly districts, or even regions in the case of crimes[4] . This is quite broad and imprecise. We then turned to only London data and noticed that they were more precise with wards, postcodes, or even geographic coordinates. Our research on immigration to the United Kingdom revealed that London had 450,000 people moving in per year (Office for National Statistics, 2020a). After careful consideration, our group concluded that we would have more impact by providing a quality offer to fewer customers, rather than a fairly generic tool for more people.

We provide an overview of the standard of living by area with the help of a dataset on real estate prices (Office for National Statistics, 2020c). This is essential as income and spending power of people vary. We also considered ease of transport using 3 datasets showing the location of metro stations (Department for Transport, 2014), bus stops (Department for Transport, 2014), and self-service bicycle points (Au, 2019). We picked these three because they are in the top 5 of the most used means of transport in London with individual cars and taxis which are accessible from anywhere (Transport for London, 2019).

[2] A smaller part of a ward. See Appendix 6 – Area types, for clarification of area codes
[3] See more in Appendix 1 – Business Model Canvas
[4] See Appendix 3 – Crime rate by area

Then, we looked at health factors using NHS datasets providing the locations of hospitals, GPs, Dentists and pharmacies (NHS, 2020). Indeed, people with health problems or high risk of accidents might be reassured by being close to a medical point.

To assess quality of life, we focused on the presence of everyday essential facilities such as markets (Greater London Authority, 2018a), supermarkets (Geolytix, 2020), parks (Greater London Authority, 2018b), restaurants (Food Standards Agency, 2018), and entertainment with museums , cinemas, theatres, galleries, casinos, pubs, clubs and dance performances (Greater London Authority, 2018c) (Gambling Commission, 2020).

Other categories included are education and safety. We considered where schools (Greater London Authority, 2018d) and libraries (Collections Trust, 2012) are located for education, and a dataset containing several types of crimes to form a global score for safety (Metropolitan Police Service, 2020).

Finally, our model helps take into account the distance between two points. University and jobs are the main reason for people moving to London (Blinder, 2018), and they usually know where they will study or work. This feature is a real value-add to our offering and a way to personalize the experience even more.

Lastly, we used a dataset to manipulate the data we collected (Greater London Authority, 2014). It provides information on each borough of London, including descriptive figures like the number inhabitants or total area. For instance, this helped us to calculate the density of metro stations per borough.

How data was used

Why focus on London?

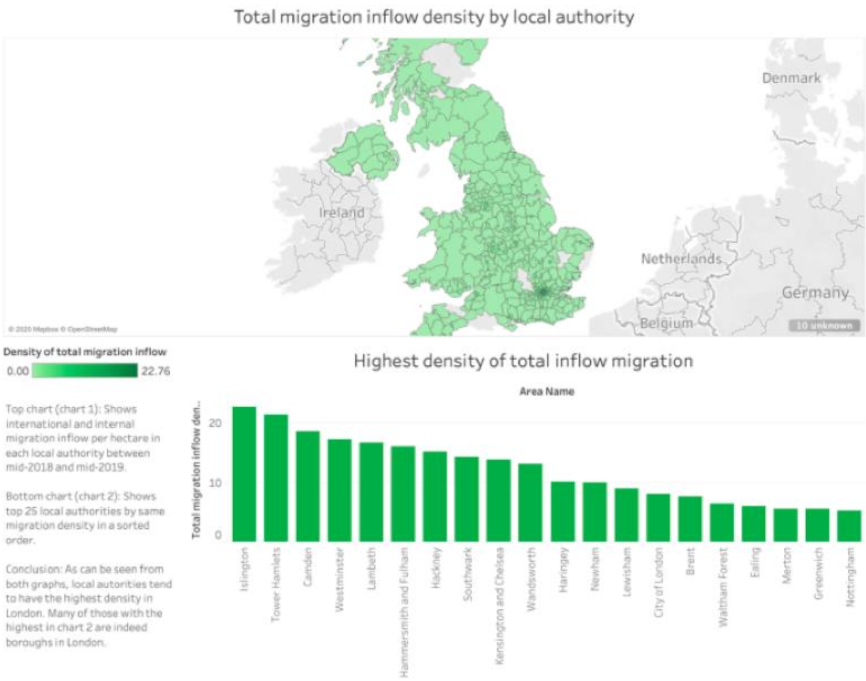
To understand where the service makes the most sense commercially, the density of long-term (12+ months) migration inflow is used (Office for National Statistics, 2020a) (Office for National Statistics, 2020b)[5]. This, as migration inflow itself would disregard the variety in area size of local authorities, which is eliminated by

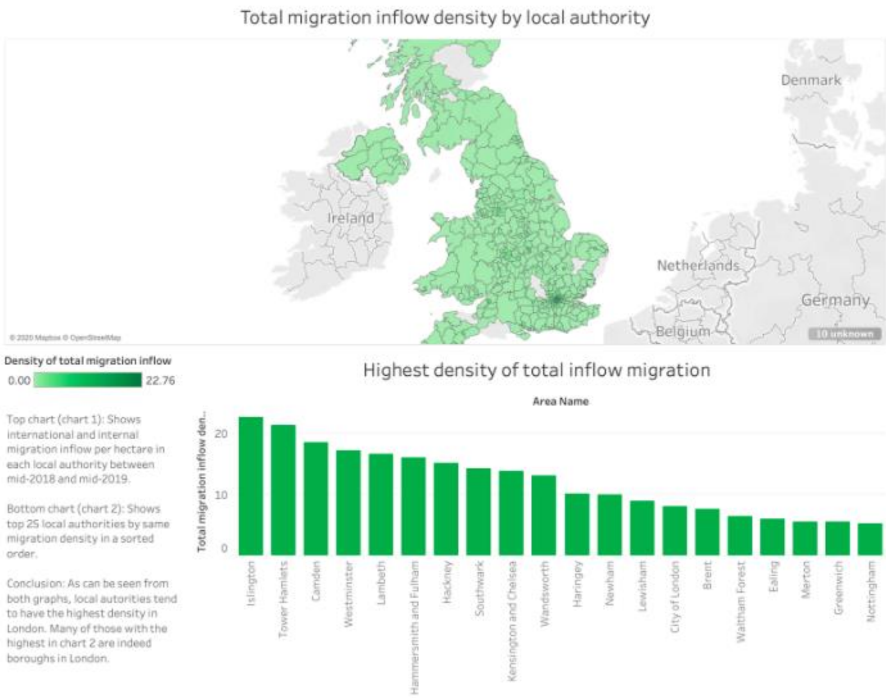
5 (Long-term international migration inflow + Long-term internal migration inflow) / Area to Mean High Water Excluding Area of Inland Water

density. Using migration inflow density, we can by chart 1 in visualisation 1 see that most of the local authorities with the highest density are those in or around London.

Chart 2 is also confirming this with a range of London boroughs in the top 20 of local authorities by the densest migration inflow, while it also shows that there is a significant variation in density even amongst the top 20.

Visualisation 1 – Charts on migration inflow density



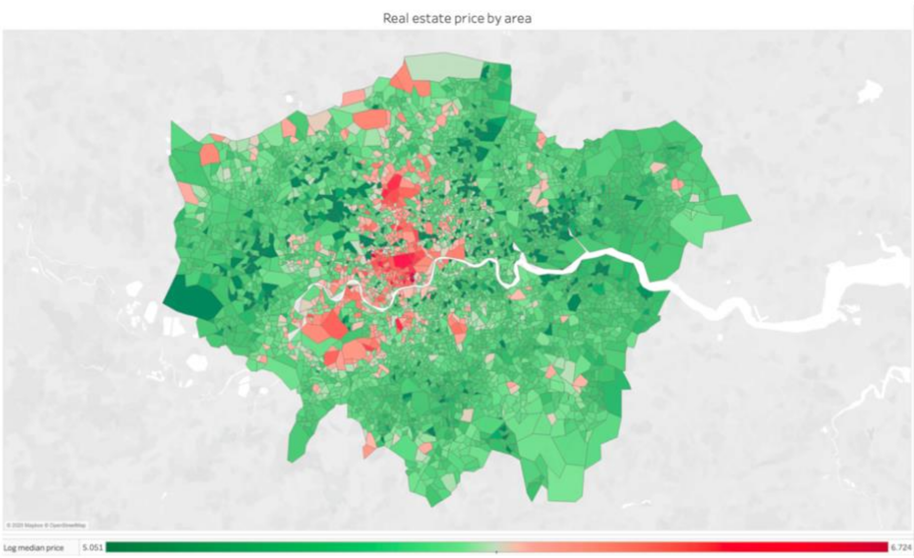


While London is the most attractive market today, another factor important in the decision is how migration inflows will change in the nearest future. Thus, using the development in migration inflow the past 3 years[6] as a proxy for future developments, the growth expectations were calculated (Office for National Statistics, 2020a). In chart 1 in visualisation 2, it is seen that the local authorities in and around London all have seen a fairly good growth over the past 3 years, and chart 4 shows that none of them have experienced negative growth.

Additional value from open-source data sets

While it is evident from "Data's role in our business" how a large range of data sets are adding value in our calculations, these are only being visualised combined in the end. However, we have in our work with mapping down the usefulness of the data sets that we have included visualised these, to understand their importance. Some examples of these includes:

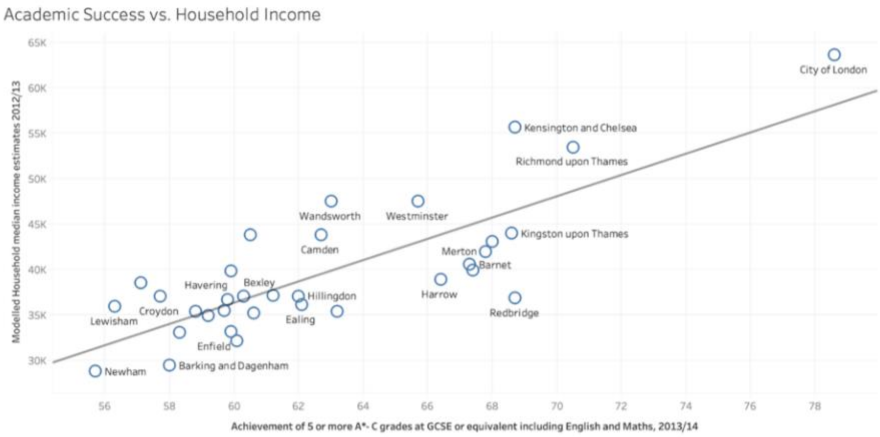
Visualisation 5 – The real estate price by area



(Office for National Statistics, 2020c)

[6] Mid 2015/16 to mid 2018/19

Visualisation 6 – Academic success vs. household income by borough



(Greater London Authority, 2014)

Technical Considerations

Datasets and challenges

Hospitals (Health)

Health datasets (NHS, 2020), such as the hospitals in figure 4.1 (see next page), only included postcodes as their address. Our goal was to calculate a density score for each LSOA. We did that in the following way:

1. Converted postcodes to LSOA codes using the London Postcode Directory provided by the Office For National Statistics 2020d
2. Count the total number of health facilities in each area by combining multiple datasets (hospitals, GPs, dentists, …)
3. Calculated the density of health facilities for each LSOA
4. Standardizing the score to our normalized 0 to 100 score

Cinemas (Entertainment)

We considered 5 different datasets similar to the cinemas dataset (Greater London Authority, 2018c) (see figure 4.2 in the next page) when calculating the entertainment score. These datasets were pretty similar, and all came with precision at the level of wards. We calculated the amount of entertainment venues (cinemas, theatres, etc) in each ward and calculated their density. We then standardized an entertainment score for each ward (see figure X in the next page). However, we saw that when comparing the areas on a map, Westminster had such a high score that all the other areas were close to zero as seen in figure 4.2.1.

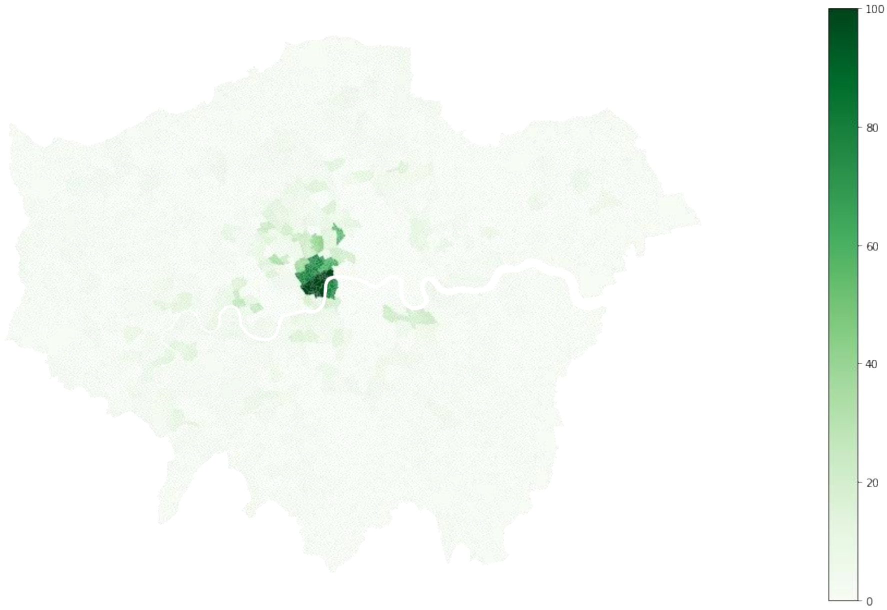


Figure 4.2.1, Map of Standardized Entertainment Score

To fix this issue, we took log of the score, which makes it more comparable as seen in figure 4.2.2

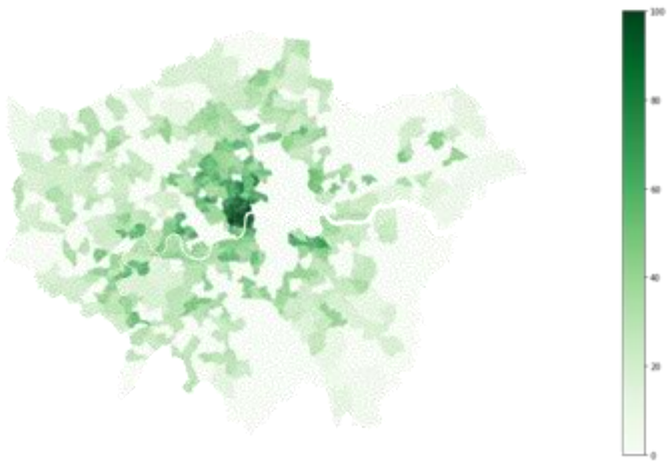


Figure 4.2.2, Map of Standardized Log Score of Entertainment

Crime Rate

The crime rate dataset (Office for National Statistics, 2020d) was one of the easiest ones to work with. It came with area codes at the LSOA level, so it was relatively easy to integrate with our framework. The values were also already normalized to per population levels, so we didn't have to do any special calculations (see figure 4.3 in the next page).

In general, we mainly faced these challenges when working with the datasets:

- Various area accuracy levels - postcodes, geographic coordinates, LSOA codes, etc. This required us to use an additional dataset (see Appendix 4.1) to convert, for example, postcodes into OA area codes, which only then could be used for further analysis (we used Pandas to complete conversion).

- As datasets usually contained a lot of information which had to be either deleted or reformatted, we used python to clean them for further use. For example, in figure 4.1, we deleted unnecessary information such as Addresses, Sectors, Parent Organisation, Phone, Email, etc.

- Another issue stemmed from the fact that some datasets did not have a value for all areas as seen in figure 4.4. For example, our hospitals dataset was missing values for quite a few. We fixed this by filling the missing entries with the mean value for said category. This minimizes the impact missing data has on the final score.

E00001848	E01000378	E05000071	Bexley	15.3	0.10598	0.093457944
E00001685	E01000340	E05000066	Bexley	4.1	0.195406667	0.093457944
E00001604	E01000326	E05000064	Bexley	5.4	0.0548	0.142857143
E00001686	E01000340	E05000066	Bexley	4.6	0.200946666	0.163934426
E00001607	E01000326	E05000064	Bexley	11.9	0	
E00002172	E01000440	E05000080	Bexley	5	0.259153334	
E00002250	E01000461	E05000083	Bexley	16.5	0.152053333	

Figure 4.4, Missing values in Facilities dataset

We used over 20 different datasets in our analysis, here are samples of a couple of them.

OrganisationID	Organisi	OrganisationType	SubType	Sector	Organis	IsPimsManaged	OrganisationName	Address1	Address2	Address3	City	County	Postcode	Latitude	Longitude
17970	NDA07	Hospital	Hospital	Independent Sector	Visible	TRUE	Walton Community Hospital - Virgin Care Services Ltd		Rodney Road		Walton-on-Thames	Surrey	KT12 3LD	51.379997253418000	-0.40604206919670100
17981	NDA18	Hospital	Hospital	Independent Sector	Visible	TRUE	Woking Community Hospital (Virgin Care)		Heathside Road		Woking	Surrey	GU22 7HS	51.315132141113300	-0.55628949403762800
18102	NLT02	Hospital	Hospital	NHS Sector	Visible	TRUE	North Somerset Community Hospital	North Somerset Community Hospital	Old Street		Clevedon	Avon	BS21 6BS	51.43719482421880	-2.8471927642822300
18138	NMP01	Hospital	Hospital	Independent Sector	Visible	FALSE	Bridgewater Hospital	120 Princess Road			Manchester	Greater Manchester	M15 5AT	53.459743499755900	-2.2454688549041700
18142	NMV01	Hospital	Hospital	Independent Sector	Visible	TRUE	Kneesworth House	Old North Road	Bassingbourn		Royston		SG8 5JP	52.078121185302700	-3.06040551513433E-02
18143	NMV02	Hospital	Hospital	Independent Sector	Visible	TRUE	Stockton Hall Hospital	Stockton Hall	The Village	Stockton On The Forest	York	North Yorkshire	YO32 9UN	53.995403289794900	-1.0025526285171500
18271	NQ106	Hospital	Hospital	Independent Sector	Visible	TRUE	Fryatt Hospital, Harwich	419 Main Road			Harwich	Essex	CO12 4EX	51.934696197509800	1.261444091796880
18272	NQ108	Hospital	Hospital	Independent Sector	Visible	TRUE	Clacton Hospital		Tower Road		Clacton	Essex	CO15 1LH	51.786079406738300	1.148187518119810

Figure 4.1 Hospitals dataset

name	borough_name	website	os_addressbase_uprn	borough_code	ward_2018_name	ward_2018_code	address1	address2	address3	longitude	latitude	easting	northing	runtime
Cineworld Cinema London Wembley							Stadium Way			-0.28371052470183800	51.55685619721720	519078.5460051070	185614.34662150300	11/22/2020
Showcase Newham	Newham		10008989589	E09000025	Beckton	E05000475		Jenkins Ln		0.08120937490833860	51.52634159580010	544470.9608409240	182877.92324266900	11/22/2020
David Lean Cinema	Croydon		200001222331	E09000008	Fairfield	E05011468		Katherine St		-0.09898276583837270	51.37216277286960	532416.9641191130	165392.92799365400	11/22/2020
Vue Croydon	Croydon		10001004802	E09000008	Fairfield	E05011468		14 High St		-0.10106037448928100	51.37299684357660	532269.9641586470	165481.92796988000	11/22/2020
Vue Croydon Purley Way	Croydon		100022906148	E09000008	Broad Green	E05011465	Valley Park Leisure Complex	21 Hesterman Way		-0.12450070304906800	51.37812234849380	530623.9646087350	166009.92782629300	11/22/2020
Odeon Beckenham	Bromley		10003626016	E09000006	Copers Cope	E05000113		The High Street		-0.031376362986825400	51.407106814193100	537017.9628681400	169402.92690423900	11/22/2020
Odeon Kingston	Kingston upon Thames		128004717	E09000021	Grove	E05000408		Clarence Street		-0.2993418794085	51.41170694687230	518369.96793997100	169447.92689036300	11/22/2020
Odeon Wimbledon	Merton		48105080	E09000024	Dundonald	E05000459		39 The Broadway		-0.20489884217573900	51.41961437263560	524915.9661586690	170483.92660923200	11/22/2020

Figure 4.2 Cinemas dataset

Code	Borough	Year	Offences	Rate	Number_of_offences
E09000002	Barking and Dagenham	1999-00	All recorded offences	120.5	19,567
E09000003	Barnet	1999-00	All recorded offences	98	30,708
E09000004	Bexley	1999-00	All recorded offences	95.1	20,680
E09000005	Brent	1999-00	All recorded offences	127.7	33,253

Figure 4.3 Crime Rate dataset

Merging Data

After normalizing our datasets to a common area format, we joined them using Pandas' Merge function. We combined different datasets to provide a total score for any specific area. For example, for Health facilities we merged datasets regarding GPs, Hospitals, Dentists and Pharmacies. This allowed us to calculate a total score for this category. Similar process was done for all other categories as well. See Figure 4.6.

OA11CD	LSOA11CD	WD11CD_BF	LAD11NM	Total	Score	ScoreLog
E00023264	E01004612	E05000626	Wandsworth	0	0	0
E00003359	E01000692	E05000111	Bromley	1	0.54709719	9.45544958
E00023266	E01004615	E05000626	Wandsworth	0	0	0
E00020264	E01004027	E05000548	Southwark	0	0	0
E00023263	E01004613	E05000626	Wandsworth	0	0	0

Figure 4.6 - Example for the entertainment category.

In the end, we merged scores from all the categories to one final data frame, which we use to calculate customized scores based on our user's preferences. See Figure 4.7.

OA11CD	LSOA11CD_x	WD11CD_BF	WD11NM_B	LAD11CD	LAD11NM_x	area	EntScore	EntScoreLog	FacScore	TransportScc	CrimeScore	HealthScore	EduScore	CostScore	DistScore	Score	ScoreLog
E00023264	E01004612	E05000626	Tooting	E09000032	Wandsworth	4.10	8.05	2.20	13.19	35.17	4.69	18.85	0.00	4.87	63.43	32.50	76.09
E00003359	E01000692	E05000111	Chislehurst	E09000006	Bromley	7.30	0.47	0.39	12.14	18.85	1.78	21.51	9.13	3.25	68.20	31.35	75.33
E00023266	E01004615	E05000626	Tooting	E09000032	Wandsworth	1.50	8.05	2.20	15.30	35.17	3.94	21.51	0.00	1.19	61.40	34.62	77.42
E00020264	E01004027	E05000548	Riverside	E09000028	Southwark	4.40	24.81	3.25	13.69	58.76	16.30	11.62	18.45	3.98	88.81	52.27	86.14
E00023263	E01004613	E05000626	Tooting	E09000032	Wandsworth	1.90	8.05	2.20	13.19	35.17	9.44	28.75	0.00	3.85	64.15	34.54	77.37

Figure 4.7 - Combined dataset with scores for all the categories.

Challenges in Python and Tableau

Python:

- Merge datasets across area levels, which sometimes left gaps in datasets (see figure 4.4)

- Standardisation of data as datasets had very different magnitudes. We ended up normalising data from 0-100 in every step.

Tableau:

- To assign geographic roles, we needed to unify external datasets so that the area codes would represent the correct location.

- Along with local super output areas, we tried showing boroughs on the map using borders for clear understanding. However, it was not possible to do so. Additionally, some graphs were poor quality due to missing values in the dataset we were looking into.

- Lastly, for using Tableau in our real-life product, we would need to automate importing Python calculations into Tableau and then embed the visualisations in our website.

Conclusion

In conclusion, thanks to various and numerous open data we were able to propose our business idea without making any concession on the quality of the chosen categories. Also, the accuracy of the data in terms of geographical scale allowed us to build a precise model by breaking down London into small geographical areas. In terms of data manipulation, Python and Tableau allowed us to exploit our data, standardize our model and offer relevant and intuitive visualizations. We believe that the accuracy of our offering and the advertising campaigns we want to run will allow us in the short term to significantly increase our number of customers, and thus our revenues.

The limitations that our business is facing are the ability to retain our customers and the presence of 2 datasets that are 8 and 10 years old, whereas many changes can have occurred in the city since then. In the future, the major challenge will be to have recent data in order to propose a model that is always up to date. Our development objectives are to offer an even more personalized experience to our consumers while keeping it simple to use, to find a way to retain our consumers over the long term, and to extend our offer to another Megalopolis such as Paris or New York.

Appendix

Appendix 1 – Business Model Canvas

PROBLEM

-Choosing an appropriate area to live in London

-Unclarity about activities and facilities in a determined area

-Complex gathering, reading & understanding of the open data available

EXISTING SOLUTIONS

-Randstad's "What Is The Best Borough For You?" platform

-Accommodation forums (The Student Room)

-Real estate agencies & forums

SOLUTION

- Customized area recommendations

-Activities suggestions by borough

-Graphs/ visualisations leveraging various datasets

KEY METRICS

-Website visits

-Conversion rate

-Newsletter subscriptions

-Generated revenue

UNIQUE VALUE PROPOSITION

-Helping customers with an important decision

-Considering a large range of categories (7), made of 25 sub-categories

-Datasets with several hundreds of observations at least, providing a high accuracy degree

-Different types of graphs/visualisations

-High level of precision from boroughs to output areas

UNFAIR ADVANTAGE

-Preferences customisation service

-Recommendation model

-Weekly updates with customized offers for members

CHANNELS

-Social media presence

-Online & outdoor advertising

CUSTOMER SEGMENTS

-London newcomers: university students, employees looking for a career change...

- 18-35 years old: main age target group

- Families: secondary target

- Market size of 450,000 people/year [1]

COST STRUCTURE

- Website development & maintenance: £10,000 [2]

- Advertising campaign: £5,000 [3]

- Workforce (CEO, marketing & PR personnel, data scientists): £7,500/month

FINANCIAL YEAR

-Targeted advertising based on people's preferences & area recommendations

-Estimated annual revenue: £38,000 (Y1), £68,000 (Y2[7])

[7] See more in Appendix 2 – Revenue estimation

Appendix 2 – Revenue estimation

In order to determine our expected annual revenue for Year 1 and Year 2, we had to consider market share, unique users and page views as key performance indicators.

Firstly, considering an approximate target market of 450,000 people/year, [4] we estimated a reach of 10% for the first year and 20% for the second year. We attribute this growth in unique users to our advertising campaigns and the general expansion of our business as people get familiar with it.

Drawing a comparison with the news website My London, which notifies its users of relevant news in their area, and averages 6.7m unique users and 80m monthly page views, [5] we estimated 10 page views per user for our revenue calculation.

Finally, we used Google AdSense [6] to obtain our expected revenue:

Monthly views - 450,000*[10% in Year 1, 20% in Year 2] * 10/12 =

[37.500 monthly users in Year 1, 75,000 monthly users in Year 2]

First year revenue - $4,000 a month = $48,000 a year = £38,000

Second year revenue - $7,500 a month = $90,000 a year = £68,000

Appendix 3 – Crime rate by area

Area Code	Area Name	Total recorded crime (excluding fraud[3])	**Violence against the person**	Homicide	Violence with injury	Violence without injury	Stalking and harassment[4]	Death or serious injury - unlawful driving	**Se offer**
K04000001	**ENGLAND AND WALES [6]**	**5032544**	**1750750**	**725**	**512743**	**717824**	**518688**	**770**	**152**
E92000001	**ENGLAND**	**4724104**	**1643872**	**689**	**483459**	**675647**	**483337**	**740**	**142**
E12000001	**North East**	**265771**	**94697**	**27**	**24017**	**34642**	**35996**	**15**	**7**
E23000013	Cleveland	67183	23675	11	5485	9135	9040	4	1
E23000008	Durham	61889	25891	6	5428	8741	11705	11	1
E23000007	Northumbria	136699	45131	10	13104	16766	15251	0	4
E12000002	**North West**	**675164**	**243216**	**114**	**67484**	**101174**	**74289**	**155**	**20**
E23000006	Cheshire	92174	38613	13	9910	13901	14784	5	2
E23000002	Cumbria	36137	15408	7	4410	5889	5085	17	1
E23000005	Greater Manchester	278374	85465	46	23911	38347	23097	64	8
E23000003	Lancashire	134788	55076	28	15051	22734	17223	40	4
E23000004	Merseyside	133691	48654	20	14202	20303	14100	29	3
E12000003	**Yorkshire and The Humber**	**555135**	**204090**	**61**	**55941**	**76877**	**71076**	**135**	**16**
E23000012	Humberside	92732	33488	10	10315	13220	9913	30	2
E23000009	North Yorkshire	44129	16386	6	5993	6477	3883	27	1
E23000011	South Yorkshire	143171	47865	18	14017	15441	18369	20	4

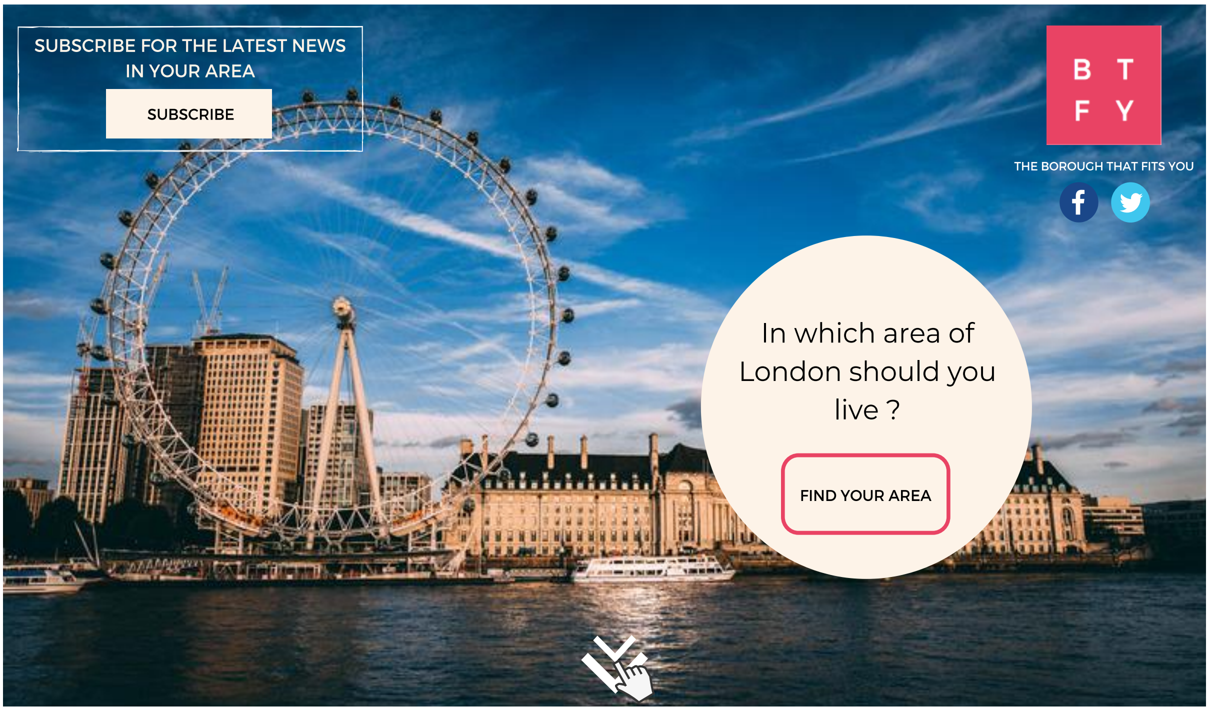
Appendix 4 – Competitor comparison: What is the best borough for you?

Category	Added Value
Granularity	WBFY shows results at output area scale, while Randstad's tool only covers boroughs. Therefore, WBFY offers more precision.
Preference Categories	Randstad's website only allows users to choose different filters, but not score or rate them. That is, it doesn't consider all filters, and it does so in an absolute way. In comparison, WBFY's approach ensures that the user picks the relative importance of all categories, resulting in a higher accuracy.
Map Visualisation	WSBY's result map is interactive and allows the user to look at the data and performance of each area by hovering over them. Randstad's map is fixed and doesn't highlight the recommended borough.
Results Layout	Randstad's results section is poorly optimised (missing pictures, inoperative buttons, results 2-18 not shown), heavily damaging the user experience. WSBY's website has a simple, intuitive and up-to-date design.

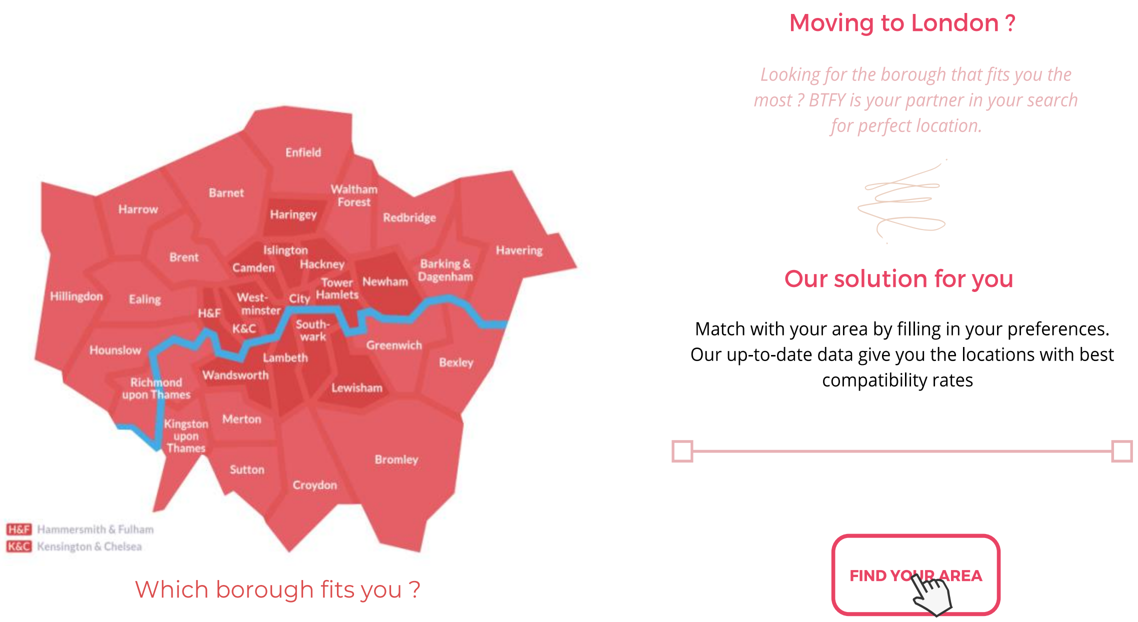
(Randstad, 2020)

Appendix 5 – Website mock-up

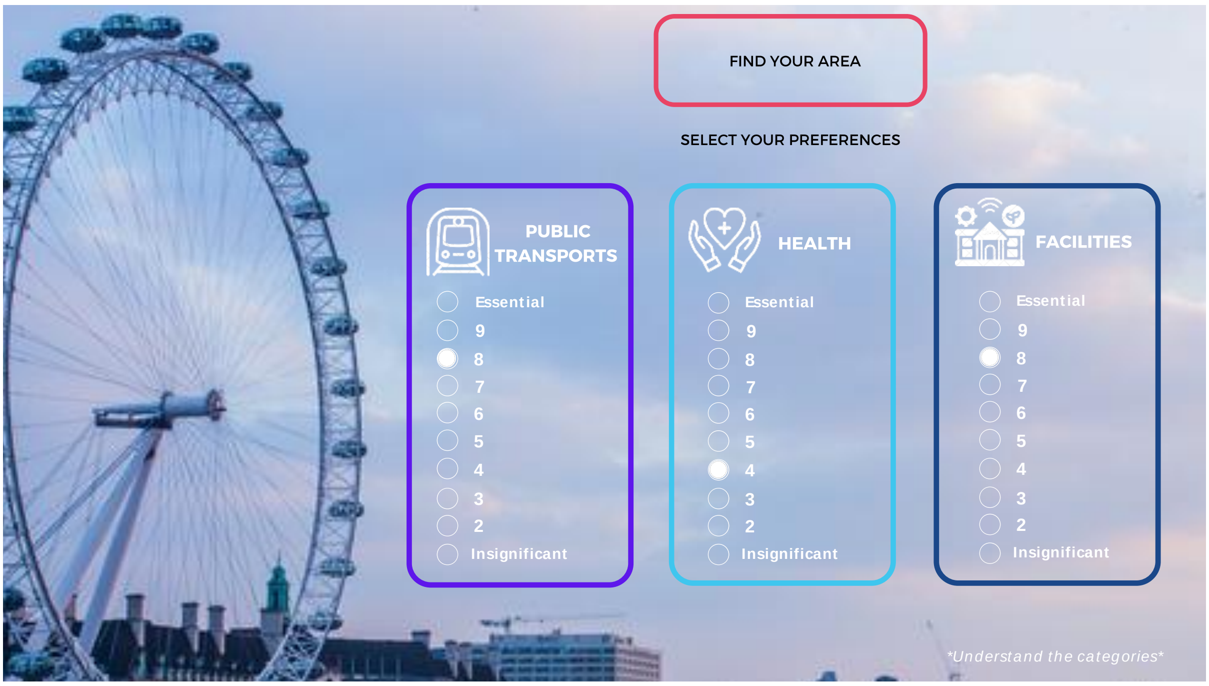
Welcome page:



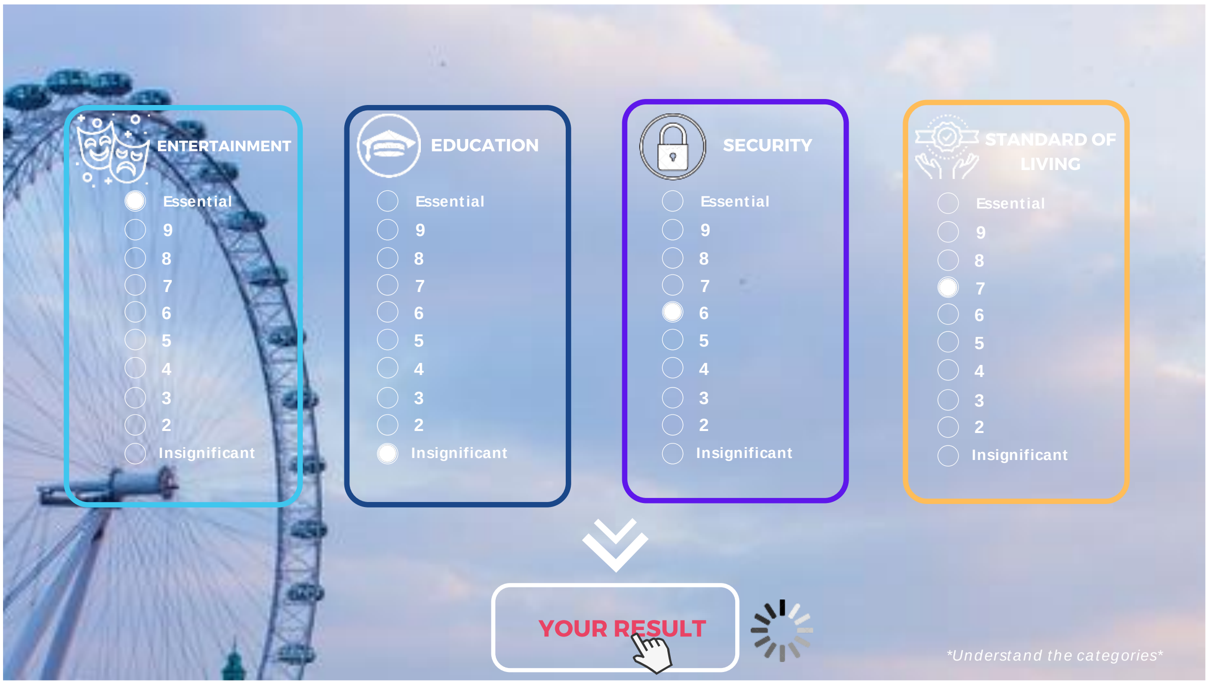
Product explanation page:



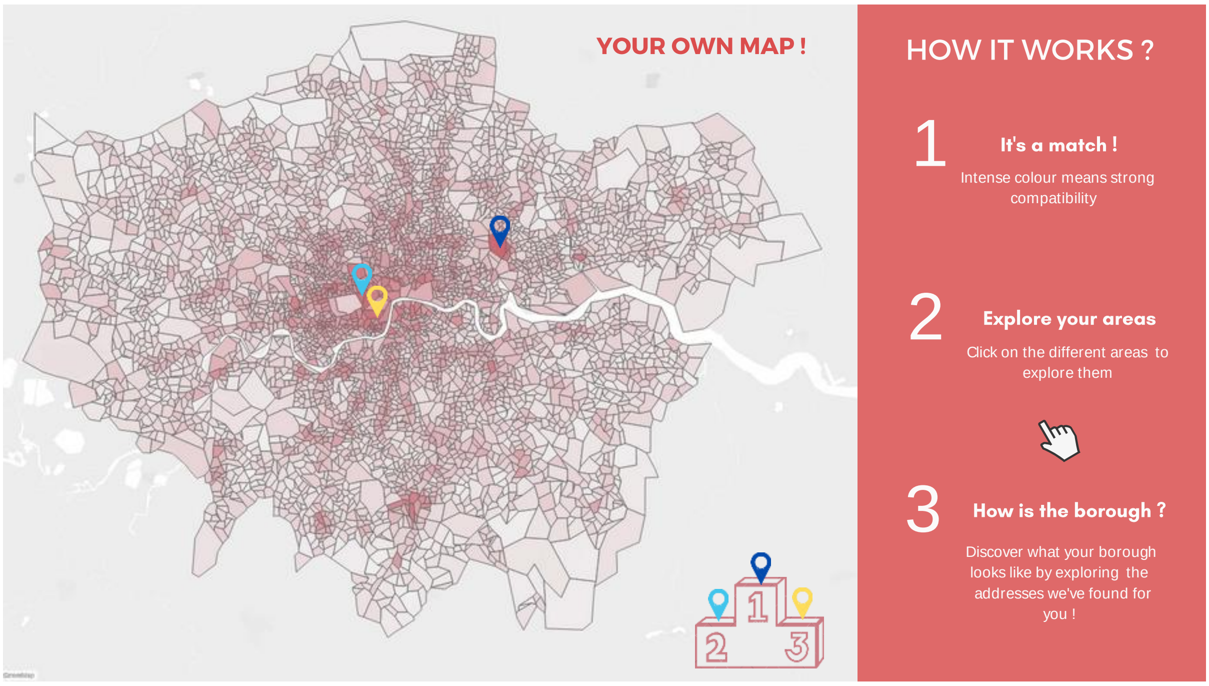
Individualised weightings page 1:



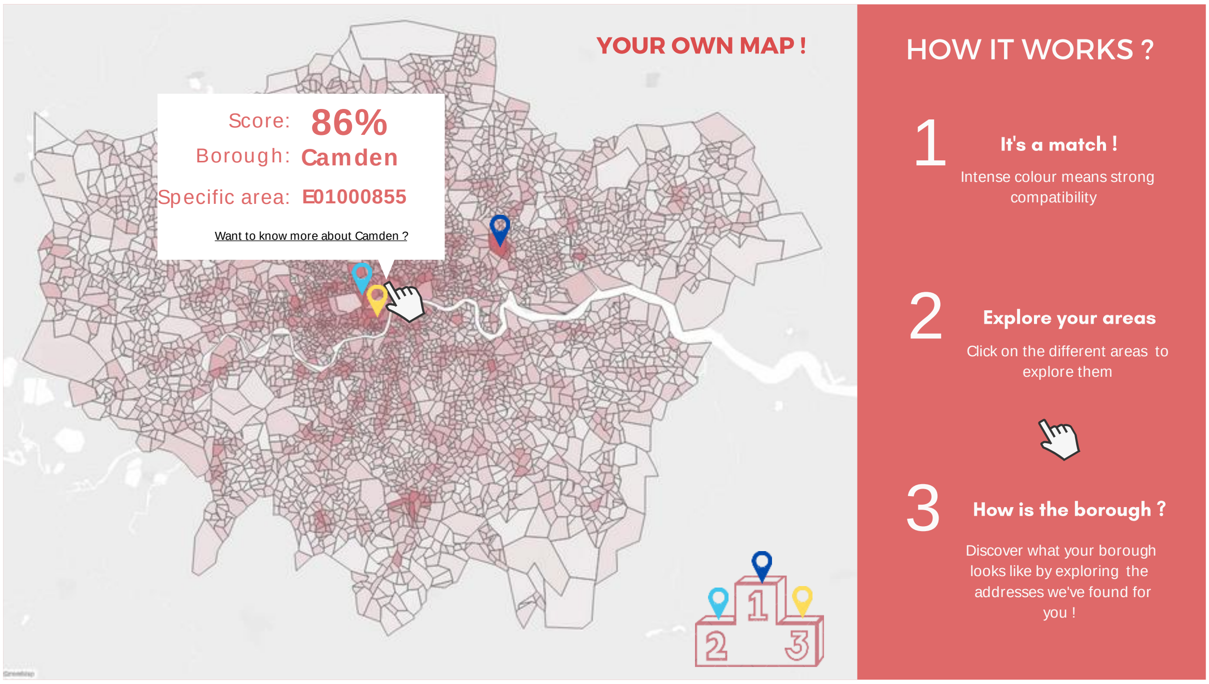
Individualised weightings page 2:



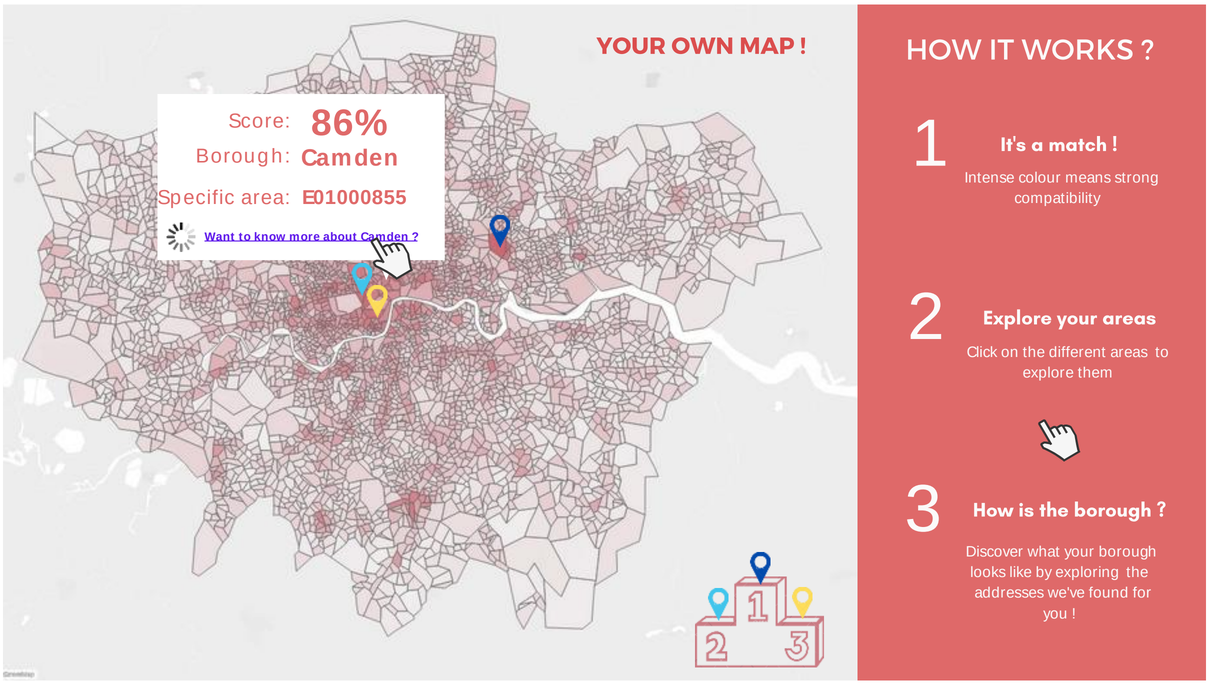
Your individualised map:



The individualised map is interactive:



From the individualised map, you can deep-dive into interesting areas for you:



Extra information for a specific area chosen from individualised map:

Step off the tube and enter the colourful world of Camden, undoubtedly one of the most vibrant boroughs in London. Original stomping ground to punks and goths, this grungy street food mecca is home to one of the best live music scenes in London. But while Camden is known for its heady atmosphere and sprawling markets, it has a softer side too. Meander away from the hustle and bustle of the main high street to discover broad, tree-lined streets filled with whitewashed Victorian houses and the kind of quiet atmosphere you could live in.

Restaurants :


Pappa Ciccia-Fulham High Street


Shaka Zulu


Mildreds Camden


Wagamama

Live music venues :

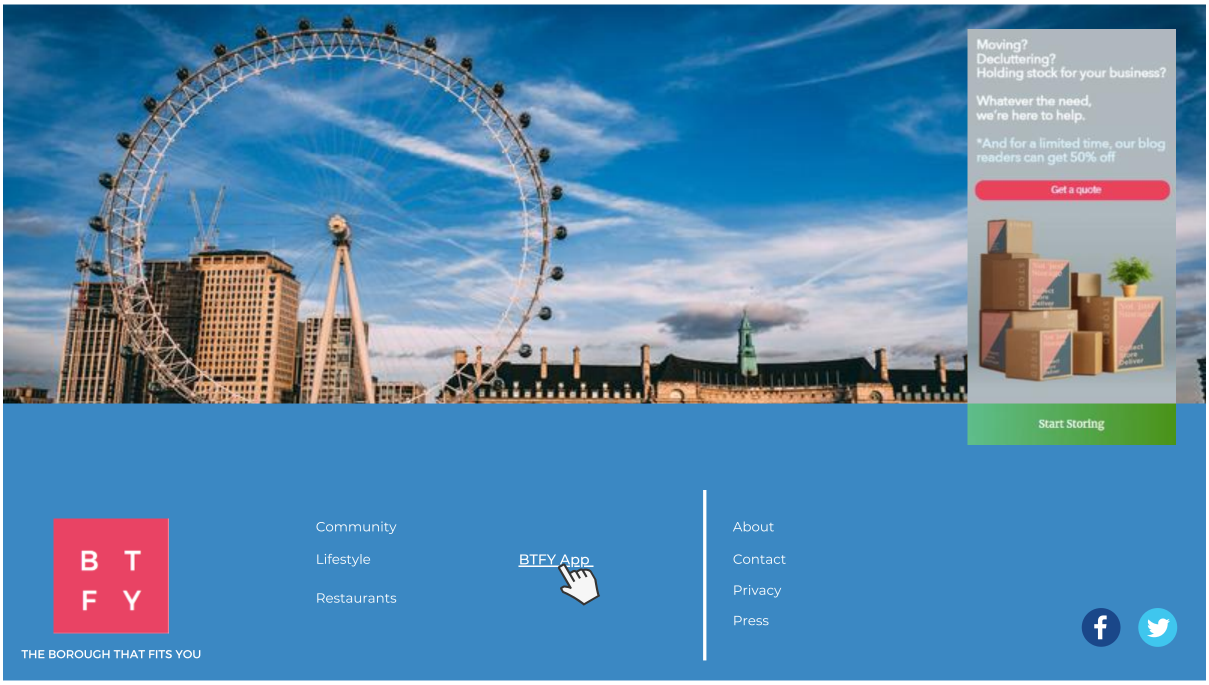

Jazz Cafe


KOKO

Want to know more about your borough and your future neighbours ?

App Store

We also have an app:



Appendix 6 – Area types

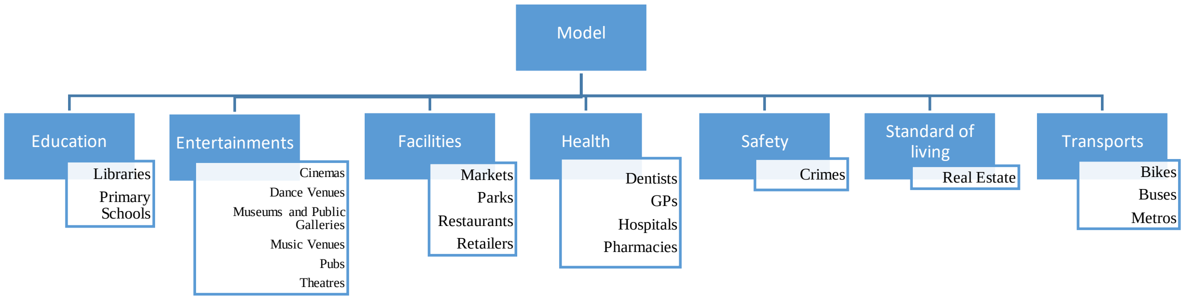
Full Name	Abbreviation	Description
Output Area	OA	Lowest geographical level at which census estimates are provided, with an average population of about 310 residents.
Lower Layer Super Output Area	LSOA	Geographical hierarchy level with an average population of 1500 people.
Middle Layer Super Output Area	MSOA	Geographical hierarchy level with an average population of about 7500 residents.

Appendix 7 – Overview of data sets used

Categories / Subcategories	Purpose	Content	Date	Size	Format
Education	-------------------	--	---------------	------------------- -	-----------
Libraries	Libraries in the UK	Description / Identification: ID/ Name/Type Location: Country/ County/ Locality/ Postcode/ Street/ Town/ Contact: Email/ Phone/ Website	12/2012	4,274 rows 14 columns 51,124 data	CSV
Primary Schools	Primary Schools in London	Description / Identification: Gender/ ID/ Name/ Phase/ Status/ Type/ URN Location: Address/ Easting/ LA/ LSOA/ Northing/ Postcode/ Town/ Ward/ Contact: Website	2016	3,890 rows 25 columns 96,250 data	CSV
Entertainments	-------------------	--	---------------	------------------- -	-----------
Cinemas	Cinemas in London	**Access:** Address/ Borough/ Latitude/ Longitude/ Ward/ Website Name	01/2019	80 rows 16 columns 971 data	CSV
Dance Performances	Dance Performances in London	**Access:** Address/ Borough/ Latitude/ Longitude/ Ward/ Website Name	12/2018	190 rows 16 columns 2,459 data	CSV
Museums and public galleries	Museums and Galleries in London	**Access:** Address/ Borough/ Latitude/ Longitude/ Ward/ Website Name	01/2019	164 rows 16 columns 2,461 data	CSV
Music venues	Music venues in London	**Access:** Address/ Borough/ Latitude/ Longitude/ Ward/ Website Name	03/2018 Updated immanent	798 rows 16 columns 10,770 data	CSV
Pubs	Pubs in London	**Access:** Address/ Borough/ Latitude/ Longitude/ Ward/ Website Name	01/2019 Updated immanent	4,099 rows 16 columns 52,386 data	CSV
Theatres	Museums and Galleries in London	**Access:** Address/ Borough/ Latitude/ Longitude/ Ward/ Website Name	01/2019	265 rows 16 columns 3,970 data	CSV
Facilities	-------------------	--	---------------	--------------------	-----------
Markets	Fresh food markets in London	**Description / Identification:** ID/ Name/Type **Location** Opening hours/ Postcode **Contact:** Website	28/12/2018	24 rows 7 columns 136 data	XLS
Parks	Parks in London	**Location:** Output areas/ Local authority districts/ Lower layer super output areas	30/11/2018	25,054 rows 8 columns	CSV

		Diverse info: Green space/ Rent price/ Safety/ Schools/ Travel		200,432 data	
Restaurants	Restaurants in London	**Description / Identification:** IDs Name/ Local Authority info (name/ email...)/ Scores of satisfactions/ / Type **Location:** Postcode/ Latitude/ Longitude **Contact:** Website	27/07/2020	2397 rows 37 columns 74,964 data	XML
Retailers	Retailers in the UK	**Description / Identification:** Fascia/ ID/ Name/ PQI/ Retailer/ Size **Location:** Address/ Easting/ Latitude/ Longitude/ Northing/ Postcode/ Suburb/ Town		16,688 rows 18 columns 268,703 data	CSV
Health	-------------------	--	---------------	--------------------	-----------
Dentists	Dentists in the UK	**Description / Identification:** Code/ ID/ Name/ Organization status/ PIMS Management?/ Subtypes/ Type **Location:** Address/ City/County/ Latitude/ Longitude/ Postcode **Contact:** Email/ Fax/ Phone/ Website	05/2020	7,378 rows 21 columns 133,173 data	CSV
GPs	GPs in the UK	**Description / Identification:** Code/ EPS Enabled?/ ID/ Name/ Organization status/ PIMS Management?/ Sector/ Subtypes/ Type **Location:** Address/ City/ County/ Latitude/ Longitude/ Postcode **Contact:** Email/ Fax/ Phone/ Website	05/2020	9,074 rows 24 Columns 176,818 data	CSV
Hospitals	Hospitals in the UK	**Description / Identification:** Code/ ID/ Name/ Organization status/ PIMS Management?/ Sector/ Subtypes/ Type **Location:** Address/ City/ County/ Latitude/ Longitude/ Postcode **Contact:** Email/ Fax/ Phone/ Website	05/2020	1,207 rows 20 Columns 22,271 data	CSV
Pharmacies	Pharmacies in the UK	**Description / Identification:** Code/ ID/ Name/ Organization status/ PIMS Management?/ Sector/ Subtypes/ Type **Location:** Address/ City/ County/ Latitude/ Longitude/ Postcode **Contact:** Email/ Fax/ Phone/ Website	05/2020	11,026 rows 24 columns 211,010 data	CSV
Safety	Crimes in London	**Description:** Category / Date **Location:** / Borough/ LSOA Code	From 11/2018 to 10/2020	103,846 rows 28 columns 2,907,688 data	CSV
Standard of living	Cost of real estate in London	**Description:** Rent for 1/2/3/4 room(s) **Location:** Output Areas	11/2018	25,054 rows 5 columns 125,270 data	CSV
Transports	-------------------	--	---------------	--------------------	-----------
Bike Stations	Bike Stations in London	**Description / Identification:** Capacity/ ID/ Name **Location:** Longitude/ Latitude	2018	774 rows 5 columns 3870 data	CSV
Bus and Metro Stations	Bus and Metro Stations in the UK	**Description / Identification:** Code/Creation/ ID/ Modification/ Name/ Status / Type **Location:** Easting and Northing	2010	104,114 rows 13 columns 1,233,553 data	CSV

Appendix 8 – Overview of categories and subcategories

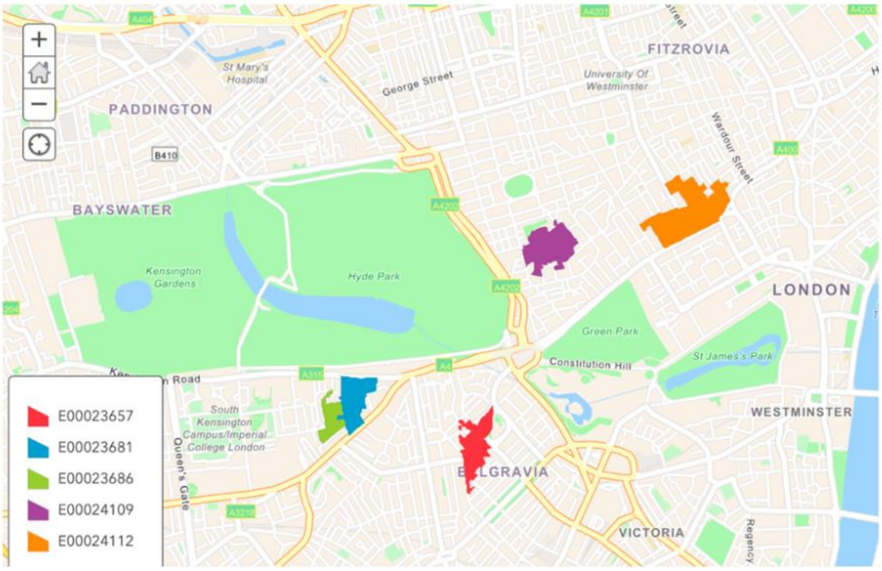


Appendix 9 – Why does the service make sense in London?

To investigate whether our service makes sense in London, we decided to compare 5 very similar Output Areas. Output Areas compared were:

- E00023657
- E00023681
- E00023686
- E00024109
- E00024112

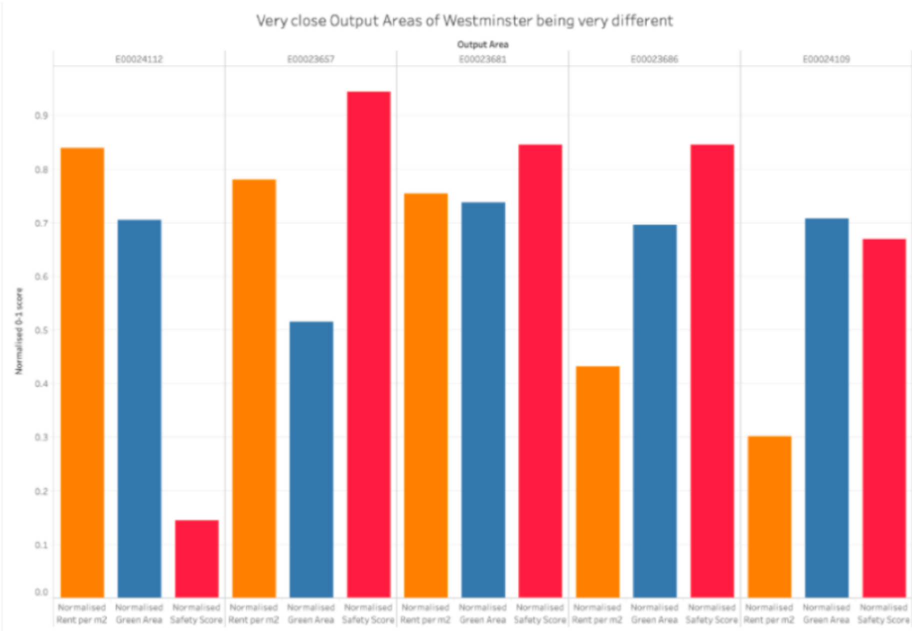
Visualisation 3 - The 5 Output Areas



(ArcGIS, 2020)

For the Output Areas in visualisation 3, the standardised score on 3 different datasets were calculated.

Visualisation 4 - The 5 Output Areas compared



Normalised scores from 0-1 for rent per m2 (orange), green area (blue), and safety (red).

(Greater London Authority, 2018b)

As can be seen from visualisation 4, E00024112, E00023657 and E00023681, all more or less have the same rent price per m2, though the areas differ widely in how much green area and how safe the areas are. Similarly, E00023686 and E00024109 score very high on both safety and green area, though they are way less expensive. Finally, particularly interesting is E00023681 and E00023686 who border each other, though E00023686 is almost half as expensive while keeping the same amount of green area and outperforming E00023681 significantly on safety. Thus, it is evident that there are significant differences between individual Output Areas, and informing about these differences on Output Level or close to for people migrating to London can add significant value for them.

Appendix 10 – Distance to target area

We also wanted to give our customers the ability to choose a target location (in the form of a postcode) as we believe it's an important factor when moving in. The goal of this was to find an ideal area for our user close to where they need to be (for example as a university student).

We did this in the following way:

1. Get OA code of the postcode our user submitted using the London Postcode Directory provided by the Office For National Statistics 2020d
2. Calculate centre points in coordinates for all OAs using geopandas, from a map file containing the borders of each OA (ArcGIS, 2020)
3. Calculate the distance between the target location and each OA

This resulted in a map such as figure 4.5.

Figure 4.5, OA distance to School of Management building in Canary Wharf.

Combining this with the scores for each area allowed us to find the best area close to where our user needs to be.

Bibliography

ArcGIS, 2020. *Customisable map for UK.* [Online]
Available at: https://www.arcgis.com/home/webmap/viewer.html?panel=gallery&suggestField=true&url=https%3A%2F%2Fons-inspire.esriuk.com%2Farcgis%2Frest%2Fservices%2FAdministrative_Boundaries%2FWards_December_2015_Boundaries%2FMapServer%2F0
[Accessed 26 November 2020].

Au, E., 2019. *London Bike Sharing System.* [Online]
Available at: https://www.kaggle.com/edenau/london-bike-sharing-system-data?select=stations.csv
[Accessed 19 November 2020].

Barrat, 2018. *How to choose the right area of London to live in.* [Online]
Available at: https://www.barratthomes.co.uk/new-homes/london/advice-and-inspiration/how-to-choose-right-area-of-london/
[Accessed 19 November 2020].

Blinder, S., 2018. *Immigration by Category: Workers, Students, Family Members, Asylum Applicants.* [Online]
Available at: https://migrationobservatory.ox.ac.uk/resources/briefings/immigration-by-category-workers-students-family-members-asylum-applicants/
[Accessed 22 November 2020].
Collections Trust, 2012. *UK Public Libraries contacts.* [Online]
Available at: https://data.gov.uk/dataset/de509172-9bf3-4e2b-bea3-518dc70217cb/uk-public-libraries-contacts
[Accessed 21 November 2020].
Department for Transport, 2014. *National Public Transport Access Nodes (NaPTAN).* [Online]
Available at: https://data.gov.uk/dataset/ff93ffc1-6656-47d8-9155-85ea0b8f2251/national-public-transport-access-nodes-naptan
[Accessed 18 November 2020].
Dourou, S., 2019. *London continues to be most miserable place in the UK Read more: https://metro.co.uk/2019/10/23/london-continues-to-be-most-miserable-place-in-the-uk-10971619/?ito=cbshare Twitter: https://twitter.com/MetroUK | Facebook: https://www.facebook.com/MetroUK/.* [Online]
Available at: https://metro.co.uk/2019/10/23/london-continues-to-be-most-miserable-place-in-the-uk-10971619/
[Accessed 23 November 2020].
Food Standards Agency, 2018. *UK food hygiene rating data (London) - Food Standards Agency.* [Online]
Available at: https://data.gov.uk/dataset/e945d1f5-72fe-4b08-8b58-8e68d83441b5/uk-food-hygiene-rating-data-london-food-standards-agency
[Accessed 22 November 2020].
Gambling Commission, 2020. *Licensed gambling premises.* [Online]
Available at: https://data.gov.uk/dataset/8dc9bef0-ad46-497b-9a74-30a6ce9d2f98/licensed-gambling-premises
[Accessed 23 November 2020].
Geolytix, 2020. *Retail Places.* [Online]
Available at: https://www.geolytix.co.uk/#!geodata
[Accessed 21 November 2020].
Greater London Authority, 2014. *London Borough Profiles and Atlas.* [Online]
Available at: https://data.london.gov.uk/dataset/london-borough-profiles
[Accessed 20 November 2020].
Greater London Authority, 2018a. *Number of fresh produce food markets in London.* [Online]
Available at: https://data.gov.uk/dataset/cb31e5e2-77d9-4583-8e0e-16dc4e3270a7/number-of-fresh-produce-food-markets-in-london
[Accessed 22 November 2020].
Greater London Authority, 2018b. *MyLondon data.* [Online]
Available at: https://data.gov.uk/dataset/65449de8-fd8a-41a0-b2db-13879b0a48f2/mylondon
[Accessed 19 November 2020].
Greater London Authority, 2018c. *Cultural Infrastructure Map.* [Online]
Available at: https://data.london.gov.uk/dataset/cultural-infrastructure-map
[Accessed 21 November 2020].
Greater London Authority, 2018d. *London Schools Atlas.* [Online]
Available at: https://data.gov.uk/dataset/6b776872-c786-4960-af1d-dab521aa4ab0/london-schools-atlas
[Accessed 21 November 2020].
Kinleigh Folkard & Hayward, 2020. *Identifying location and requirements.* [Online]
Available at: https://www.kfh.co.uk/resources/buyers/buying-process/identifying-location-and-

requirements
[Accessed 20 November 2020].
Lee, K.-Y., 2020. *The Effect of Residential Environmental Satisfaction on Quality of Life and the Moderating Effect of Housing Type: The Case of Gyeonggi, Korea.* [Online]
Available at: https://www.ajpor.org/article/13055-the-effect-of-residential-environmental-satisfaction-on-quality-of-life-and-the-moderating-effect-of-housing-type-the-case-of-gyeonggi-korea
[Accessed 24 November 2020].
Metropolitan Police Service, 2020. *Recorded Crime: Geographic Breakdown.* [Online]
Available at: https://data.london.gov.uk/dataset/recorded_crime_summary
[Accessed 22 November 2020].
NHS, 2020. *NHS website datasets.* [Online]
Available at: https://www.nhs.uk/about-us/nhs-website-datasets/
[Accessed 20 November 2020].
Office for National Statistics, 2020a. *Local area migration indicators UK.* [Online]
Available at:
https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/migrationwithintheuk/datasets/localareamigrationindicatorsunitedkingdom
[Accessed 18 November 2020].
Office for National Statistics, 2020b. *Standard Area Measurements (2019) for Administrative Areas in the United Kingdom.* [Online]
Available at: https://geoportal.statistics.gov.uk/datasets/standard-area-measurements-2019-for-administrative-areas-in-the-united-kingdom
[Accessed 18 November 2020].
Office for National Statistics, 2020c. *Median house prices by lower layer super output area: HPSSA dataset 46.* [Online]
Available at:
https://www.ons.gov.uk/peoplepopulationandcommunity/housing/datasets/medianpricepaidbylowerlayersuperoutputareahpssadataset46
[Accessed 19 November 2020].
Office for National Statistics, 2020d. *Postcode Directory for London.* [Online]
Available at: https://data.london.gov.uk/dataset/postcode-directory-for-london?fbclid=IwAR2aCPiPLYvh9jhI4qMo7TXljZONUee6ApH9ah8uQft9qLQg6c0tvZd8FL8
[Accessed 19 November 2020].
Quilty, D., 2019. *Where Should I Live? – 14 Factors When Deciding the Best Place to Live.* [Online]
Available at: https://www.moneycrashers.com/where-should-i-live-decide-best-places/
[Accessed 20 November 2020].
Randstad, 2020. *Which London borough are you best suited to?.* [Online]
Available at: https://www2.randstad.co.uk/best-london-borough/
[Accessed 17 November 2020].
Transport for London, 2019. *Travel in London.* [Online]
Available at: http://content.tfl.gov.uk/travel-in-london-report-12.pdf
[Accessed 21 November 2020].