

House Prices & Neighbourhood Venues

Data Analysis of Birmingham, UK

IBM Data Science Coursera Capstone Project Week 4

Introduction

Birmingham is the second-largest city of the United Kingdom (UK) with roughly 1.1 million inhabitants within the city area and 3.8 million inhabitants within the metropolitan area. This also makes Birmingham the 17th largest city in the world. Located approximately 100 miles from central London, Birmingham, as one of the United Kingdom's major cities, is considered to be the social, cultural, financial, and commercial center of both the East and West Midlands.

There are some facts about the city

1. **Population:** 1,111,307 (2017 estimate), 3.8 million in the Greater Metropolitan area, most ethnically diverse city in the UK.
2. **Official Language:** English (*de facto*) with 108 languages being spoken in schools, including Urdu, Punjabi, Bengali, and many others.
3. **GDP Per Capita:** \$46,296.98 USD
4. **Youthful Population:** 45.7% of Birmingham's population is under 30
5. **Currency:** Pound Sterling or Great Britain Pound (£, GBP). As of early 2017, \$1 USD = £0.80, €1 EUR = £0.85, \$1 CAD = £0.61, \$1 AUD = £0.62

Compared with other major cities in the world, houses and flats and flat prices in Birmingham are reasonable. From a new person's point of view, if they want to move to this city they want to invest in such places where the housing prices are low and the facilities (shops, restaurants, parks, hotels, etc.) and social venues are nearby.

Keeping these things into consideration it is very difficult for an individual to find such a place in such a large city and gather this much information. To solve these problems, we can create a map and information chart where the real estate index is placed on the city and each area is clustered according to the venue density.

Target audience

This analysis is usually to help new-comers to the city and existing people who want to buy a home. This analysis will give them a better understanding of which area they can live based on their affordability. The analysis may help them to decide to buy their homes.

Hence there are two major categories of the target audience -

New Comers

People from across the country and world coming to the city and planning to buy a home in a neighborhood of their choice.

Existing city people

Planning to invest in buying their new home based on choices of the venue they want in their neighborhood.

Data Collection

There are various sources are used to collect the data for this exercise -

Source	Description	Method	Results
https://en.wikipedia.org/wiki/B_postcode_area	First, we need all the neighborhoods of the city. Neighborhoods can be achieved in various ways. The one is chosen in this exercise to get the details from Wikipedia. Wikipedia has a page for all the areas (Coverages) for Birmingham. The B postcode area , also known as the Birmingham postcode area , is a group of postcode districts.	HTML DOM parsing	Neighbourhood names with postal code being a key
https://propertydata.co.uk/cities/Birmingham	After getting the first step done, we need the average price of houses in these neighborhoods. We use propertydata.co.uk for the data collection. This service gives us average house price in the city (based on postal code)	HTML DOM parsing	Average price of the houses with postal code for neighbourhoods
https://opencagedata.com/	We need to find the latitude and longitude for the neighborhoods of the postal code areas of the city. We use this service to find geocodes.	REST API Call / JSON	Lat / Lng for the neighbourhood areas
https://developer.foursquare.com	Foursquare API to find the venues of the neighbourhoods	REST API Call / JSON	Venue details of the neighbourhood

Methodology & Data Preprocessing

Data is collected from Wikipedia for the list of area codes and details. City area details are collected after parsing HTML DOM via Python.

Postcode district ↕	Post town ↕	Coverage ↕	Local authority area ↕
B1	BIRMINGHAM	Birmingham City Centre, Broad Street (east)	Birmingham
B2	BIRMINGHAM	Birmingham City Centre, New Street	Birmingham
B3	BIRMINGHAM	Birmingham City Centre, Newhall Street	Birmingham
B4	BIRMINGHAM	Birmingham City Centre, Corporation Street (north)	Birmingham
B5	BIRMINGHAM	Digbeth, Highgate, Lee Bank	Birmingham
B6	BIRMINGHAM	Aston, Witton	Birmingham
B7	BIRMINGHAM	Nechells	Birmingham
B8	BIRMINGHAM	Washwood Heath, Ward End, Saltley	Birmingham
B9	BIRMINGHAM	Bordesley Green, Bordesley	Birmingham
B10	BIRMINGHAM	Small Heath	Birmingham
B11	BIRMINGHAM	Sparkhill, Sparkbrook, Tyseley	Birmingham
B12	BIRMINGHAM	Balsall Heath, Sparkbrook, Highgate	Birmingham
B13	BIRMINGHAM	Moseley, Billesley	Birmingham
B14	BIRMINGHAM	Kings Heath, Yardley Wood, Druids Heath, Highter's Heath, Warstock	Birmingham
B15	BIRMINGHAM	Edgbaston,	Birmingham
B16	BIRMINGHAM	Ladywood	Birmingham
B17	BIRMINGHAM	Harborne, Edgbaston	Birmingham

HTML DOM of this page consists of a table and few hyperlinks. So, after collecting HTML DOM table , collecting data is converted into dataframe. And new dataframe looked like -

	Town	Coverage	Area	Postal
1	BIRMINGHAM	Birmingham City Centre, Broad Street (east)	Birmingham	B1
2	BIRMINGHAM	Birmingham City Centre, New Street	Birmingham	B2
3	BIRMINGHAM	Birmingham City Centre, Newhall Street	Birmingham	B3
4	BIRMINGHAM	Birmingham City Centre, Corporation Street (no...	Birmingham	B4
5	BIRMINGHAM	Digbeth, Highgate, Lee Bank	Birmingham	B5

After collecting the data of Bermingham areas (coverage), we need to collect the average price of these areas. For that price data from is used from PropertyData (<https://propertydata.co.uk/cities/Birmingham>) . Data is read from the URL PropertyData. HTML DOM of PropertyData is parsed and property price data is collected like -

	Postal	AvgYield	AvgPrice	PerSqftPrice	FiveYearChange	Links
1	B1	4.8%	£218,216	£317	+19%	Explore data
2	B3	4.3%	£239,971	£286	+23%	Explore data
3	B5	5.2%	£211,124	£258	+19%	Explore data
4	B9	5.3%	£139,358	£134	+23%	Explore data
5	B11	4.0%	£171,299	£143	+22%	Explore data

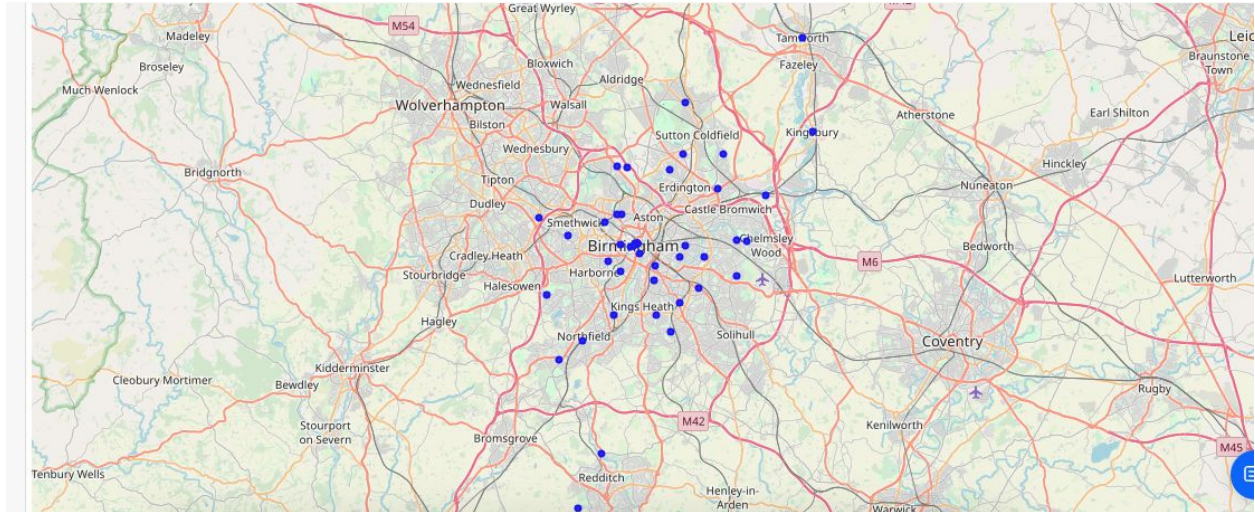
A data cleaning needs to be done here. From the table above we only require Postal and AvgPrice. So a new data set is created merging these two -

	Town	Coverage	Area	Postal	AvgPrice
0	BIRMINGHAM	Birmingham City Centre, Broad Street (east)	Birmingham	B1	218216
1	BIRMINGHAM	Birmingham City Centre, Newhall Street	Birmingham	B3	239971
2	BIRMINGHAM	Digbeth, Highgate, Lee Bank	Birmingham	B5	211124
3	BIRMINGHAM	Bordesley Green, Bordesley	Birmingham	B9	139358
4	BIRMINGHAM	Sparkhill, Sparkbrook, Tyseley	Birmingham	B11	171299

Another challenge is to find the geolocation (latitude and longitude). For that Opencagedata service is used.

	Town	Coverage	Area	Postal	AvgPrice	lat	lng
0	BIRMINGHAM	Birmingham City Centre, Broad Street (east)	Birmingham	B1	218216	52.478955	-1.905994
1	BIRMINGHAM	Birmingham City Centre, Newhall Street	Birmingham	B3	239971	52.480859	-1.900922
2	BIRMINGHAM	Digbeth, Highgate, Lee Bank	Birmingham	B5	211124	52.886993	-2.886859
3	BIRMINGHAM	Bordesley Green, Bordesley	Birmingham	B9	139358	52.479293	-1.838810
4	BIRMINGHAM	Sparkhill, Sparkbrook, Tyseley	Birmingham	B11	171299	52.463967	-1.875828

At this point we have all the details required like Town, Coverage, Area and GeoDetails (Lat, Lng). To represent these geolocations and area on a map we use Folium. Folium is a python package to show the map. Folium is used to visualize geographic details of Birmingham and its coverage. All the latitude and longitude are passed to the Folium map method to achieve following visual -



We use Foursquare API to explore the coverage and segment them. For our project, we designed the limit as 100 venue and the radius 1400 meter for each borough from their given latitude and longitude information. Here is the head of the list Venues name, category, latitude and longitude information from Foursquare API.

	name	categories	lat	lng
0	Birmingham Museum & Art Gallery	History Museum	52.480385	-1.903905
1	Birmingham Town Hall	Concert Hall	52.479509	-1.903560
2	Hyatt Regency Birmingham	Hotel	52.478096	-1.909324
3	Adams	Restaurant	52.479797	-1.901482
4	Symphony Hall	Concert Hall	52.478874	-1.910229

The same process is repeated for all the venues and we get following results -

	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Neighborhood						
Acocks Green	10	10	10	10	10	10
Balsall Heath, Sparkbrook, Highgate	15	15	15	15	15	15
Birmingham City Centre, Broad Street (east)	75	75	75	75	75	75
Birmingham City Centre, Newhall Street	79	79	79	79	79	79
Boldmere, New Oscott, Wylde Green	13	13	13	13	13	13
Bordesley Green, Bordesley	4	4	4	4	4	4
Bournville, Cotteridge, Stirchley	12	12	12	12	12	12
Bromsgrove (east)	79	79	79	79	79	79
Castle Bromwich, Smith's Wood, Bromford, Hodge Hill	81	81	81	81	81	81
Chelmsley Wood, Marston Green, Kingshurst, Fordbridge	5	5	5	5	5	5
Edgbaston,	5	5	5	5	5	5
Erdington, Stockland Green, Short Heath, Perry Common	5	5	5	5	5	5

Finally by using the Foursquare API in conjunction with the created datasets, a table of most common visited venues in the city neighborhoods is generated. We used K-means clustering for further data analysis to find the right neighbourhood.

Results

We concluded (based on our priorities for the neighborhoods) following neighbourhoods are good places to settle. These neighbourhoods are having decent average price for the households and having categories of the venues that we wanted to test for like (Grocery store, bar and train station). Here are the results -

	Coverage	Grocery Store	Bar	Train Station	Cluster Labels	Town	Area	Postal	AvgPrice	lat	lng
23	Moseley, Billesley	0.25	0.0	0.00	2	BIRMINGHAM	Birmingham	B13	211475	52.427302	-1.874673
14	Hall Green	0.20	0.0	0.00	2	BIRMINGHAM	Birmingham	B28	266405	52.436995	-1.845597
40	Tamworth, Fazeley, Kingsbury, Polesworth	0.25	0.0	0.00	2	TAMWORTH	Tamworth, North Warwickshire	B78	264294	52.564476	-1.682967
44	Water Orton, Coleshill, Nether Whitacre	0.25	0.0	0.25	2	BIRMINGHAM	North Warwickshire	B46	313723	52.517139	-1.739945
24	Northfield, Longbridge, West Heath	0.50	0.0	0.25	2	BIRMINGHAM	Birmingham	B31	197025	52.408231	-1.965252

Conclusion

As people are turning to big cities to start a business or work. For this reason, people can easily interpret where to live with all facilities and cheaply. Not only for investors but also city managers can manage the city more regularly by using similar data analysis types or platforms.