|  |  |
| --- | --- |
| **Coursera**  **IBM Data Science Professional Certificate**  **Department of Data Science, Machine Learning and Data Analysis**  Coursera's Mission, Vision, and Commitment to Our Community | Coursera    **Applied Capstone Project**  **Exploring Ethnic Diversity and Venue-Based city-to-city Similarity Measures** | |
| 1. **Quanunga**   **BEng Mechanical Engineering**  **January 2021** | |
| Signed:  Arkaanahmedkhan Quanunga | Date:  01.01.2021 |

Contents

[Abstract 3](#_Toc60516831)

[Keywords 3](#_Toc60516832)

[Introduction 3](#_Toc60516833)

[DATA 4](#_Toc60516834)

[Methodology: 5](#_Toc60516835)

[Data Visualization 5](#_Toc60516836)

[Appendix 1: Postcode Area Data Table Format 6](#_Toc60516837)

[Appendix B: Ethnic Group Distributions 9](#_Toc60516838)

[Figure 4- EC postcode area (Postcode Project, 2012) 3](file:///C:\Users\arkaa\Desktop\Capstone_Project.docx#_Toc60516343)

[Figure 1- B postcode area (Office of National Statistics, 2012) 3](file:///C:\Users\arkaa\Desktop\Capstone_Project.docx#_Toc60516344)

[Figure 2- Ethnic Group Distribution over London Boroughs ( Census, 2011) 4](file:///C:\Users\arkaa\Desktop\Capstone_Project.docx#_Toc60516345)

[Figure 3- Ethnic Group Distribution in Birmingham (Census, 2011) 4](file:///C:\Users\arkaa\Desktop\Capstone_Project.docx#_Toc60516346)

[Table 1 Postcode Area District data of East Central London ( Office of National Statistics, 2012 ) 5](#_Toc60516136)

[Table 2: Postcode district data (B) for Birmingham(Postcode Project,2012) 6](#_Toc60516137)

[Table 3- Ethnic group population, London 2001 & 2011. ( Census, 2011) 9](#_Toc60516138)

[Table 4 Racial make- up of London boroughs 2011 ( Census, 2011) 9](#_Toc60516139)

[Table 5- Ethnic Group Distribution Birmingham ( Census, 2011) 10](#_Toc60516140)

# Abstract

# Keywords

Venues, Location-Analysis, City Similarities, Geolocator API, Geocoder API, Location-Based Services (LBS), Neighborhood Comparison, Data Visualization,

# Introduction

Imagine two hypothetical cities, Birmingham, and East Central London, that are precisely the same in nearly every way, having the same venues- the same number of restaurants, parks, and gyms. Suppose further that they only differ in the venues' spatial arrangement. Birmingham's venues are distributed uniformly throughout the city, and the East Central London venues benefit the residents' movement due to their locations. East Central London, being in the capital of England- London, and it's substantial economic and cultural developments aiding many of the venue's popularity and growing competitions of new platforms. In contrast, Birmingham got results in infrastructure, including the Grand Central transformation in 2015 and the upcoming tram services project in 2021 (Temple, 2015). One might ask about the differences and similarities between these two homogeneous Neighborhoods in many ways. This report analyses preliminary concepts for representing and comparing the two cities. (̧tiuc-Pietro, et al., 2013)

Growth of Location-Based Services(LBS) during the 2010s due to selective availability of Global Positioning System(GPS) on smartphones lead to a cosmic increase in data (Huang, et al., 2018). The data collected from such services include applying optimising shipping routes in real-time, increasing revenues and profit margins for existing and new opening restaurants, entertainment locale, and Hotels, and details of peak locations during the day. Using such data with relevant machine learning algorithms can undoubtedly be a powerful tool to help automate day to day life. Examples of such data include Large corporations such as Apple, Samsung, and Facebook that are willing to expand their offices in numeral cities. However, they must consider the Neighborhood's skillset, the venues nearby, and the particular location's population diversity.

Furthermore, a new graduate student is job hunting and is willing to relocate if he lands the job. He can use such Location-Based data to narrow the applications to the cities where his skills are valued. Additionally, a person renting/buying a new home in an unknown town wants to know the Neighborhood's similarities and differences in that city or even compare the location to his district. Finally, political parties and civil engineers can investigate inside city data to pick the optimum location for expanding residential life and provide residents with everyday needs.

Comparing two different cities often gives rise to the problem of normalizing data, leading to inaccurate conclusions. This report only analyzes a part of London- East Central London and Birmingham. Doing such an investigation means that it is comparing Neighborhoods that split by grids/ clusters. This removes the redundancy of spatial distribution and normalizing data between cities. Our hypothetical assumption says that the venues in Birmingham are spatially distributed, and the outlets of East Central London benefit the everyday movement of residents in that area.

# DATA

The data used in this report include the postcode areas in East Central (EC) London and the postcode areas of Birmingham (B). They are both obtained from the Office of National Statistics, 2012: EC postcode area & B postcode area.

Figures 1 and 4 show examples of postcode area data obtained. Appendix 1 shows the entire list of data used.

Figure - EC postcode area (Postcode Project, 2012)

Figure - B postcode area (Office of National Statistics, 2012)

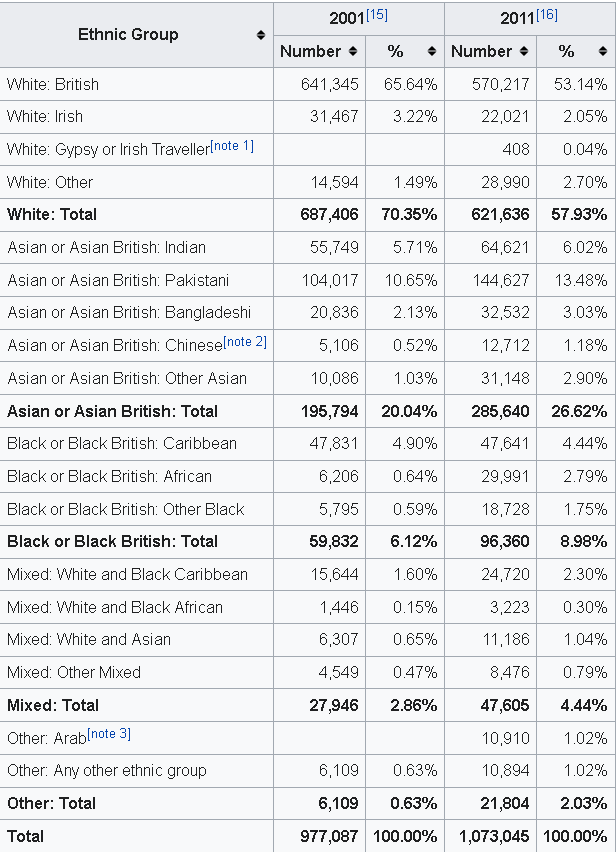
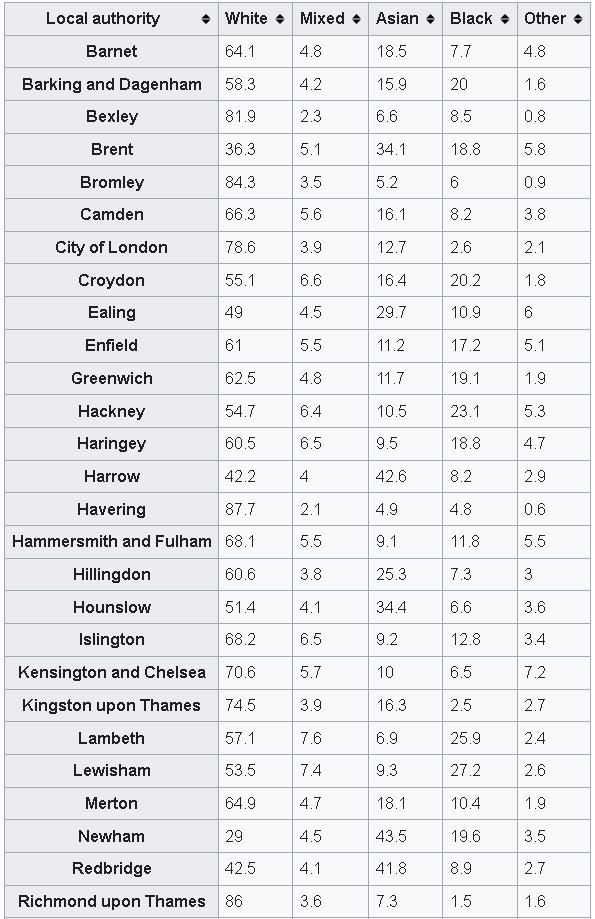
Census 2011 provides the data for the diversity in population over Birmingham and East Central London shown in Figures 2 and 3. Appendix B shows the full list of population-based data obtained from Census 2011.

Figure - Ethnic Group Distribution over London Boroughs ( Census, 2011)

Figure - Ethnic Group Distribution in Birmingham (Census, 2011)

# Methodology:

Firstly, the postcode district areas for Birmingham City and East Central London run through a Location-based service (LBS) known as Geocoder API and Geolocator API from Google to convert the districts into longitude-latitude on the world map.

The Folium library in python uses the longitude-latitude values to plot the Neighborhoods on the location-specific map. The visualization helps with data analysis by interacting with raw data of numbers and words. It helps speed the process of data analysis as the human brain tends to process visual information in a brisk than any written communication presented. Furthermore, it draws the necessary relationships between the data of the two cities.

Following the data visualization of the postcode district for both cities, Foursquare API is used to obtain the top 100 venues in a radius of 500m of each Neighborhood. This report discards any venue that either falls outside the city boundary or could not be mapped in the same Neighborhood. This ensures the data representation from both cities is applied to the same data set points when venues are considered. The venues' category obtained merges with the Neighborhood to create a binary data frame of 0s and 1s. For example, if the Neighborhood has an 'Italian Restaurant,' it has a result of 1, and if the district does not have an Antique shop within the 500m radius and top 100 venues, then the output is 0. Subsequently, this data frame statistically examines the mean frequency of the venues categories in the district. As the number of 1s for 'Italian Restaurant' increases, the mean frequency of 'Italian Restaurant' in that Neighborhood increases.

The top 100 venues in the district's name and categories of the venues allow each Neighborhood's analyses to occur. This report uses the K-means algorithm to form a number of clusters in the postcode district areas with the venue's categories' mean frequency in each Neighborhood. From the top 100 venues, the top 10 types of venues merge with the Neighborhood to create a new data frame. This data frame is fed into the K-means algorithm to form the clusters. The different groups produced are compared to each other and labelled in the analyses. Finally, inside city cluster analyses occur, and the comparison is shifted to city-to-city examination.

# Data Visualization

# Appendix 1: Postcode Area Data Table Format

Table Postcode Area District data of East Central London ( Office of National Statistics, 2012 )

|  |  |  |  |
| --- | --- | --- | --- |
| Postcode district | Post town | Coverage | Local authority area(s) |
| EC1A | LONDON | St Bartholomew's Hospital | City of London, Islington |
| EC1M | LONDON | Clerkenwell, Farringdon | Islington, Camden, City of London |
| EC1N | LONDON | Hatton Garden | Camden, City of London |
| EC1P | LONDON |  | non-geographic |
| EC1R | LONDON | Finsbury, Finsbury Estate (west) | Islington, Camden |
| EC1V | LONDON | Finsbury (east), Moorfields Eye Hospital | Islington, Hackney |
| EC1Y | LONDON | St Luke's, Bunhill Fields | Islington, City of London |
| EC2A | LONDON | Shoreditch | Islington, Hackney, City of London |
| EC2M | LONDON | Broadgate, Liverpool Street | City of London, Tower Hamlets |
| EC2N | LONDON | Old Broad Street, Tower 42 | City of London |
| EC2P | LONDON |  | non-geographic |
| EC2R | LONDON | Bank of England | City of London |
| EC2V | LONDON | Guildhall | City of London |
| EC2Y | LONDON | Barbican | City of London |
| EC3A | LONDON | St Mary Axe, Aldgate | City of London |
| EC3M | LONDON | Lloyd's of London, Fenchurch Street | City of London |
| EC3N | LONDON | Tower Hill, Tower of London | Tower Hamlets, City of London |
| EC3P | LONDON |  | non-geographic |
| EC3R | LONDON | Monument, Billingsgate | City of London |
| EC3V | LONDON | Cornhill, Gracechurch Street, Lombard Street | City of London |
| EC4A | LONDON | Fetter Lane | City of London, Westminster |
| EC4M | LONDON | St Paul's | City of London |
| EC4N | LONDON | Mansion House | City of London |
| EC4P | LONDON |  | non-geographic |
| EC4R | LONDON | Cannon Street | City of London |
| EC4V | LONDON | Blackfriars | City of London |
| EC4Y | LONDON | Temple | City of London, Westminster |
| EC50 | LONDON |  | non-geographic |

Table : Postcode district data (B) for Birmingham(Postcode Project,2012)

|  |  |  |  |
| --- | --- | --- | --- |
| Postcode district | Post town | Coverage | Local authority area(s) |
| 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, Birchfield, Birmingham , Witton | Birmingham |
| B7 | BIRMINGHAM | Nechells, Vauxhall | 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 |
| B18 | BIRMINGHAM | Winson Green, Hockley | Birmingham |
| B19 | BIRMINGHAM | Lozells, Newtown, Birchfield | Birmingham |
| B20 | BIRMINGHAM | Handsworth Wood, Handsworth, Birchfield | Birmingham |
| B21 | BIRMINGHAM | Handsworth | Birmingham |
| B23 | BIRMINGHAM | Erdington, Stockland Green, Short Heath, Perry Common | Birmingham |
| B24 | BIRMINGHAM | Erdington, Wylde Green, Tyburn | Birmingham |
| B25 | BIRMINGHAM | Yardley | Birmingham |
| B26 | BIRMINGHAM | Sheldon, Yardley | Birmingham |
| B27 | BIRMINGHAM | Acocks Green | Birmingham |
| B28 | BIRMINGHAM | Hall Green | Birmingham |
| B29 | BIRMINGHAM | Selly Oak, Bournbrook, Selly Park, Weoley Castle, California (part of) | Birmingham |
| B30 | BIRMINGHAM | Bournville, Cotteridge, Stirchley | Birmingham |
| B31 | BIRMINGHAM | Northfield, Longbridge, West Heath | Birmingham |
| B32 | BIRMINGHAM | Woodgate, Bartley Green, Quinton, California (part of) | Birmingham |
| B33 | BIRMINGHAM | Kitts Green, Stechford | Birmingham |
| B34 | BIRMINGHAM | Shard End, Buckland End | Birmingham |
| B35 | BIRMINGHAM | Castle Vale | Birmingham |
| B36 | BIRMINGHAM | Castle Bromwich, Smith's Wood, Bromford, Hodge Hill | Birmingham, Solihull |
| B37 | BIRMINGHAM | Chelmsley Wood, Marston Green, Kingshurst, Fordbridge | Solihull |
| B38 | BIRMINGHAM | Kings Norton, West Heath | Birmingham |
| B40 | BIRMINGHAM | National Exhibition Centre, Birmingham International Airport, Marston Green, Bickenhill | Solihull |
| B42 | BIRMINGHAM | Perry Barr, Great Barr, Hamstead | Birmingham |
| B43 | BIRMINGHAM | Great Barr, Hamstead | Birmingham, Sandwell, Walsall |
| B44 | BIRMINGHAM | Perry Barr, Kingstanding, Great Barr | Birmingham |
| B45 | BIRMINGHAM | Rednal, New Frankley, Rubery, Cofton Hackett, Barnt Green, Lickey, Longbridge, Northfield (part) | Birmingham, Bromsgrove |
| B46 | BIRMINGHAM | Water Orton, Coleshill, Nether Whitacre | North Warwickshire |
| B47 | BIRMINGHAM | Hollywood, Wythall | Bromsgrove |
| B48 | BIRMINGHAM | Alvechurch | Bromsgrove |
| B49 | ALCESTER | Alcester | Stratford-on-Avon |
| B50 | ALCESTER | Bidford-on-Avon | Stratford-on-Avon |
| B60 | BROMSGROVE | Bromsgrove (east) | Bromsgrove |
| B61 | BROMSGROVE | Bromsgrove (west) | Bromsgrove |
| B62 | HALESOWEN | Halesowen (east), Romsley, Hunnington, Quinton (part) | Dudley, Bromsgrove, Birmingham |
| B63 | HALESOWEN | Halesowen (west and town centre), Hayley Green, Hasbury | Dudley, Bromsgrove |
| B64 | CRADLEY HEATH | Cradley Heath, Old Hill | Sandwell |
| B65 | ROWLEY REGIS | Rowley Regis, Blackheath | Birmingham, Sandwell |
| B66 | SMETHWICK | Smethwick (east and north), Bearwood (east) | Sandwell |
| B67 | SMETHWICK | Smethwick (west), Bearwood (west), Londonderry | Sandwell |
| B68 | OLDBURY | Langley, Brandhall, Quinton (part of) | Sandwell |
| B69 | OLDBURY | Oldbury, Tividale | Sandwell |
| B70 | WEST BROMWICH | West Bromwich (south, west and town centre) | Sandwell |
| B71 | WEST BROMWICH | West Bromwich (north) | Sandwell |
| B72 | SUTTON COLDFIELD | Sutton Coldfield town centre, Maney, Wylde Green, Erdington | Birmingham |
| B73 | SUTTON COLDFIELD | Boldmere, New Oscott, Wylde Green | Birmingham |
| B74 | SUTTON COLDFIELD | Four Oaks, Mere Green, Little Aston, Streetly | Birmingham, Lichfield, Walsall |
| B75 | SUTTON COLDFIELD | Sutton Trinity, Falcon Lodge, Rectory | Birmingham |
| B76 | SUTTON COLDFIELD | Walmley, Minworth | Birmingham |
| B77 | TAMWORTH | Tamworth, Wilnecote, Amington | Tamworth, North Warwickshire |
| B78 | TAMWORTH | Tamworth, Fazeley, Kingsbury, Polesworth | Tamworth, North Warwickshire |
| B79 | TAMWORTH | Tamworth, Warton | Tamworth, North Warwickshire |
| B80 | STUDLEY | Studley | Stratford-on-Avon |
| B90 | SOLIHULL | Shirley, Solihull Lodge, Majors Green, Dickens Heath, Cheswick Green | Solihull, Bromsgrove |
| B91 | SOLIHULL | Solihull | Solihull |
| B92 | SOLIHULL | Olton, Elmdon, Bickenhill, Hampton-in-Arden | Solihull |
| B93 | SOLIHULL | Knowle, Dorridge, Bentley Heath, Tilehouse Green | Solihull |
| B94 | SOLIHULL | Hockley Heath, Earlswood | Solihull |
| B95 | HENLEY-IN-ARDEN | Henley-in-Arden | Stratford-on-Avon |
| B96 | REDDITCH | Feckenham | Redditch |
| B97 | REDDITCH | Redditch (west), Callow Hill, Webheath | Redditch |
| B98 | REDDITCH | Redditch (east), Beoley | Redditch |
| B99 | BIRMINGHAM | | non-geographic |

# Appendix B: Ethnic Group Distributions

Table - Ethnic group population, London 2001 & 2011. ( Census, 2011)

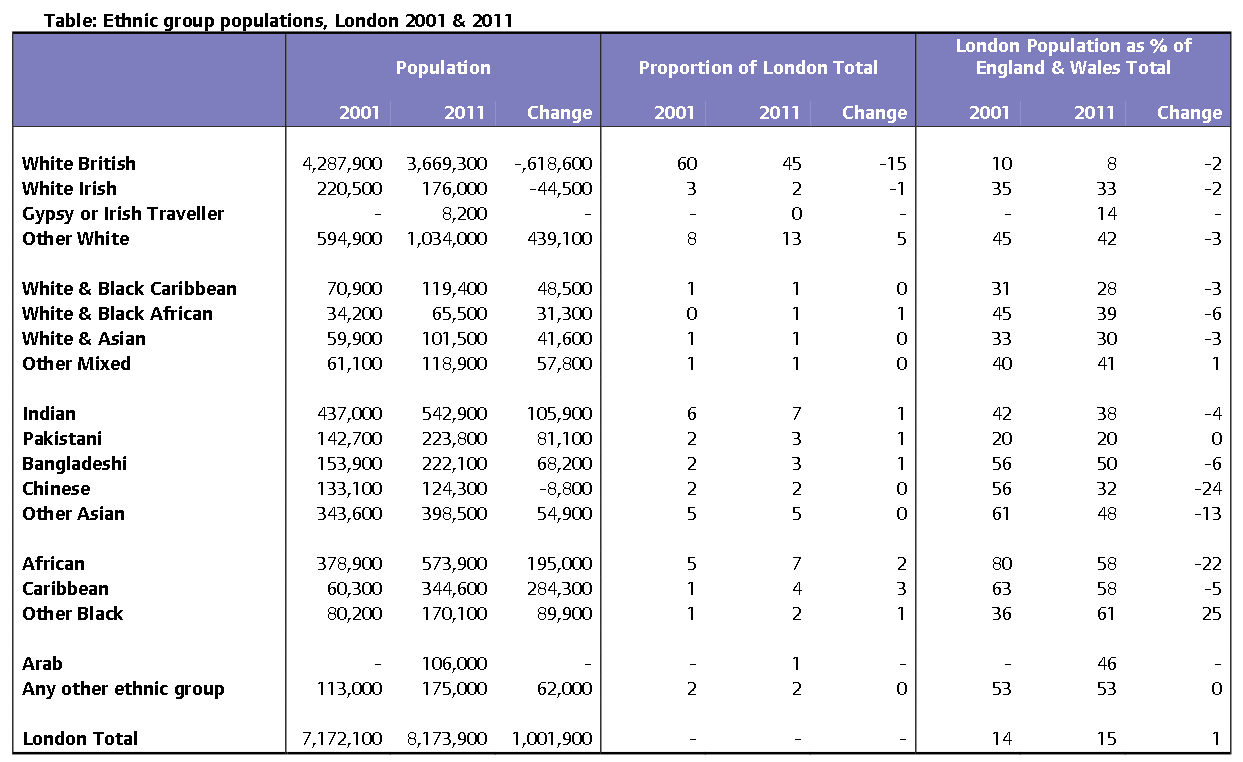


Table Racial make- up of London boroughs 2011 ( Census, 2011)

| Racial make-up of London boroughs (2011) | | | | | |
| --- | --- | --- | --- | --- | --- |
| **Local authority** | **White** | **Mixed** | **Asian** | **Black** | **Other** |
| **Barnet** | 64.1 | 4.8 | 18.5 | 7.7 | 4.8 |
| **Barking and Dagenham** | 58.3 | 4.2 | 15.9 | 20 | 1.6 |
| **Bexley** | 81.9 | 2.3 | 6.6 | 8.5 | 0.8 |
| **Brent** | 36.3 | 5.1 | 34.1 | 18.8 | 5.8 |
| **Bromley** | 84.3 | 3.5 | 5.2 | 6 | 0.9 |
| **Camden** | 66.3 | 5.6 | 16.1 | 8.2 | 3.8 |
| **City of London** | 78.6 | 3.9 | 12.7 | 2.6 | 2.1 |
| **Croydon** | 55.1 | 6.6 | 16.4 | 20.2 | 1.8 |
| **Ealing** | 49 | 4.5 | 29.7 | 10.9 | 6 |
| **Enfield** | 61 | 5.5 | 11.2 | 17.2 | 5.1 |
| **Greenwich** | 62.5 | 4.8 | 11.7 | 19.1 | 1.9 |
| **Hackney** | 54.7 | 6.4 | 10.5 | 23.1 | 5.3 |
| **Haringey** | 60.5 | 6.5 | 9.5 | 18.8 | 4.7 |
| **Harrow** | 42.2 | 4 | 42.6 | 8.2 | 2.9 |
| **Havering** | 87.7 | 2.1 | 4.9 | 4.8 | 0.6 |
| **Hammersmith and Fulham** | 68.1 | 5.5 | 9.1 | 11.8 | 5.5 |
| **Hillingdon** | 60.6 | 3.8 | 25.3 | 7.3 | 3 |
| **Hounslow** | 51.4 | 4.1 | 34.4 | 6.6 | 3.6 |
| **Islington** | 68.2 | 6.5 | 9.2 | 12.8 | 3.4 |
| **Kensington and Chelsea** | 70.6 | 5.7 | 10 | 6.5 | 7.2 |
| **Kingston upon Thames** | 74.5 | 3.9 | 16.3 | 2.5 | 2.7 |
| **Lambeth** | 57.1 | 7.6 | 6.9 | 25.9 | 2.4 |
| **Lewisham** | 53.5 | 7.4 | 9.3 | 27.2 | 2.6 |
| **Merton** | 64.9 | 4.7 | 18.1 | 10.4 | 1.9 |
| **Newham** | 29 | 4.5 | 43.5 | 19.6 | 3.5 |
| **Redbridge** | 42.5 | 4.1 | 41.8 | 8.9 | 2.7 |
| **Richmond upon Thames** | 86 | 3.6 | 7.3 | 1.5 | 1.6 |
| **Southwark** | 54.3 | 6.2 | 9.4 | 26.9 | 3.3 |
| **Sutton** | 78.6 | 3.8 | 11.6 | 4.8 | 1.3 |
| **Tower Hamlets** | 45.2 | 4.1 | 41.1 | 7.3 | 2.3 |
| **Waltham Forest** | 52.2 | 5.3 | 21.1 | 17.3 | 4.1 |
| **Wandsworth** | 71.4 | 5 | 10.9 | 10.7 | 2.1 |
| **Westminster** | 61.7 | 5.2 | 14.5 | 7.5 | 11.1 |

Table - Ethnic Group Distribution Birmingham ( Census, 2011)

|  |  |
| --- | --- |
| Birmingham Ethnic Group Distribution | |
| All categories: Ethnic group | 1,073,045 |
| White | 621,636 |
| English/Welsh/Scottish/Northern Irish/British | 570,217 |
| Irish | 22,021 |
| Gypsy or Irish Traveller | 408 |
| Other White | 28,990 |
| Mixed/multiple ethnic group | 47,605 |
| White and Black Caribbean | 24,720 |
| White and Black African | 3,223 |
| White and Asian | 11,186 |
| Other Mixed | 8,476 |
| Asian/Asian British | 285,640 |
| Indian | 64,621 |
| Pakistani | 144,627 |
| Bangladeshi | 32,532 |
| Chinese | 12,712 |
| Other Asian | 31,148 |
| Black/African/Caribbean/Black British | 96,360 |
| African | 29,991 |
| Caribbean | 47,641 |
| Other Black | 18,728 |
| Other ethnic group | 21,804 |
| Arab | 10,910 |
| Any other ethnic group | 10,894 |