# Venues data analysis – Birmingham vs London

Indian Restaurant Recommendations

# Introduction

- Birmingham and London are two of the largest UK cities that boasts multiculturalism and diversity. Therefore neighbourhoods will have some combination of different types of venues.
- Suppose that a London chain of Indian restaurants has established themselves in east London and would like to consider expanding towards another city. Then Birmingham would be an ideal candidate for the following reasons.
  - `2nd Largest City in population.
  - Great transport links and amenities. Other big cities would be much further north (e.g. Manchester)
  - Potential for lower operating costs
- Thus a study on segmenting neighbourhoods and trying to unearth similarities between neighbourhoods would aid a stakeholder into deciding their new location.

# Data collection & processing

- Data scraped postcode district tables from Wikipedia for both cities.
- Corresponding coordinates for the postcodes were obtained from Nominatim Geopy API.
- Non-geographical postcodes or anomalous values were removed.
- Cleaned data contained the following fields shown in the top-right table.
- This was then used to acquire venue data via Foursquare API for 51 postcodes.

| | Postcode district | Coverage | Latitude | Longitude |
|---|---|---|---|---|
| 0 | B1 | Birmingham City Centre, Broad Street (east) | 52.4775396 | -1.894053 |
| 1 | B2 | Birmingham City Centre, New Street | 52.4792602 | -1.8999756 |
| 2 | B3 | Birmingham City Centre, Newhall Street | 52.4832071 | -1.9054204 |
| 3 | B4 | Birmingham City Centre, Corporation Street (no... | 52.4775396 | -1.894053 |
| 4 | B5 | Digbeth, Highgate, Lee Bank | 52.4734488 | -1.8871192 |

# Venue Exploration

- In summary we requested nearby venues that were within 1km and capped at 50 venues for their given latitude & longitude pairs.
- The merged table below were returned by Foursquare.
- From the data we saw that most of the central London areas hit the 50 venues limit. Then the non-central areas had results ranging from 3-35 venues.
- Whereas for Birmingham we saw a similar result for central postcodes (such as B1) having hit the limit of 50 whilst other areas are relatively sparse.
- 198 unique venue categories within the venue dataset.
- The results heavily depends on latitude & longitude pairs used. More information about surrounding venues can be optimised if we had more coordinates or specific addresses.

| | Postcode | Postcode Latitude | Postcode Longitude | Venue | Venue Latitude | Venue Longitude | Venue Category |
|---|---|---|---|---|---|---|---|
| **115** | B3 | 52.4832071 | -1.9054204 | Birmingham Town Hall | 52.479509 | -1.903560 | Concert Hall |
| **587** | E10 | 51.8146817 | -0.356914287847605 | Cross Keys | 51.816618 | -0.356813 | Bar |
| **564** | E9 | 51.5434262 | -0.0277034 | Barge East | 51.542177 | -0.021046 | English Restaurant |
| **90** | B2 | 52.4792602 | -1.8999756 | The Lost & Found | 52.480122 | -1.900658 | Cocktail Bar |
| **945** | EC1Y | 51.52366865 | -0.0887493055891544 | Barbican Cinemas 2&3 | 51.521073 | -0.092979 | Indie Movie Theater |

# Top 5 venue category for each postcode

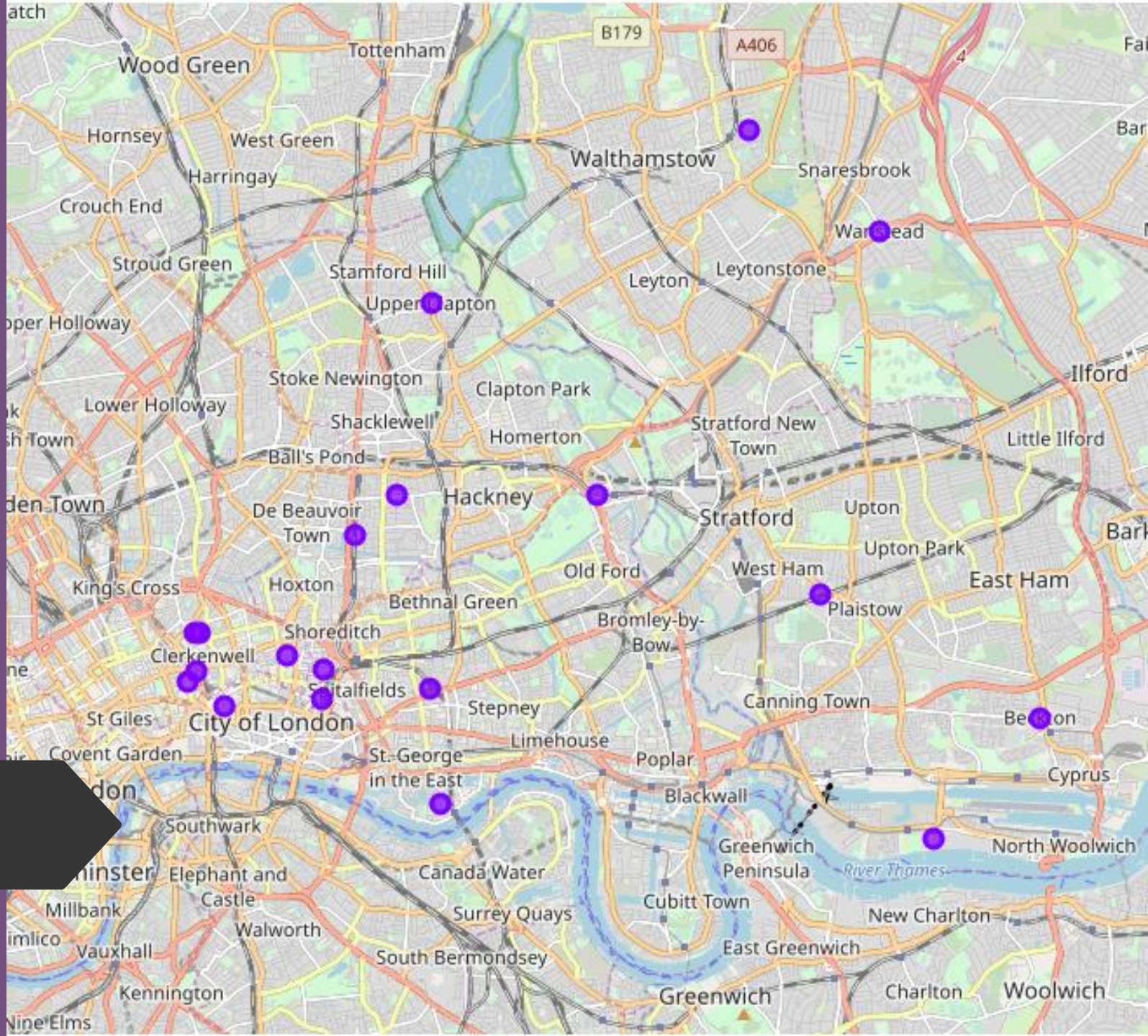
| | Postcode | 1st Most Common Venue | 2nd Most Common Venue | 3rd Most Common Venue | 4th Most Common Venue | 5th Most Common Venue |
|---|---|---|---|---|---|---|
| 0 | B1 | Burger Joint | Clothing Store | Bar | Portuguese Restaurant | Bookstore |
| 1 | B10 | Sandwich Place | Indian Restaurant | Park | Café | Turkish Restaurant |
| 2 | B11 | Pakistani Restaurant | Indian Restaurant | Convenience Store | Electronics Store | Food & Drink Shop |
| 3 | B12 | Indian Restaurant | Fast Food Restaurant | Grocery Store | Pool | Café |
| 4 | B13 | Pub | Moving Target | Golf Course | Yoga Studio | Electronics Store |
| 5 | B14 | Convenience Store | Gas Station | Grocery Store | Pharmacy | Yoga Studio |
| 6 | B15 | Tennis Court | Lake | Tennis Stadium | Golf Course | Yoga Studio |
| 7 | B16 | Italian Restaurant | Restaurant | Café | Bar | Theater |
| 8 | B17 | Cantonese Restaurant | Deli / Bodega | Coffee Shop | Gym / Fitness Center | Thai Restaurant |
| 9 | B18 | Light Rail Station | Vegetarian / Vegan Restaurant | Pizza Place | Yoga Studio | Electronics Store |

# Clustering Postcodes - Implementation

- We used sklearn's library to perform K means clustering between postcodes.
- The data of venues we have will need some pre-processing to be able to use k-means on categorical variables.
- After running K-means we examine the clusters and label each cluster based on their characteristic traits.
- Folium maps and summary statistics were used in conjunction to help visualise results.

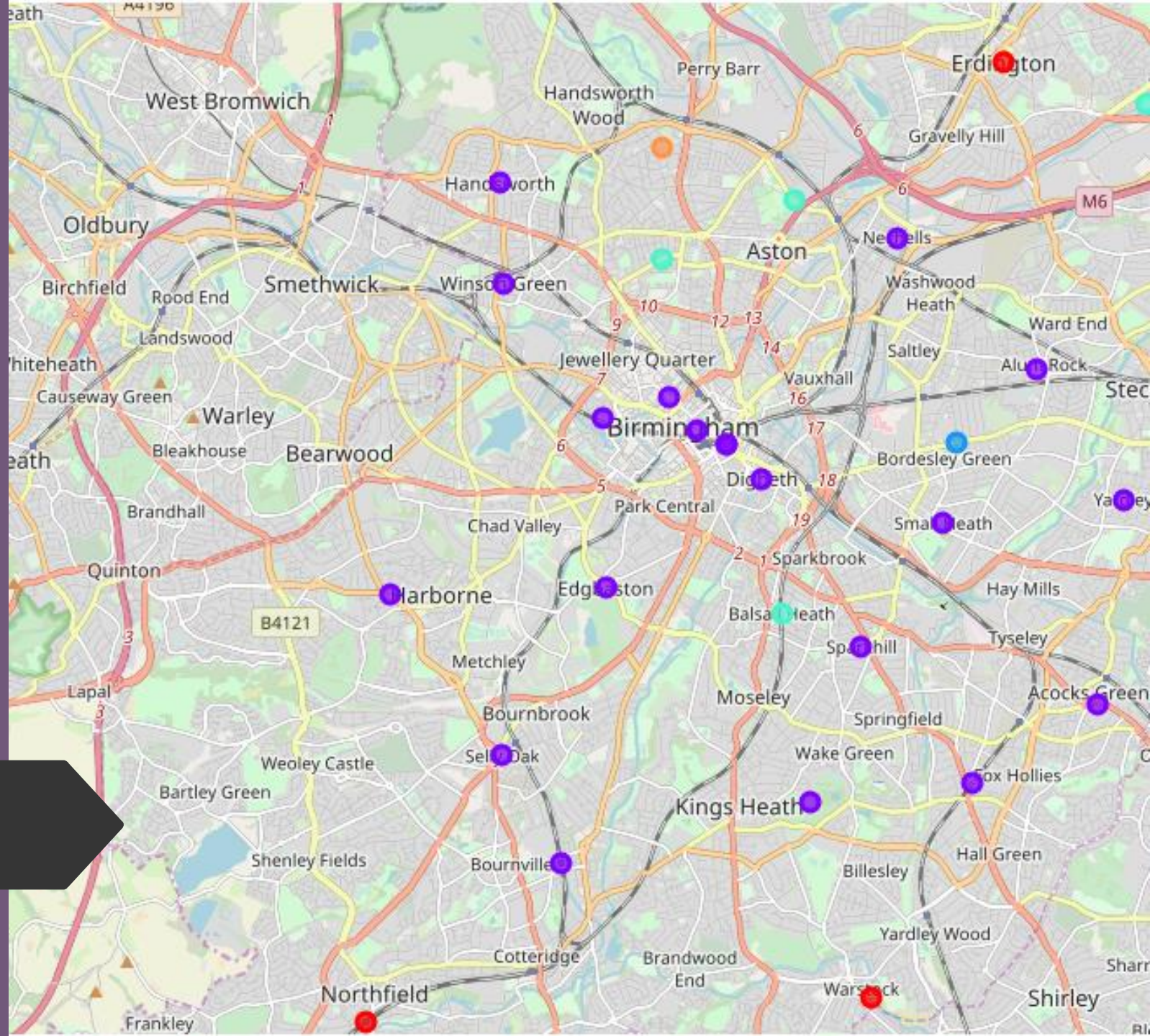
# Clustering Postcodes - Results

London



# Clustering Postcodes - Results

Birmingham



# Conclusion and Future Considerations

- Both cities in a nutshell may not be so similar. Clusters classified Birmingham town centre districts to be similar to the sample of London postcodes observed. This may suggest that population and traffic between both cities are not similar. Moreover Birmingham districts that were further from the centre were distinct from the London postcodes. I suspect underlying complexity with clustering postcodes and going forwards include few other important features for classification.
- Despite this we have some recommendations in the notebook from the results.
- Ideas to explore could include:
  - Attractiveness of location (proximity to tourist spots)
  - Socio-economic analysis
  - Property prices