London Boroughs Battle

May2020



Contents (week 1)

- 1. Business problem definition
- 2. Data and analytic approach
- 3. Methodology
 - 3.1 Data collection
 - 3.2 Data preparation
 - 3.3 Boroughs clustering (k-means)
- 4. Results
- 5. Conclusions

1. Business problem definition

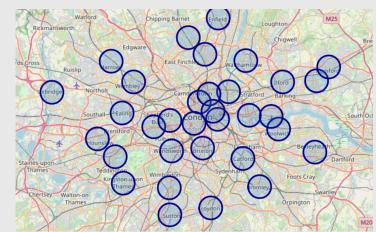
Our customer BestIndSpice is a start-up company which main activity is import of **Indian spices**. They intend to open a **brick and mortar shop** in London and ask us to evaluate the London boroughs and **identify the borough with highest potential for this business**.

They target British Indian households from London that use to cook at home.

London Boroughs







2. Data and Analytic Approach

- London is considered to be one of the world's most important global cities and has been called the world's most powerful, most desirable, most influential, most visited, most expensive,innovative,sustainable, most investment-friendly, and most-popular-forwork city" (https://en.wikipedia.org/wiki/London)
- London population was about 8.2 millions according to latest census from 2011.
 Around 6.6% of London's population is of Indian origin, largest ethnic minority group (more than 540 thousands in 2011).
- From administrative perspective, London has 32 boroughs.
- To identified the boroughs with the highest potential for Indian spices business we are going to use cluster analysis (k-means) on the following dimensions considered with impact:
 - number of British Indians living in London and the share in total population (source: https://data.london.gov.uk/dataset/ethnic-group-population-projections)
 - number of Indian restaurants (source: Fousquare API)
 - number of specialized spices stores (source: Fousquare API)

3. Methodology

3.1 Data collection

For this analysis we used the following data sources:

- List of London borough and their geographic coordinates (latitude and longitude)downloaded from https://en.wikipedia.org/wiki/List of London boroughs and loaded to https://labs.cognitiveclass.ai/ - > df_london_borough
- London population data forecast py barough from ID https://data.london.gov.uk/dataset/ethnic-group-population-projections. We extracted population forecast for 2025, for British Indians (-> df_Indian_pop) and All persons (-> df_all_pop)
- List of Indian restaurants in each borough from Foursquare API, using Categoryld filtering - > df_london_indian_rest
- List of spices stores (possible competitors) from Foursquare API, using Categoryld filtering - > df_london_spices_ID

3. Methodology

3.2 Data preparation

For boroughs clustering we needed a centralized dateframe with info at borough level and the following processing steps were done:

- Data from list of Indian restaurants/ spices stored were aggregated using groupby an count
- All sources were centralized in df_London_Cum_Info
- Missing values for Number_of_spices_stores were replaced by 0
- New feature was created: Number_of_Indian_rest_for10tho number of Indians restaurants on 10 thou British Indians

3. Methodology

3.3 Borough clustering (k-means)

For clustering we used **k-means algorithm**.

The **input variables were standardized** using preprocessing.StandardScaler.

Variables used for clustering:

- ➤ Indian_pop_fcast_2025
- > Share_of_Indian_pop_in_total
- > Number of Indian rest for 10 tho

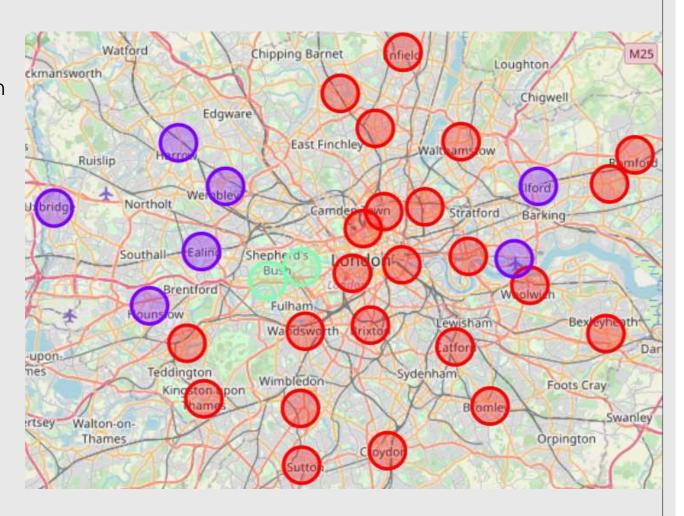
We used silhouette_score metric to determine number of clusters (k)

➤ Results shows highest silhouette_score for k=3

4. Results

We identified 3 clusters:

- Cluster0 with 23 boroughs with medium share of British Indian population and # of Indian restaurants/10thou pop
- Cluster1 with 7 boroughs with high share of British Indian population and low # of Indian restaurants / 10 thou pop
- Cluster2 with only 2 boroughs quite central, very low share of British Indian population and high number of Indian restaurant /10thou pop



5. Conclusion

- Cluster1 with 7 boroughs with high share of British Indian population and low # of Indian restaurants /10thou

pop

We recommend to focus on the 5 boroughs within Cluster1 that are situated in NW

