Analysing London Boroughs

IBM DATA SCIENCE CAPSTONE PROJECT

JAKE SMITH

Business Problem & Interest

Business Problem:

- How can specific boroughs within London be identified as ideal locations to set up a new craft beer bar?

Interest in the Problem:

- Anyone intending to enter the hospitality industry in London
- Those wanting to expand their existing hospitality business
- Tourists visiting London
- Those wanting to move to London to live

Data Sources

- Introductory background text from London **Wikipedia** page [1]
- Borough names, latitude and longitude co-ordinates web scraped using ***Beautiful Soup*** from **Wikipedia** London boroughs page [2]
- Commonly found venues based on location sourced from ***Foursquare API*** [3]
- Annual tourist trips to London csv file from **UK government data site** [4]
- **Geo json data** marking borough boundaries from **London Data Store** website [5]

Data Cleaning / Wrangling

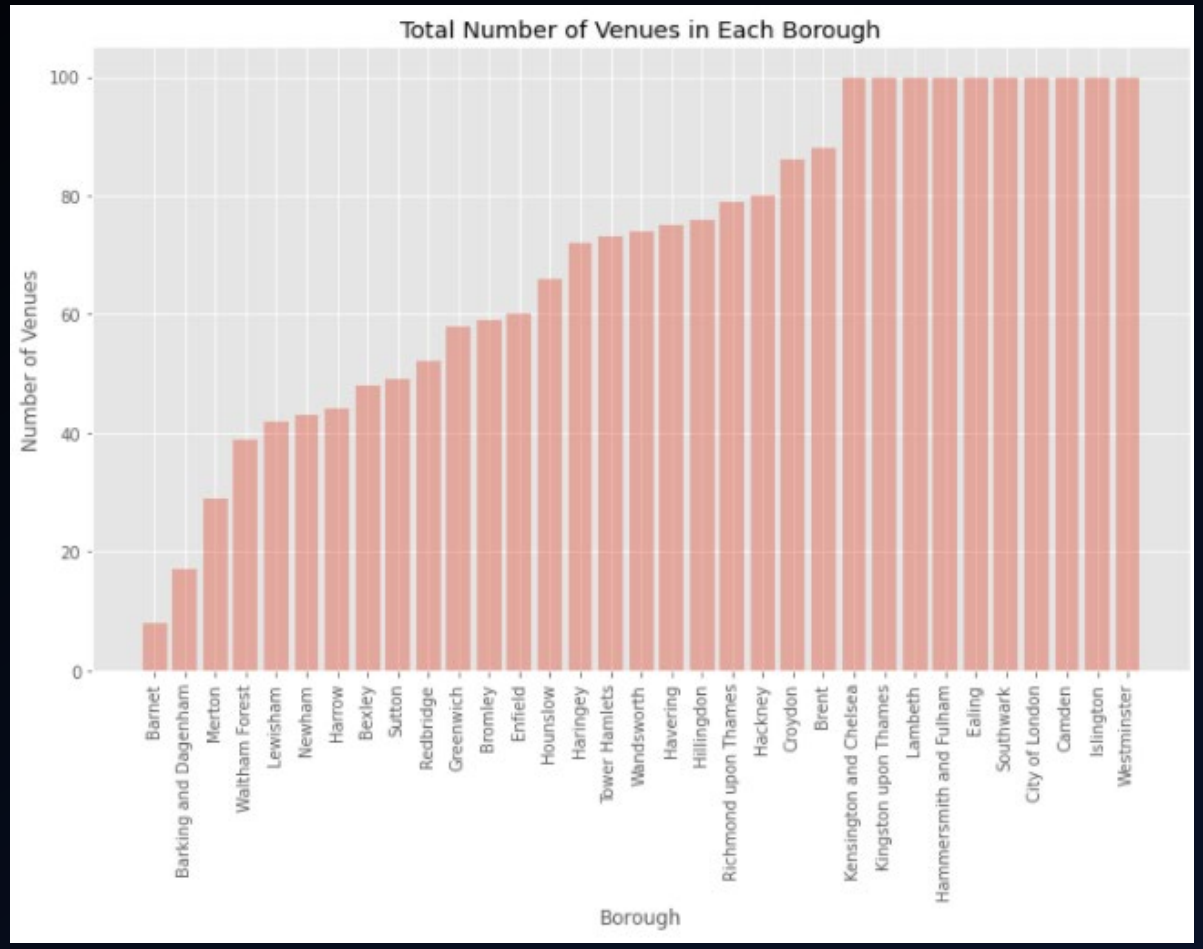
Cleaning involved:

- Removing hidden and special characters
- Replacing extraneous characters from web scraped HTML tables
- Recoding outlier values

Wrangling involved:

- Splitting columns
- Correcting data types to integer and float
- Joining and appending various data frames

Venue Frequency



- Venue types sourced from Foursquare API for each borough
- Radius of **1000** metres to borough lat and long co-ords
- Results limited to **100** per borough
- **264** unique venue categories sourced in total

Most Common Venues

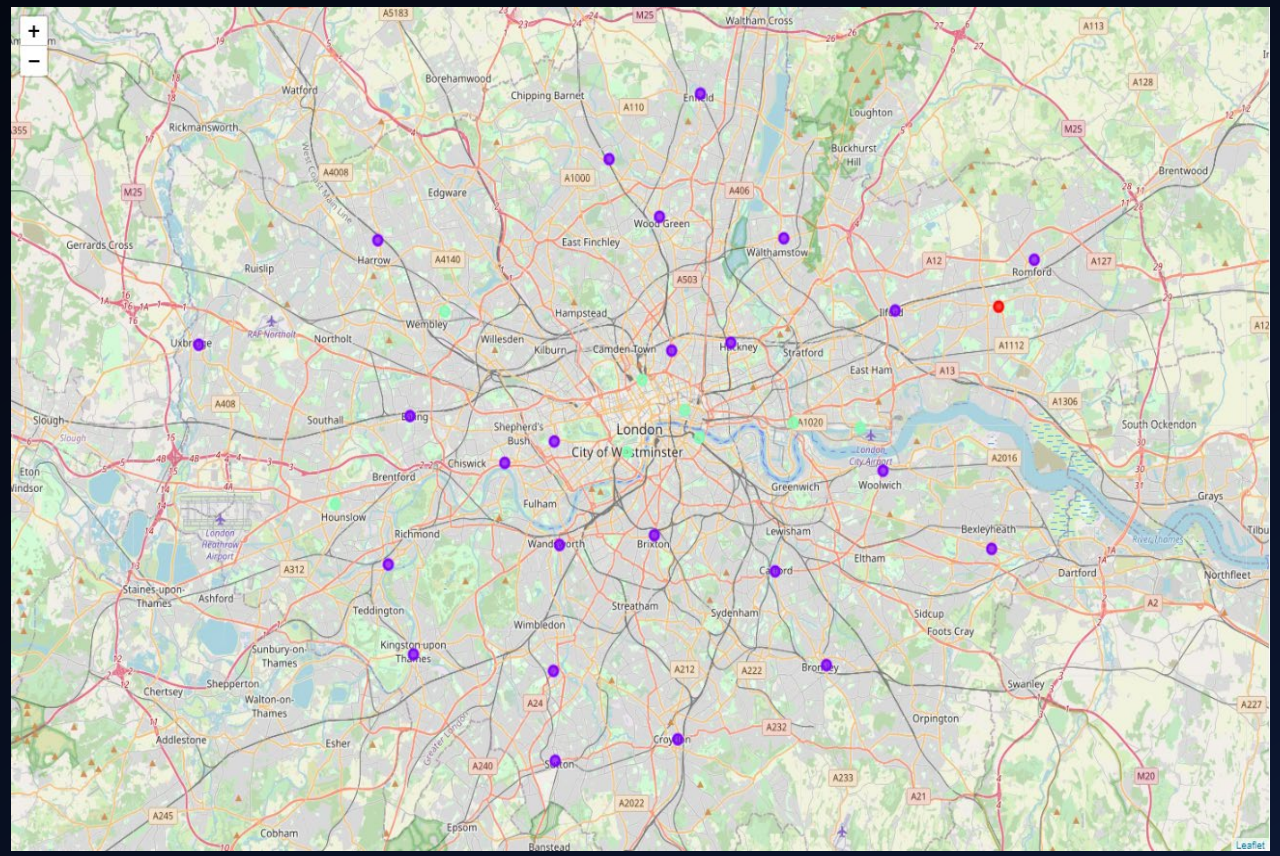
One hot encoding used to determine frequency with which different types of venues are found near each borough.

Data frame created containing **10 most common venues** found near each borough:

	Borough	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Barking and Dagenham	Bus Stop	Construction & Landscaping	Grocery Store	Diner	Discount Store	Chinese Restaurant	Park	Soccer Field	Pool	Supermarket
1	Barnet	Pub	Park	Rental Car Location	Gym	Bus Stop	Fish & Chips Shop	Financial or Legal Service	Film Studio	Filipino Restaurant	Fast Food Restaurant
2	Bexley	Pub	Clothing Store	Coffee Shop	Hotel	Fast Food Restaurant	Supermarket	Italian Restaurant	Pharmacy	American Restaurant	Bowling Alley
3	Brent	Coffee Shop	Hotel	Clothing Store	Bar	Indian Restaurant	Pizza Place	Sporting Goods Shop	Burger Joint	Grocery Store	Italian Restaurant
4	Bromley	Pub	Clothing Store	Coffee Shop	Indian Restaurant	Supermarket	Burger Joint	Electronics Store	Bar	Pizza Place	Gym / Fitness Center

K-Means Clustering Analysis

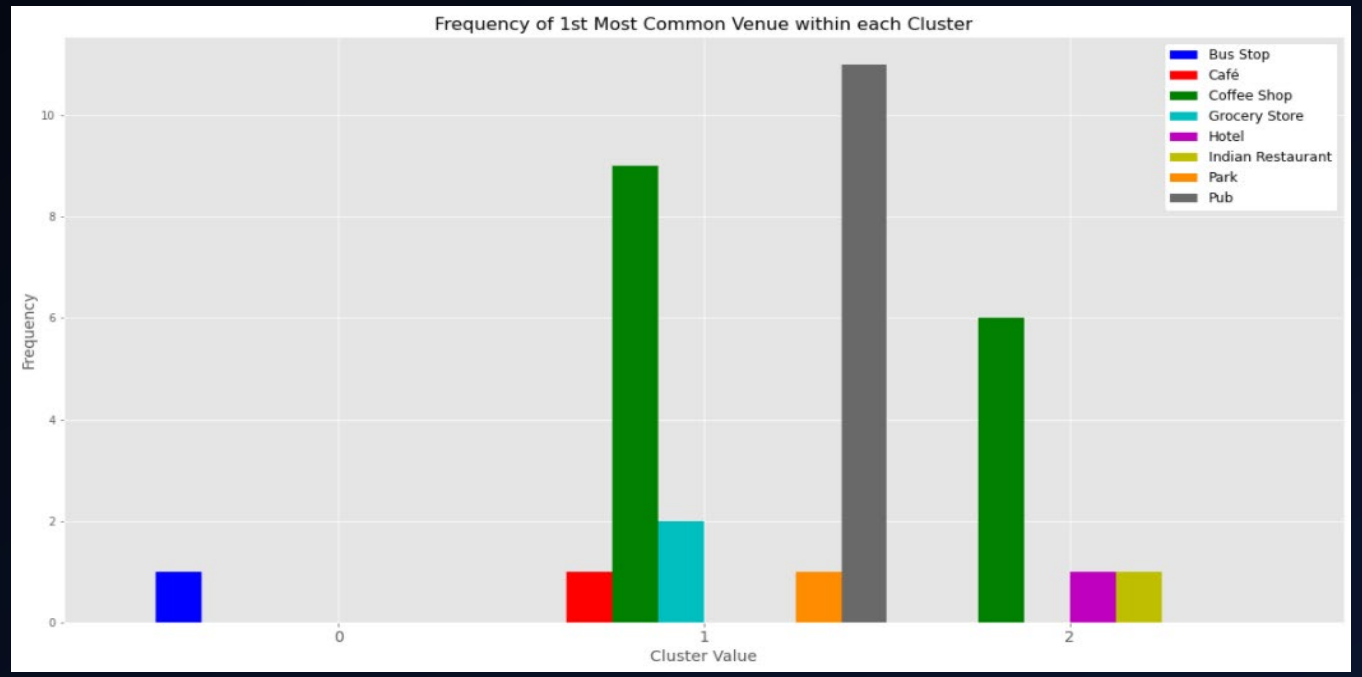
- Unsupervised **machine learning** clustering technique
- K value of **three** selected as optimum value for k
- Therefore the 33 boroughs clustered into three groups based on **similarity of venues** at each, visualised through a ***Folium*** map
- Relationship between venues and geographic location seems apparent



Exploring Clusters

Investigating the cluster venues alongside geographic location enables descriptive labels to be assigned to each

Frequency of 1st most common venues for each cluster visualised:



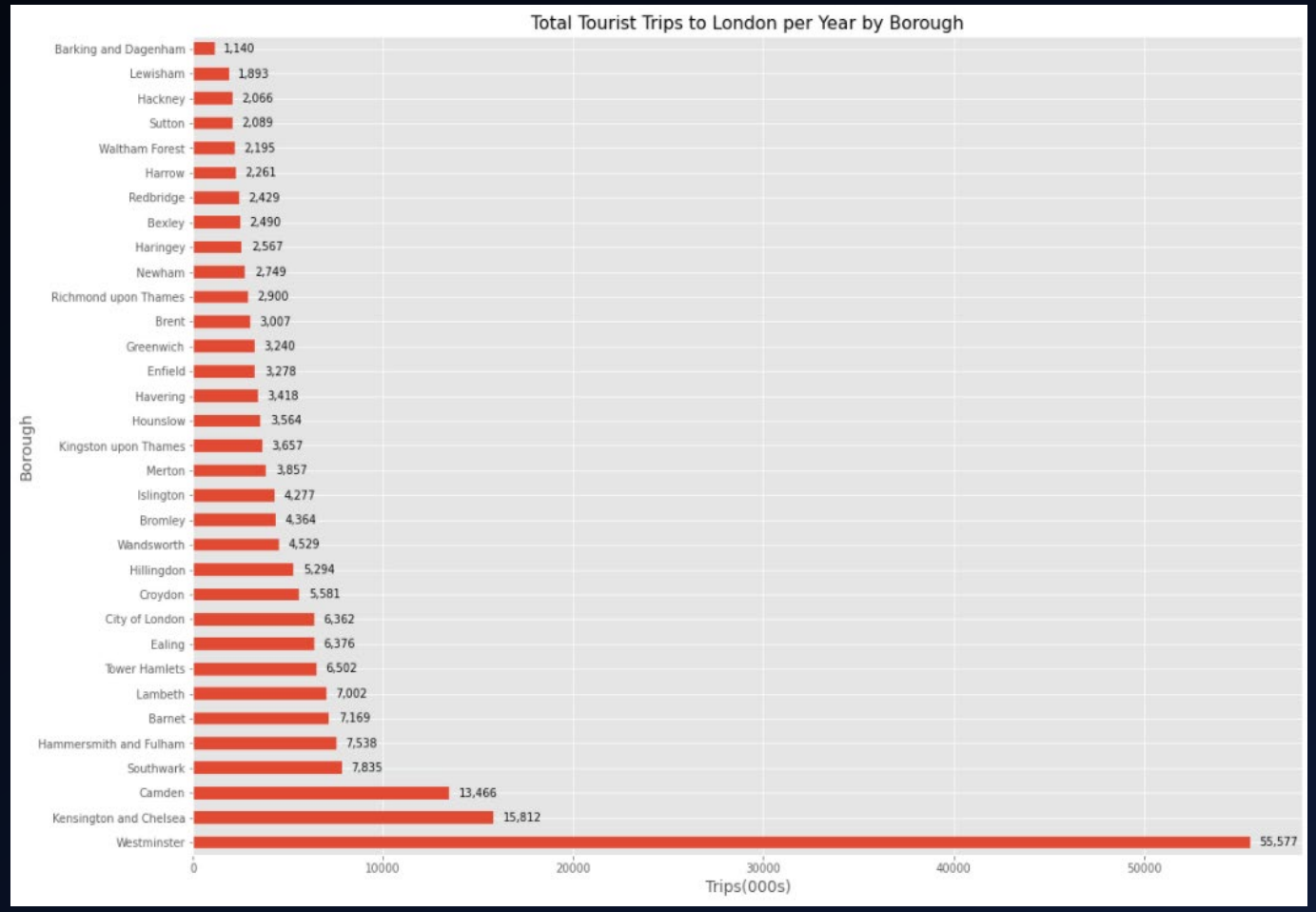
Cluster Descriptions

Based on geographical and venue information, cluster descriptions assigned:

- Cluster 0: “Atypical, non-hospitality”
  - One borough in this cluster located east, away from central London
- Cluster 1: “Pubs and coffee shops”
  - 24 boroughs located outside central London, with pubs being typically the most common venue found
- Cluster 2: “Coffee shops & hotels”
  - Eight boroughs typically centrally located with coffee shops the most encountered venue

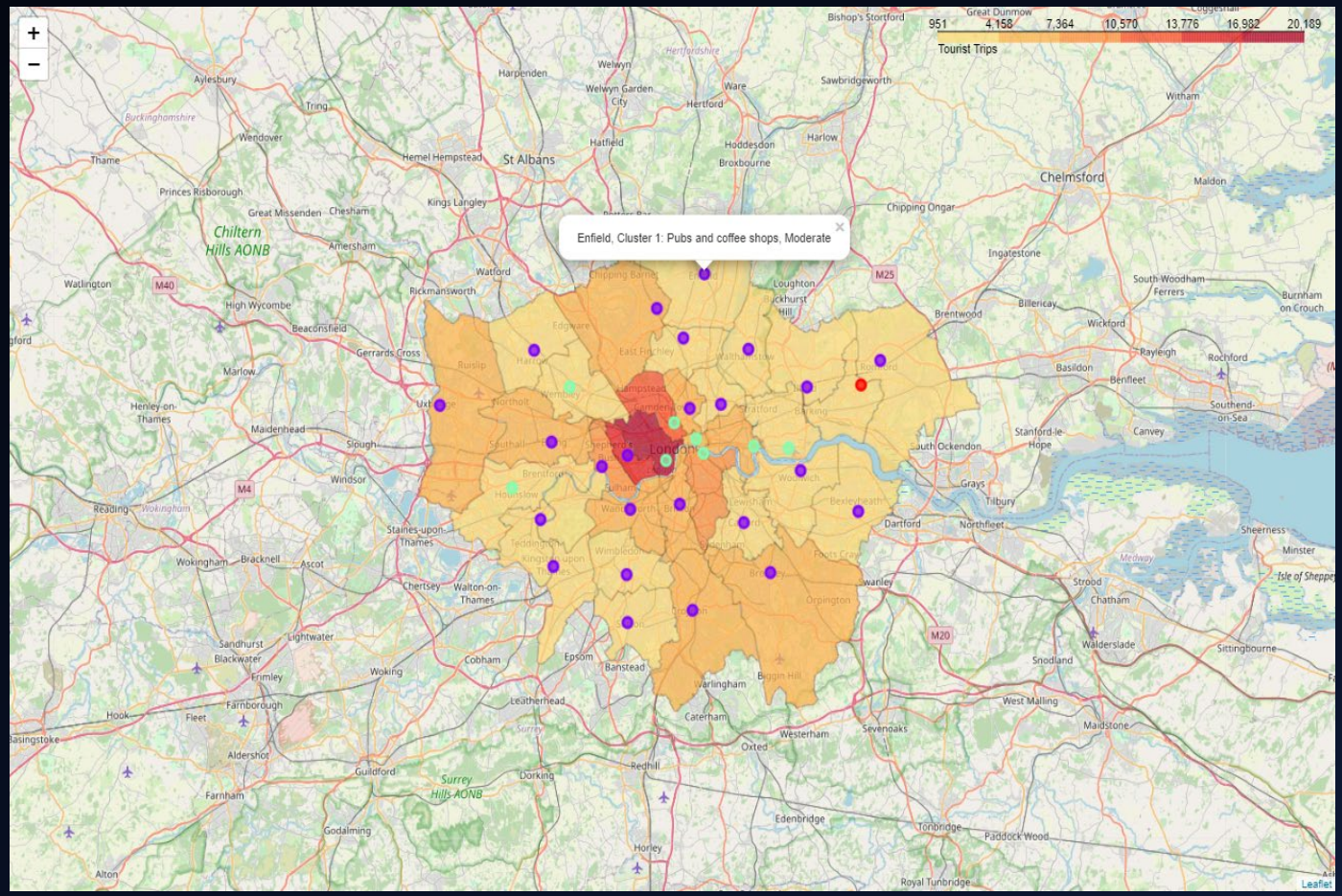
Investigating Tourist Trips Data

- Typically between 2000 and 8000 visits to each borough per year
- Westminster an outlier with over 55,000
- Categories created for adding to marker labels:
  - Below 3000: "Low"
  - 3000-5000: "Moderate"
  - 5000-8000: "High"
  - Above 8000: "Very High"



Choropleth Map with Cluster Markers

- Final choropleth map integrates and visualises the three clusters alongside the number of tourist trips for each borough
- Marker labels provide descriptive information including borough name, cluster value and description, and tourist popularity



Discussion Points

- Clear geographical difference with Cluster 1 boroughs located further out and Cluster 2 boroughs more central
- Both have a high number of coffee shops, though Cluster 1 has more pubs and Cluster 2 more hotels
- Centrally located boroughs experience more tourist trips, but variation still observed in outer boroughs

Answering the Business Problem

The following criteria can be used to answer the business problem of where would be a good location to open a craft beer bar:

1. Tourist trips: a ‘Popularity’ label of ‘High’ or ‘Very High’, indicating above 5000 tourist trips per year

2. Location / commonly found venues: any boroughs belonging to Cluster 1, where a borough has a less central location and a majority of pubs in the area

Conclusions

- An interactive map can be a valuable tool for those interested in the business problem, such as hospitality business owners, tourists, and even locals
- Future research could focus on integrating additional sources of data, such as commercial rental costs
- Understanding the demographic make up of local residents within each borough could illuminate any potential interest and likely customer base