



Visiting London City

Capstone Project – The Battle of Neighborhoods

Jorge Álvarez de la Fuente



LONDON

Business Problem

- London is a world tourist destination with full potential to become the best city to visit in Europe but it still has challenges for visitors.
- This project will analyse the venues available for people visiting London in order to look for the best recreational activities and locations in the variety of districts in London city.
- This project is interesting to visitors and expats who are considering visiting and relocating to London as they will be able to identify the most attractive locations and venues there.



LONDON

Data Acquisition and Sources

Data Acquired

- Postal Code
- District
- Postal Code Latitude
- Postal Code Longitude
- Venue
- Venue Latitude
- Venue Longitude
- Venue Category
- Most common venue

Sources

- Milesfaster website
- ArcGIS
- Foursquare API

LONDON



Data Cleaning

Scraping process of postal codes with their respective districts using Milesfaster website:

0	1	2	3
0	E1 Whitechapel, Stepney, Mile End	SE1 Waterloo, Bermondsey, Southwark, Borough	
1	E1W Wapping	SE2 Abbey Wood	
2	E2 Bethnal Green, Shoreditch	SE3 Blackheath, Westcombe Park	
3	E3 Bow, Bromley-by-Bow	SE4 Brockley, Crofton Park, Honor Oak Park	

Results after the data cleaning process:

	Postal Code	District
0	E1	Whitechapel, Stepney, Mile End
1	E1W	Wapping
2	E2	Bethnal Green, Shoreditch
3	E3	Bow, Bromley-by-Bow

LONDON

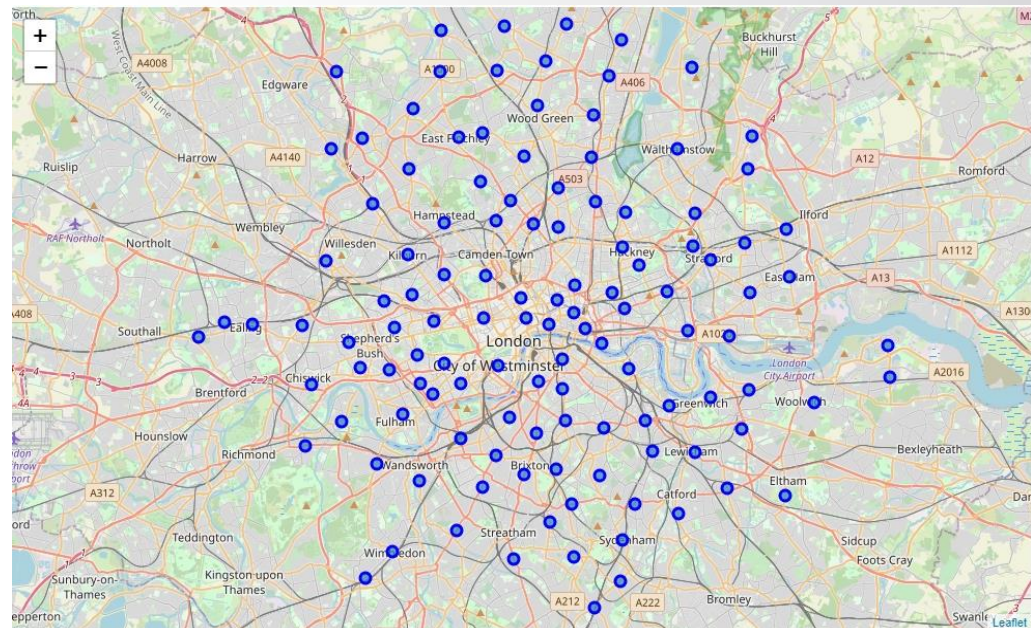


ArcGIS and Foursquare API

ArcGIS was used to get the coordinates for all districts:

	Postal Code	District	Latitude	Longitude
0	E1	Whitechapel, Stepney, Mile End	51.520220	-0.054310
1	E1W	Wapping	51.506282	-0.069426
2	E2	Bethnal Green, Shoreditch	51.526690	-0.062570
3	E3	Bow, Bromley-by-Bow	51.527020	-0.025940

Foursquare API was used to explore the districts and segment them. The map of London city was visualized:



LONDON



Foursquare API

- The function `getNearbyVenues` obtains information per venue and contains the API request url, the GET request. A DataFrame was created with new columns: district, district latitude, district longitude, venue name, venue latitude, venue longitude and venue category.
- The function `getNearbyVenues` retrieved a total of 11,145 venues.

District	District Latitude	District Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Whitechapel, Stepney, Mile End	51.52022	-0.05431	Rinkoff's Bakery	51.519964	-0.053238	Bakery
Whitechapel, Stepney, Mile End	51.52022	-0.05431	Mouse Tail Coffee Stories	51.519471	-0.058573	Coffee Shop
Whitechapel, Stepney, Mile End	51.52022	-0.05431	Genesis Cinema	51.521036	-0.051073	Movie Theater
Whitechapel, Stepney, Mile End	51.52022	-0.05431	Second Shot	51.527412	-0.056625	Coffee Shop
Whitechapel, Stepney, Mile End	51.52022	-0.05431	Old Street Brewery & Taproom	51.526950	-0.056426	Brewery
Whitechapel, Stepney, Mile End	51.52022	-0.05431	Stepney Green Park	51.517768	-0.047054	Park
Whitechapel, Stepney, Mile End	51.52022	-0.05431	Renegade London Wine	51.527005	-0.056381	Wine Bar
Whitechapel, Stepney, Mile End	51.52022	-0.05431	Mother Kelly's Bottle Shop and Tap Room	51.528413	-0.055843	Beer Bar



LONDON

Methodology

- Foursquare API was used to explore all venues of all districts and analyse each district in London.
- The most relevant venues to visit were retrieved from the entire DataFrame and visualized.
- The most common venues were analysed using word cloud to look for the most common words and translate them into venues.
- Each district was grouped with the 10 most common venues.
- All districts in London city were clustered, visualized and examined using k-means algorithm.
- The resulting clusters were analysed using word cloud to validate the previous analyses.

LONDON

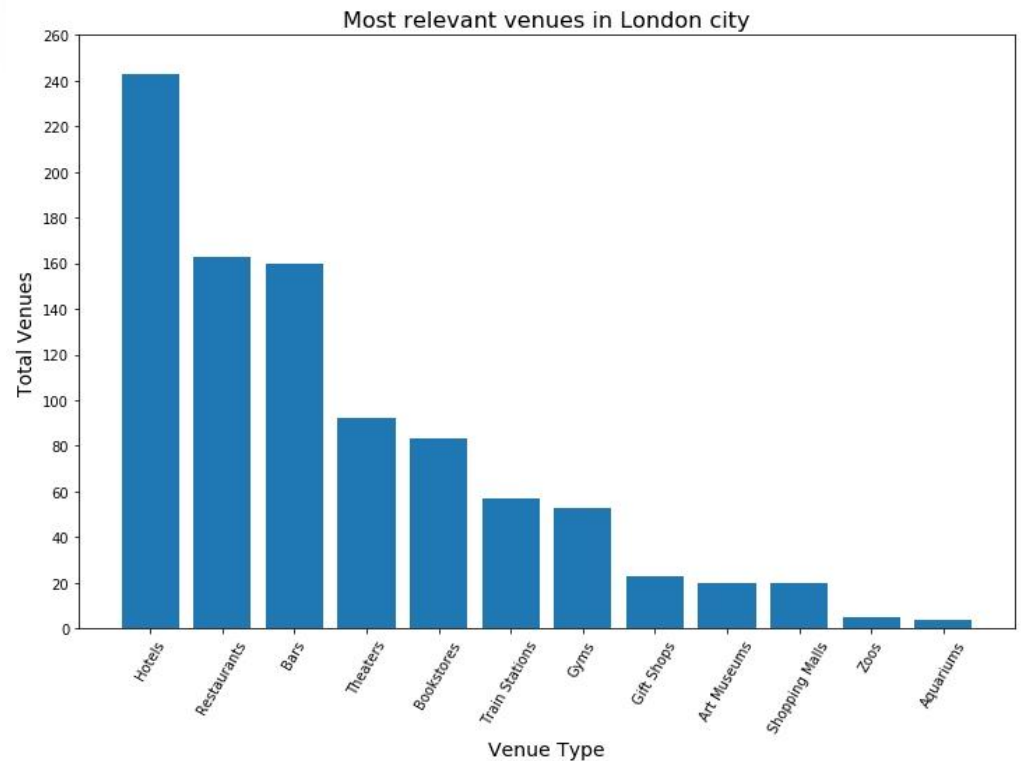


Explanatory Analysis. Analyzing districts

This section consists in performing an explanatory data analysis and deriving some additional information.

A new DataFrame containing random chosen relevant venue categories was created to evaluate whether or not London is a good place to visit.

The visualizations were developed using bar plots as follow.

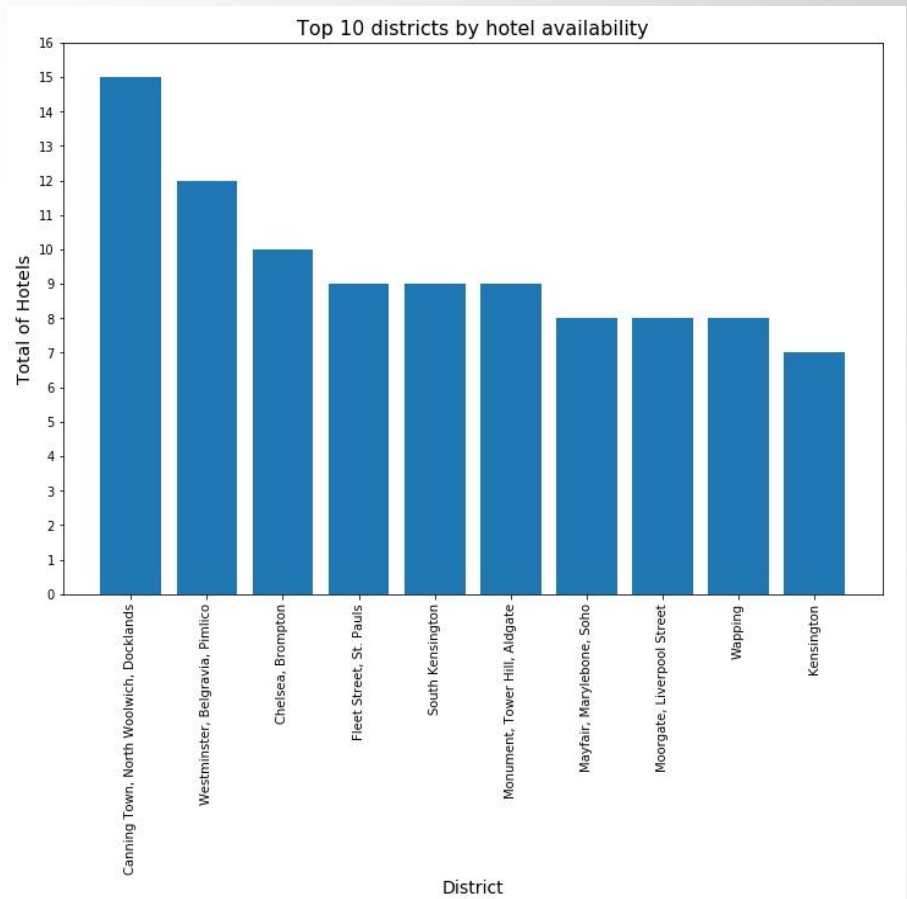


LONDON



Explanatory Analysis. Analyzing districts

From the venues randomly selected, hotels are the most common venue in London city, followed by restaurants. Hotels were grouped by district to find the top 10 districts that have more hotels in London city.

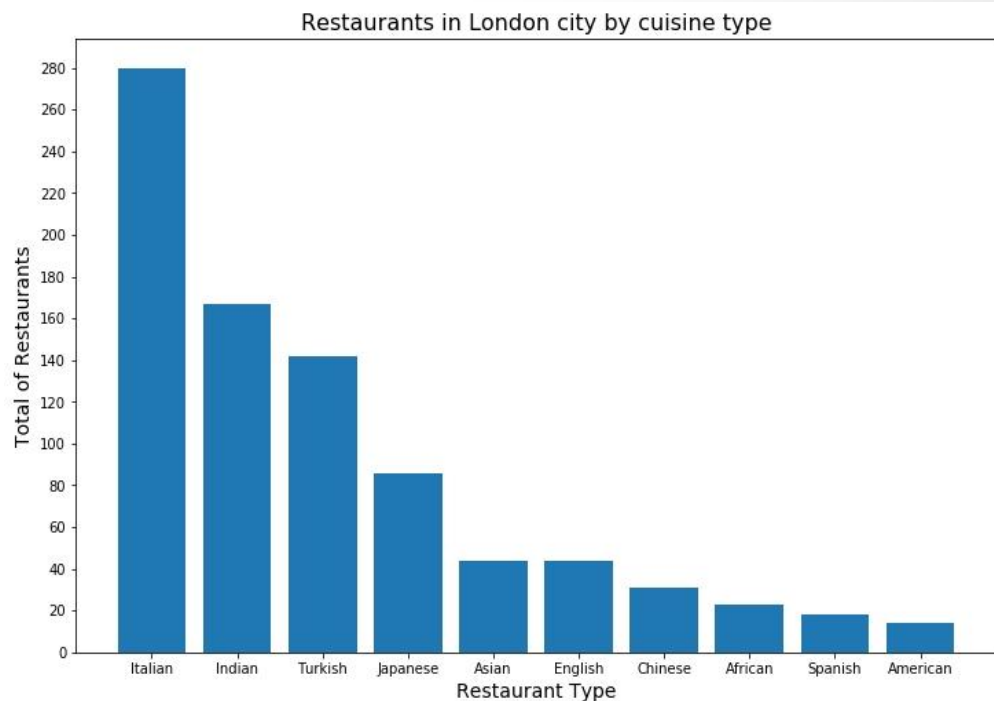


LONDON



Explanatory Analysis. Analyzing districts

Restaurants is the second most popular venue from the random categories selected.

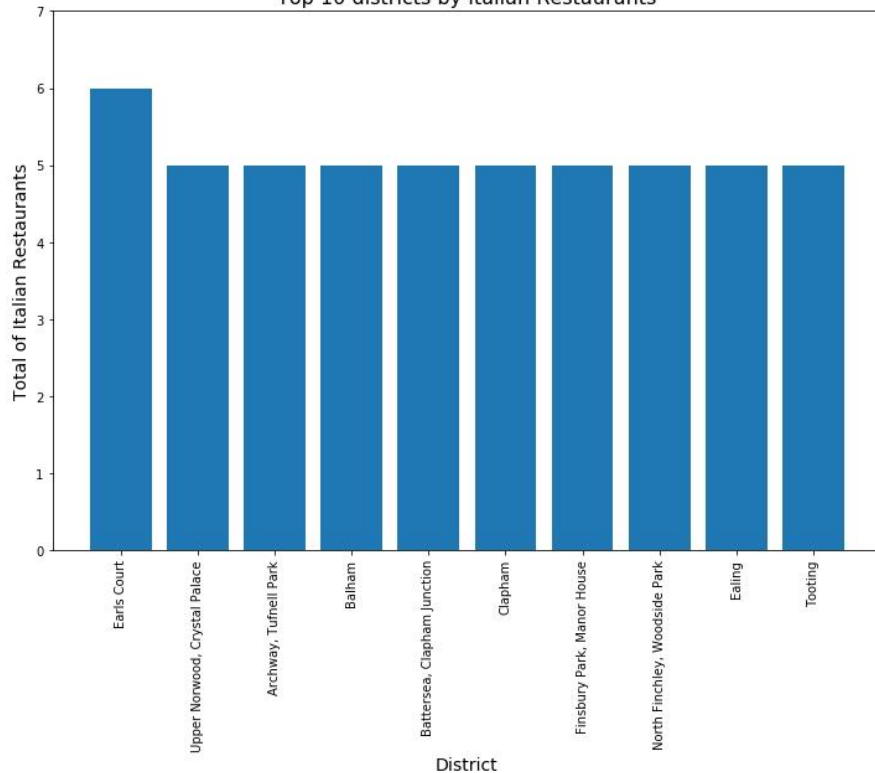


LONDON

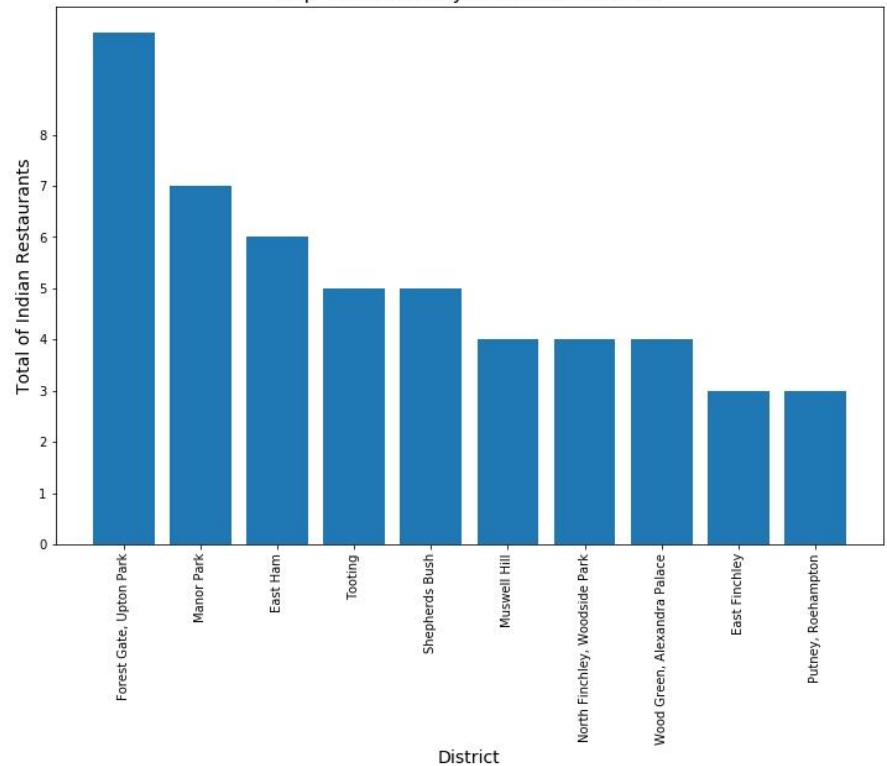


Explanatory Analysis. Analyzing districts

Top 10 districts by Italian Restaurants



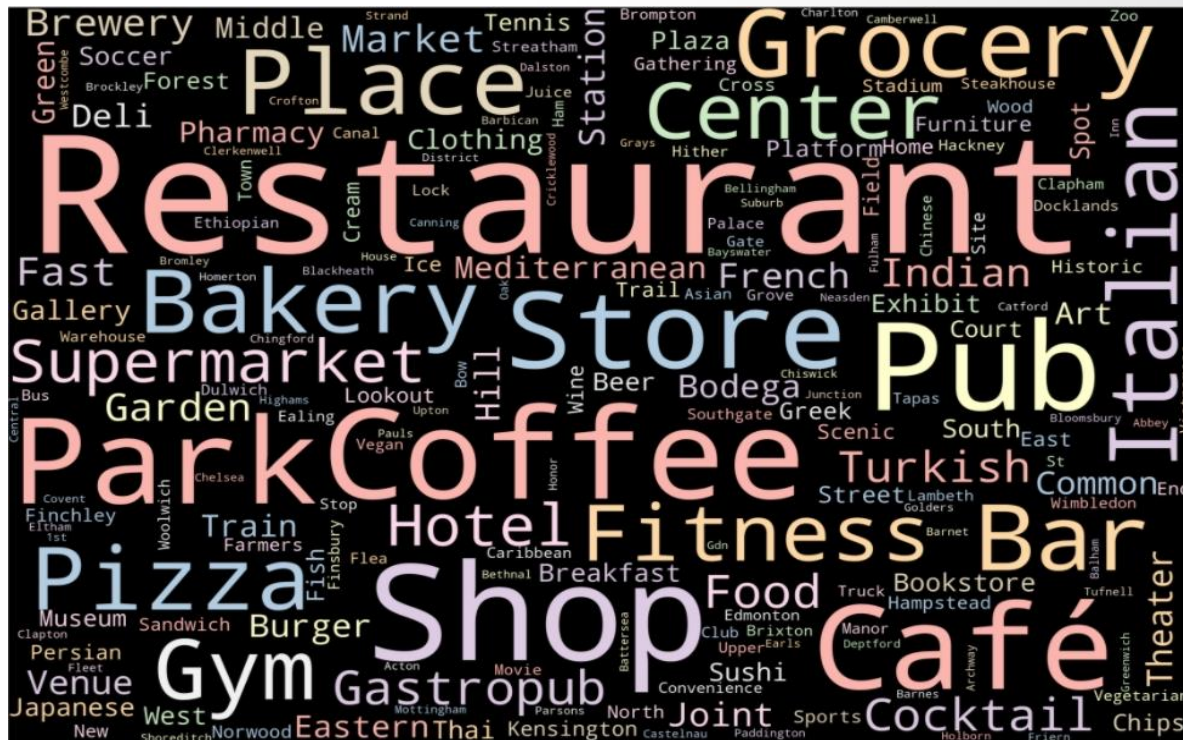
Top 10 districts by Indian Restaurants



A close-up, low-angle shot of a roller coaster track. The track is a vibrant blue and curves gracefully over a white, ornate building facade. The building features classical architectural elements like columns and arched windows. The sky is a clear, bright blue.

Explanatory Analysis. Analyzing districts

Word clouds visualization was applied to find the most common venues in London city.

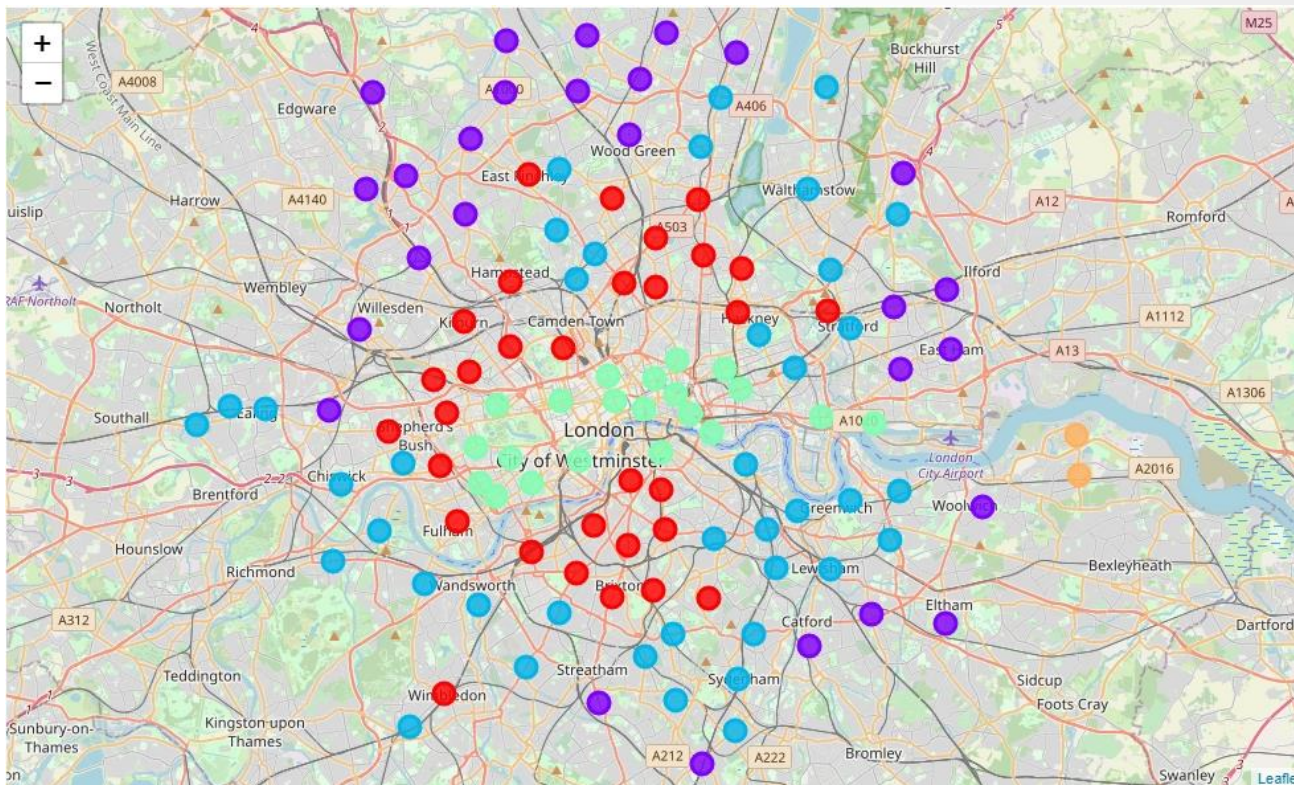


LONDON



Explanatory Analysis. Clustering districts

London city map with all clusters represented with different colors



Map legend:

Cluster 0: Red

Cluster 1: Purple

Cluster 2: Light Blue

Cluster 3: Light Green

Cluster 4: Light Orange

Explanatory Analysis. Examining districts

Cluster 0 Word Cloud



Cluster 1 Word Cloud

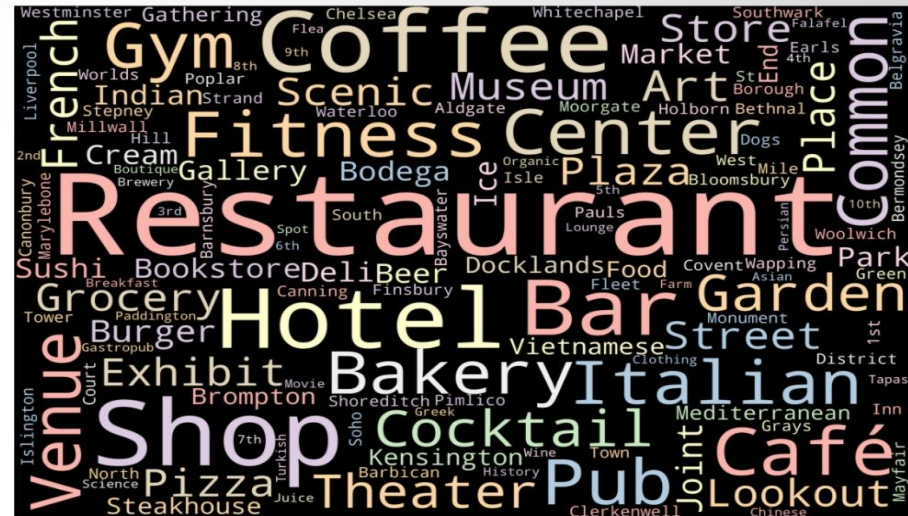


Explanatory Analysis. Examining districts

Cluster 2 Word Cloud



Cluster 3 Word Cloud





LONDON

Conclusions

- From the visitors perspective, the most relevant venues found in the explanatory analysis were hotels, restaurants, bars, theaters, pubs, bakeries, groceries, coffee shops, gyms.
- The most popular cuisine types were italian food, indian food, turkish food and japanese food.
- Most hotels were found in districts located in southwest and north of London city, while most restaurants were found in districts located in east and northeast of London city.
- It was found that most of the venues revealed during the explanatory analysis are located in London center, center-east, south and north periphery too.

A banner image at the top of the slide. On the left, the word "LONDON" is written in large, white, sans-serif capital letters against a blue background with a white cloud pattern. To the right of the text, the image transitions into a photograph of a blue roller coaster track curving over a stone building, likely a London landmark.

LONDON

Limitations

- There are a number of gaps in the present project that could be benefitted from further research.
- The explanatory analysis results depend on the accuracy of Foursquare data and the Milesfaster website has limitations as per the number of postal codes included.
- A more detailed evaluation should have included prices and iconic attractions in London city such as Buckingham Palace and Big Ben.