A decorative graphic on the left side of the slide consisting of two overlapping parallelograms. The front one is blue and the back one is a light green. They are positioned diagonally, with the blue one partially covering the green one.

Exploring areas around London Underground Stations

Edwin Nicolaas Asberg



LIVING NEARBY METRO STATIONS

- Traffic jams are a major issue in big cities
- People tend to search for places nearby public transportation to live

PROBLEM:

- How to characterize areas around metro stations to help individuals to pick a region to live?
 - Using London as a case study



DATA AND METHODOLOGY

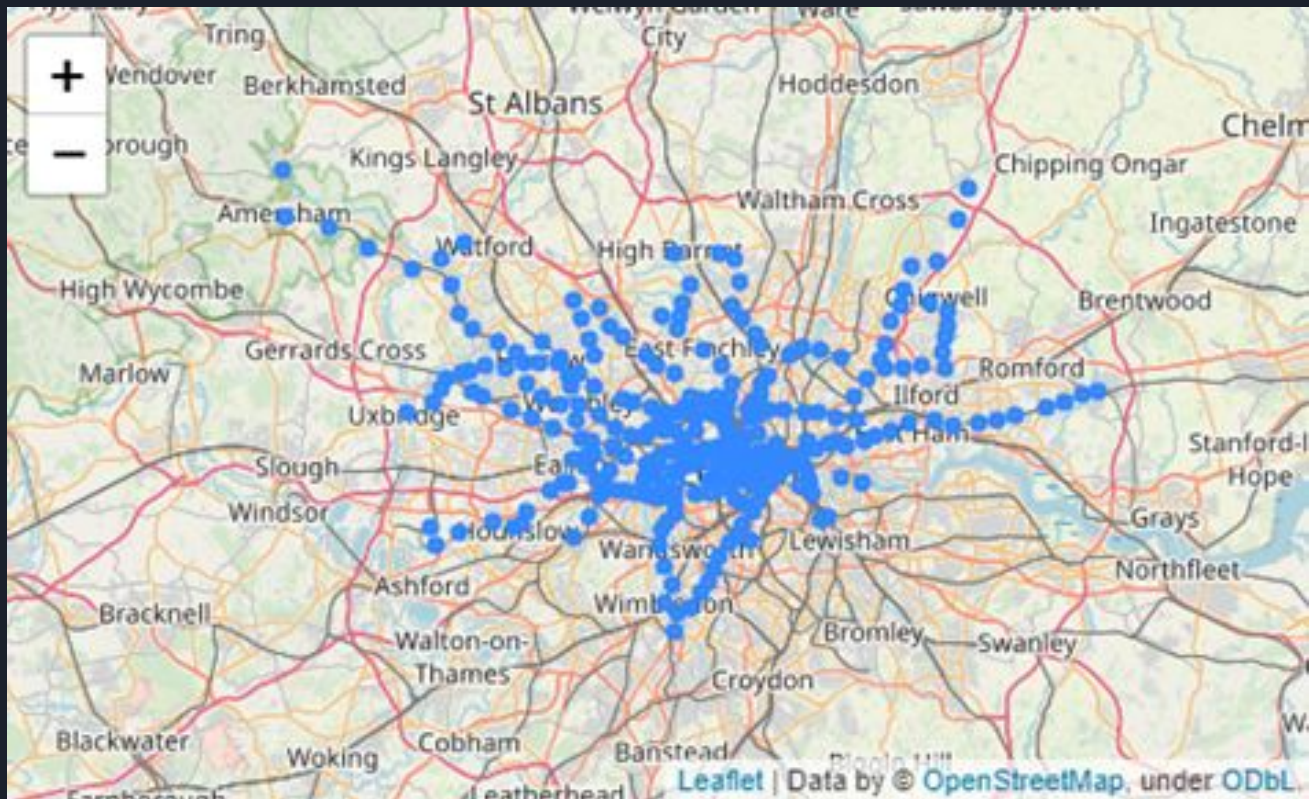
DATA

- https://wiki.openstreetmap.org/wiki/List_of_London_Underground_stations#Tube_Stations_List
- Foursquare Places API

METHODOLOGY

- K-means clustering

LONDON UNDERGROUND OVERVIEW



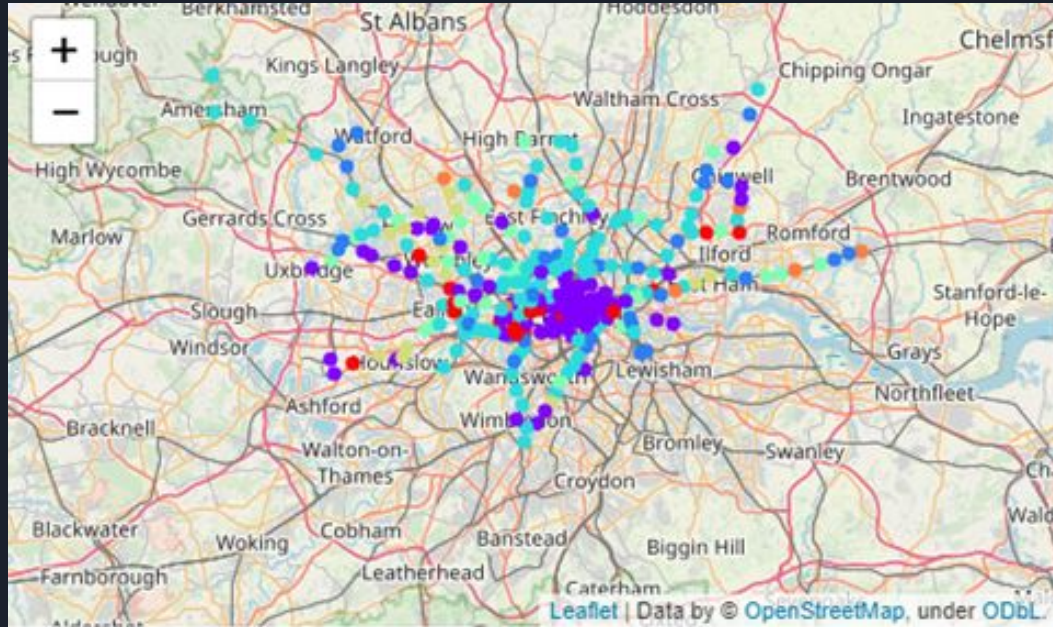
DATA TREATMENT

- Merging Underground geotagged data with venues retrieved from Foursquare (nearby 500 meters of each station)
- Selecting the top-5 most common venues for each underground station

	Underground	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Acton Central	Gym / Fitness Center	Pub	Coffee Shop	Wine Bar	Bakery
1	Acton Town	Museum	Kebab Restaurant	Pizza Place	Coffee Shop	Fish & Chips Shop
2	Aldgate	Hotel	Coffee Shop	Cocktail Bar	Gym / Fitness Center	Salad Place
3	Aldgate East	Coffee Shop	Pub	Hotel	Indian Restaurant	Café
4	Alpertton	Supermarket	Gym / Fitness Center	Indian Restaurant	Middle Eastern Restaurant	Asian Restaurant

LETTING THE ALGORITHM WORK

- K-means clustering algorithm ($k = 7$):





UNDERSTANDING EACH CLUSTER

- **Cluster 0**: Stations within this cluster are mainly surrounded by Hotels. Pubs are also important venues;
- **Cluster 1**: Around the stations in this cluster, there are a variety of venues with emphasis on Coffee Shops;
- **Cluster 2**: This cluster has Pubs and Cafés as the most common venues;
- **Cluster 3**: This is the largest cluster, meaning the most heterogeneous as well. Coffee Shops and Pubs are well offered around the stations here;
- **Cluster 4**: In this cluster, stations are surrounded by Grocery Stores;
- **Cluster 5**: If you are into Indian food, these stations are the place to live. Indian Restaurants are the most common venues here;
- **Cluster 6**: Gym and Fitness Centre and Parks are well offered in this small cluster (only 7 stations).

Tables available here:

https://github.com/enasberg/IBM_Data-Science_Certificate/blob/master/Capstone%20Project%20-%20Week%202.ipynb



CONCLUSION

- Unsupervised algorithm (k-means) had a good response;
- Innovative way to tackle the problem;
- And most important...

... People are empowered with more information to take their own decision.

Thank
You!

