

Coursera Capstone Project

London stations in context of Business development and Tourist relations.

Background:

- How to establish the best London subway stations with the highest density of amenities around the given stations in the inner London area?
- The goal is twofold:
 - What stations areas have room for further Business Development and what areas the developers may focus on?
 - Learning where is the highest number of eateries and other amenities required to satisfy all the necessities of life in a densely populated area such as London both for residents and tourists alike.

Who could be interested to learn about this?

- This is targeting both Business developers who can expand into areas that still haven't been accommodating local residents needs and Tourists alike and Tourists who are looking for the station areas with the highest level of amenities around train stations they can have a quick access to due to their lack of familiarity with inner London
- [London economy](#) is almost 17% of the whole UK GDP, thus Business development in this UK city could have multiple economical impacts and benefits
- Tourism is a vital part of the UK economy, accounting for [9.6% of jobs](#) ;London represents [55% of the tourism market](#), hosting a record 19.8m visits in 2017 ().
- Bearing in mind that visitors spending reached between [2013-2017 over 17 billion Pounds](#), it is prudent to understand the capabilities that transportation grid may offer in the field of Business Development

Why should this be relevant for the target group?

- Finding a variety of amenities (eateries, convenience stores working 24/7, shops of all sorts) can be a frustrating, so the key is a quick access around the transport grid that is vital both for the busy workforce and students, but also tourists who also are not familiar with the

neighbourhood they are in at the moment, and they very often rely on the subway or trains because not feeling to be in charge of the unknown space.

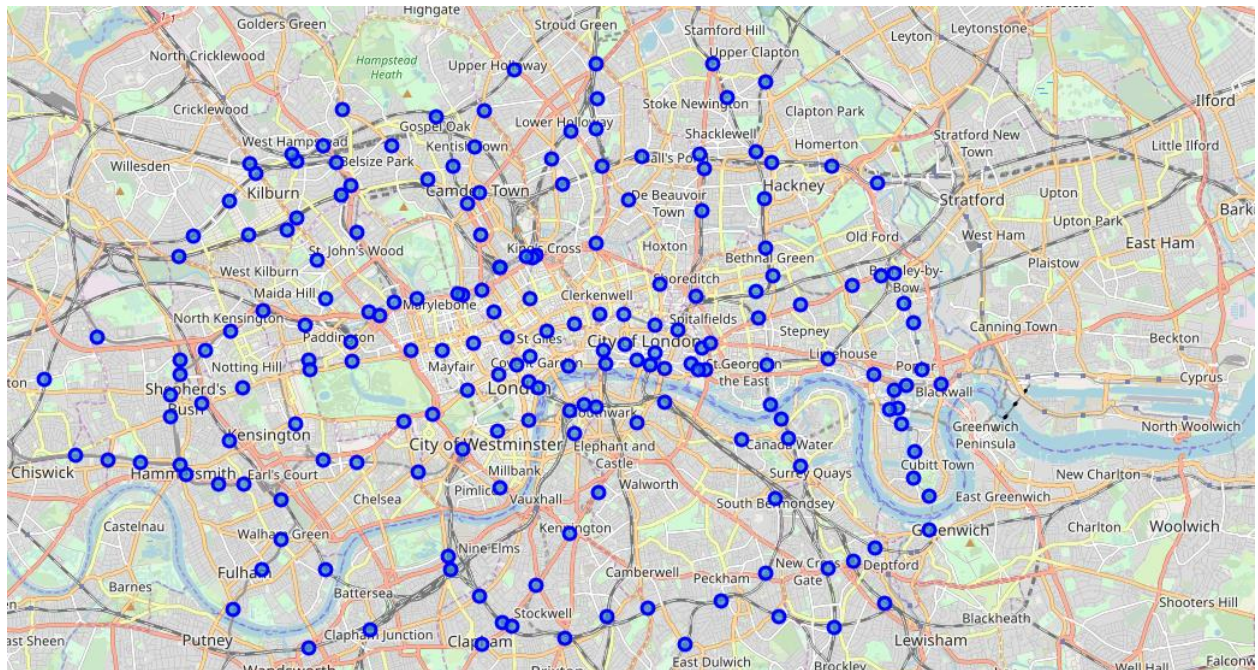
- So, this little project should help also interested parties who are looking for places around the transportation areas where there are still opportunities for Business development, and on the other side for the people who are looking for the aforementioned quick access to the city and all it has to offer.
- This solution can be extended to any neighbourhood in any other big city.

This report will cluster London Underground stations in zones 1 and 2 to help tourists and businesses plan the best route around the city.

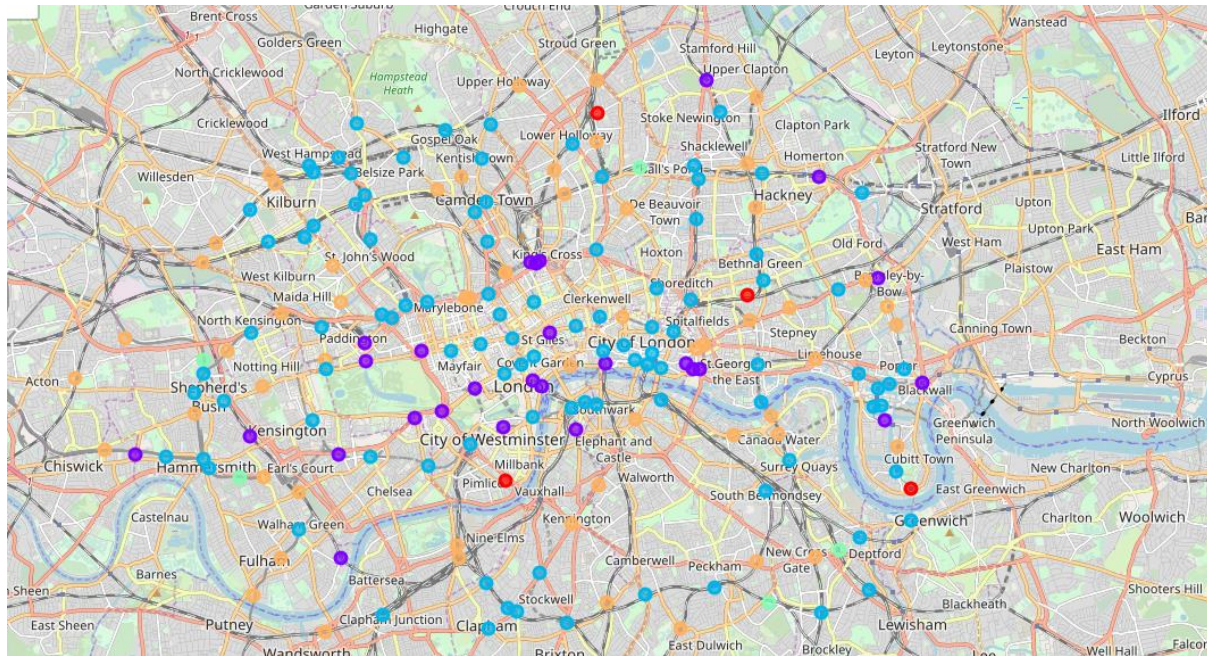
Method and Data Source

- The project has been developed by being carried out on the project Jupyter Notebook importing all relevant Python libraries required for various computation:
 - *Pandas* for data manipulation and analysis. In particular, it offered good data structuring and operations for manipulating numerical tables and series
 - *Numpy* for supporting the large dataset, helping with multi-dimensional arrays and matrices, along with a large collection of high-level mathematical functions to operate on these arrays and help us to move forward
 - We decided to go with *K means clustering (unsupervised learning)* so we can get into the data obtaining the info that may not occur us if we went for supervised learning models. We imported *scikit-learn* library to help us with K means clusters
 - In order to fully visualize the stations areas and clusters we used *Folium* and *Matplotlib* that helped with the precise and accurate plotting.
- The data sources that we have used were:
- Popular geospatial app with venues -Foursquare API that we used for extracting venues nearby underground stations. We focused on 10 most common venues around them to shorten the time for computation
- [Doogal](#) datasets, out of which the one with London stations helped us with which providing the list of all London Underground stations in a recognizable CSV file we could further manipulate with Pandas.
- We had the chance to send multiple queries to FourSquare was queried using their API and Latitudes and Longitudes of London stations as major input.

London Stations visualized in unclustered manner (Inner London Zone 1 and 2)



Map of Clustered Stations Area linked with venue categories in London inner area Zone 1 and 2



- **Red Cluster** (1) = 4 stations areas

This Cluster is showing us that the most frequent venues are parks and there might be a place for Business development (sport facilities could be an option) as this is our smallest Cluster and it shows that these areas are not high on tourists agenda.

- **Purple Cluster** (2) = 30 station areas

Cluster 2 is listing hotels as most frequent ones, as this is the most frequented area by Tourists, so the market might be saturated for further development in the hospitality filed.

- **Blue Cluster** (3)= 99 station areas

This is our cluster with highest venue density-it's an eatery hotspot , it is a great spot for Tourists but not a winner in terms of Business development.

- **Green Cluster** (4) = 5 stations areas

This Cluster is on a low point for Tourists and there might be some opportunities within Business Development but not as high as Cluster 1 as this area is not really accessible and its rather suburbia, the development should be focused on accommodating needs of the area residents (i.e. no fast food restaurants in top 5 venues)

- **Orange Cluster** (5) =63 station areas

- This is an area with Pubs being nr one venue category and there are fast food restaurants so the venues are catering obviously to Tourists and highly commuting workforce. However high density is not giving any room for long-term Business Development