# Introduction & business problem

London is the capital and largest city of England and the United Kingdom. Standing on the River Thames in the south-east of England, at the head of its 50-mile (80 km) estuary leading to the North Sea, London has been a major settlement for two millennia. Londinium was founded by the Romans. The City of London, London's ancient core − an area of just 1.12 square miles (2.9 km2) and colloquially known as the Square Mile − retains boundaries that closely follow its medieval limits.

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-for-work city. London exerts a considerable impact upon the arts, commerce, education, entertainment, fashion, finance, healthcare, media, professional services, research and development, tourism and transportation. London ranks 26th out of 300 major cities for economic performance. It is one of the largest financial centres and has either the fifth or the sixth largest metropolitan area GDP. It is the most-visited city as measured by international arrivals and has the busiest city airport system as measured by passenger traffic. It is the leading investment destination, hosting more international retailers and ultra high-net-worth individuals than any other city. London's universities form the largest concentration of higher education institutes in Europe, and London is home to highly ranked institutions such as Imperial College London in natural and applied sciences, the London School of Economics in social sciences, and the comprehensive University College London and King's College London.

The Accomp Corp is an investment group, they have a plan to open hotel in London. Therefore, Accomp Corp has requested me to do research on London city neighborhoods to find some location that suitable to theirs new hotel in London.

# Data Extraction & Preliminary Analysis

As I am requested by the Director of the Accomp Corp to explore neighborhoods in the city before choosing location for future hotel in London city areas, I have decided to explore locations and surrounding venues through third party data source provider FourSquare through their API for fetching data. Foursquare is a technology company that uses location intelligence to build meaningful consumer experiences and business solutions. They have app developers can install and they also provide more than 75 million short tips from local experts around the world.

The London's Districts, Wards data included theirs location I get from https://www.doogal.co.uk/london\_postcodes.php. In this website they provide many more data features as well.

# Methodology

A part of my methodology, I have used Foursquare API for this project analysis, identifying existing venues to investigate further the standard of living in London city neighborhoods.

As the initial data was not clean, I had to do data aggregation, transformation and clean up for exploratory data analysis.

Then, I have used K-means clustering to organize neighborhoods into 5 different clusters based on venue's categories similarity and frequencies.

The K-means clustering will find similarities among neighborhoods in the same cluster and differences among neighborhoods different clusters. I have downloaded folium map to visually identify neighborhoods in each cluster.

The analysis mainly applies the following Python libraries:

Pandas, Numpy – Libraries for data storage, manipulation and array computing

Geopy – Library to retrieve locational data

Matplotlib, Folium – Libraries for representing numeric and locational data

Requests, Urllib – Libraries to retrieve data and handle http exchange with the Foursquare API

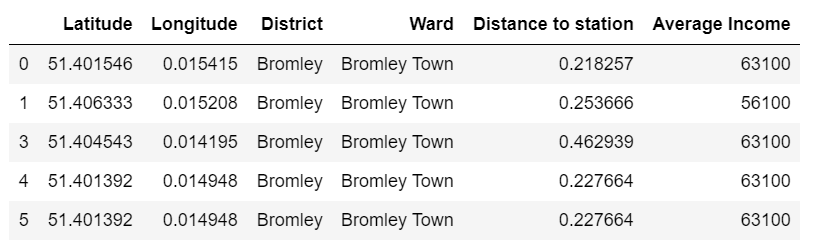
## Data process & Exploration

The London's Districts, Wards data included theirs location I get from <https://www.doogal.co.uk/london_postcodes.php>. In this website they provide many more data features as well.

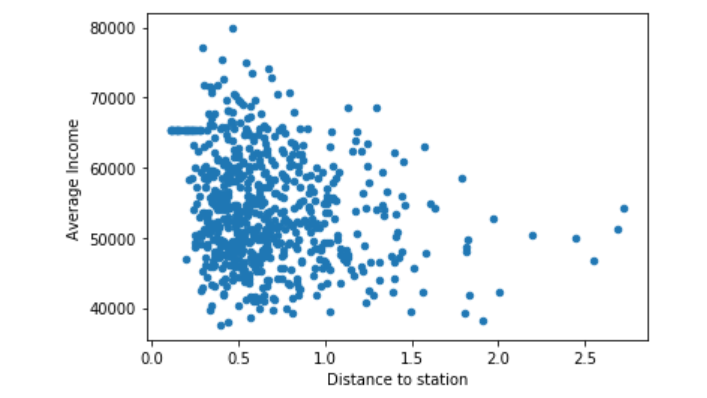
For convenience, I downloaded data to the csv file named London postcodes.csv



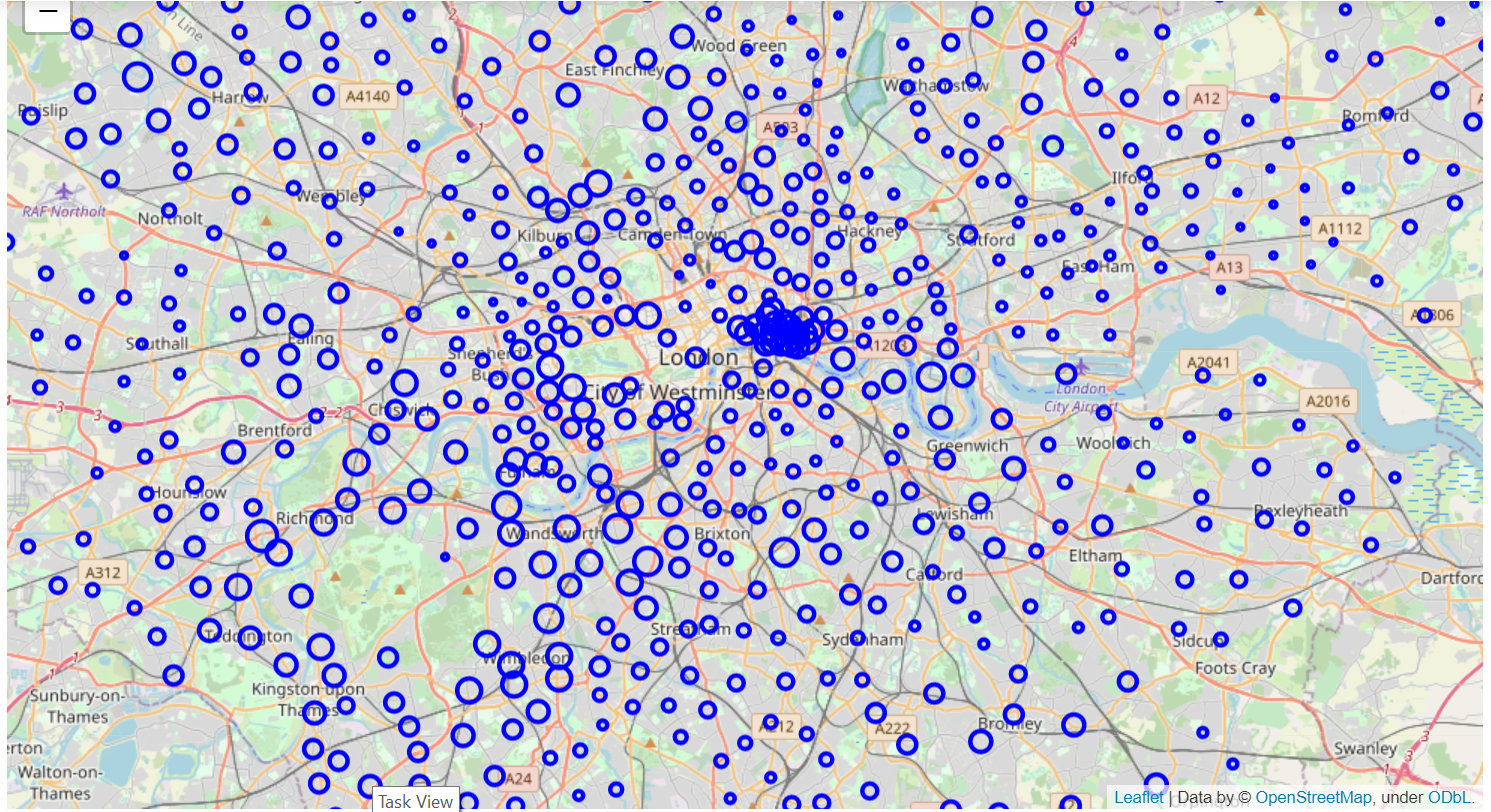
Only get useful columns



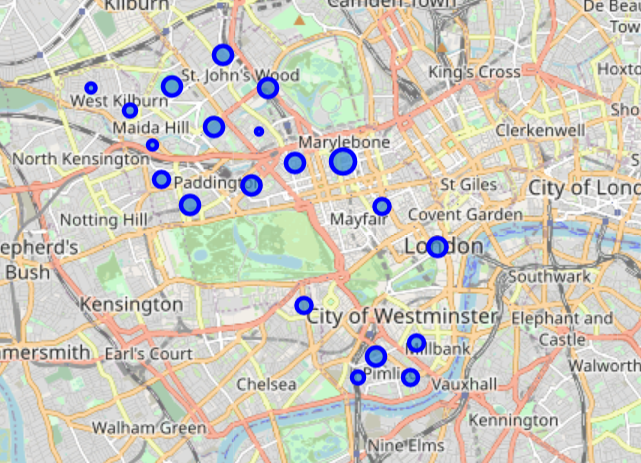
The plot show the impact of Distance to Station to Average Income. I found that the closer to station the higher income, but highest income in range 0.5 miles to the stations.



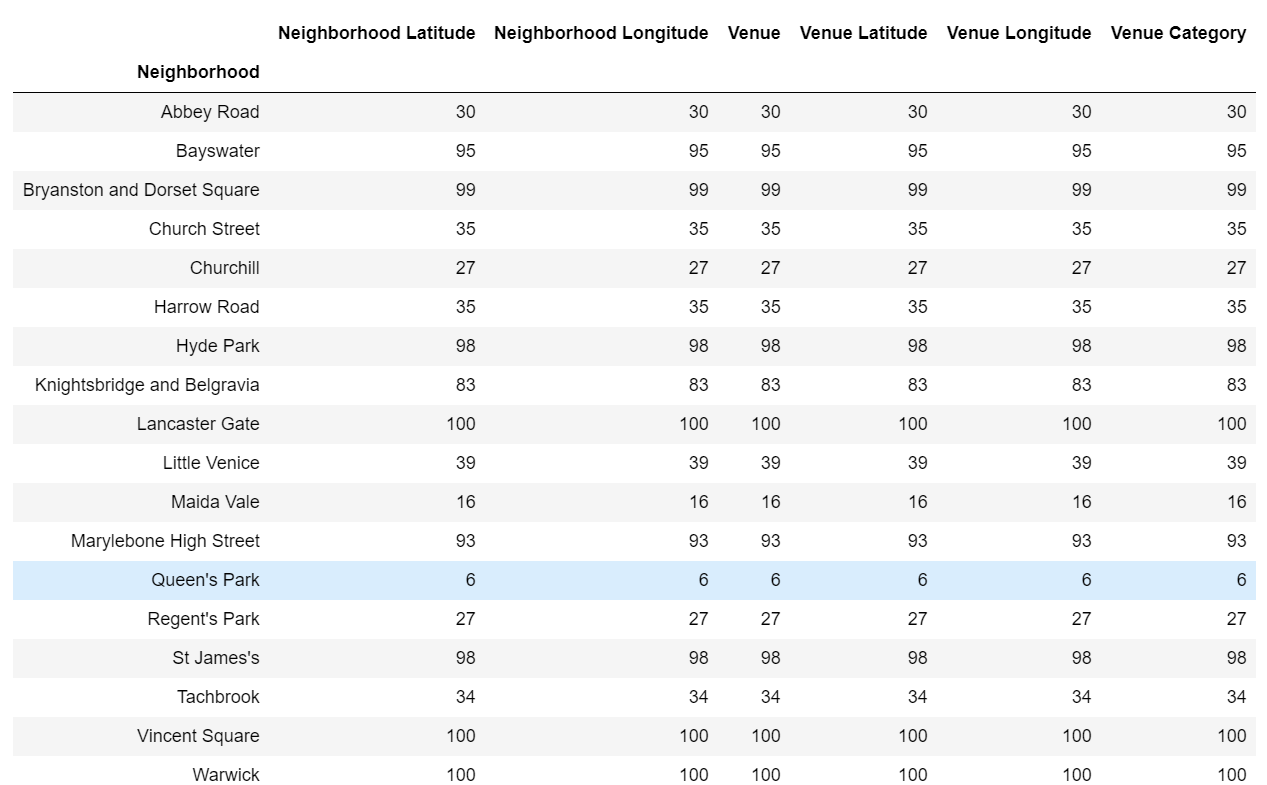
I using Folium map to display all neighborhoods of London city. The bigger circles the higher average income.



Then I focus in Westminster only to minimize the data.



The using Foursquare to find venues in Westminster, group by each neighborhood



The most frequent venue types in each neighborhood

----Abbey Road----

venue freq

0 Café 0.10

1 Grocery Store 0.07

2 Deli / Bodega 0.07

3 Coffee Shop 0.07

4 Garden 0.03

----Bayswater----

venue freq

0 Hotel 0.11

1 Pub 0.09

2 Pizza Place 0.05

3 Café 0.05

4 Garden 0.04

----Bryanston and Dorset Square----

venue freq

0 Middle Eastern Restaurant 0.09

1 Coffee Shop 0.07

2 Café 0.04

3 Lebanese Restaurant 0.04

4 Hotel 0.04

----Church Street----

venue freq

0 Coffee Shop 0.09

1 Theater 0.06

2 Grocery Store 0.06

3 Sandwich Place 0.06

4 Yoga Studio 0.03

----Churchill----

venue freq

0 Hotel 0.15

1 Deli / Bodega 0.07

2 Pub 0.07

3 Garden 0.07

4 Bakery 0.07

----Harrow Road----

venue freq

0 Grocery Store 0.11

1 Coffee Shop 0.09

2 Park 0.09

3 Café 0.06

4 Indian Restaurant 0.06

----Hyde Park----

venue freq

0 Hotel 0.13

1 Pub 0.06

2 Italian Restaurant 0.06

3 Coffee Shop 0.06

4 Café 0.04

----Knightsbridge and Belgravia----

venue freq

0 Café 0.11

1 Italian Restaurant 0.10

2 Hotel 0.07

3 Coffee Shop 0.06

4 Boutique 0.05

----Lancaster Gate----

venue freq

0 Hotel 0.17

1 Coffee Shop 0.06

2 Café 0.06

3 Pub 0.06

4 Chinese Restaurant 0.05

----Little Venice----

venue freq

0 Pub 0.15

1 Café 0.13

2 Coffee Shop 0.05

3 Lebanese Restaurant 0.05

4 Pizza Place 0.03

----Maida Vale----

venue freq

0 Deli / Bodega 0.12

1 Thai Restaurant 0.12

2 Coffee Shop 0.12

3 Japanese Restaurant 0.06

4 Wine Shop 0.06

----Marylebone High Street----

venue freq

0 Burger Joint 0.05

1 Hotel 0.05

2 French Restaurant 0.04

3 Italian Restaurant 0.04

4 Coffee Shop 0.03

----Queen's Park----

venue freq

0 Yoga Studio 0.33

1 Gym 0.33

2 Garden 0.17

3 Italian Restaurant 0.17

4 Wine Bar 0.00

----Regent's Park----

venue freq

0 Café 0.15

1 Cricket Ground 0.15

2 French Restaurant 0.07

3 Deli / Bodega 0.04

4 Hotel 0.04

----St James's----

venue freq

0 Theater 0.07

1 Hotel 0.07

2 Monument / Landmark 0.04

3 Burger Joint 0.04

4 Plaza 0.04

----Tachbrook----

venue freq

0 Hotel 0.12

1 Pizza Place 0.09

2 Garden 0.06

3 Park 0.06

4 Pub 0.06

----Vincent Square----

venue freq

0 Sandwich Place 0.12

1 Pub 0.09

2 Café 0.07

3 Hotel 0.05

4 Coffee Shop 0.05

----Warwick----

venue freq

0 Hotel 0.10

1 Pub 0.08

2 Italian Restaurant 0.07

3 Coffee Shop 0.05

4 Café 0.05

----West End----

venue freq

0 Clothing Store 0.12

1 Boutique 0.05

2 Cosmetics Shop 0.04

3 Indian Restaurant 0.03

4 Coffee Shop 0.03

----Westbourne----

venue freq

0 Pub 0.25

1 Gym / Fitness Center 0.12

2 Bus Station 0.06

3 Convenience Store 0.06

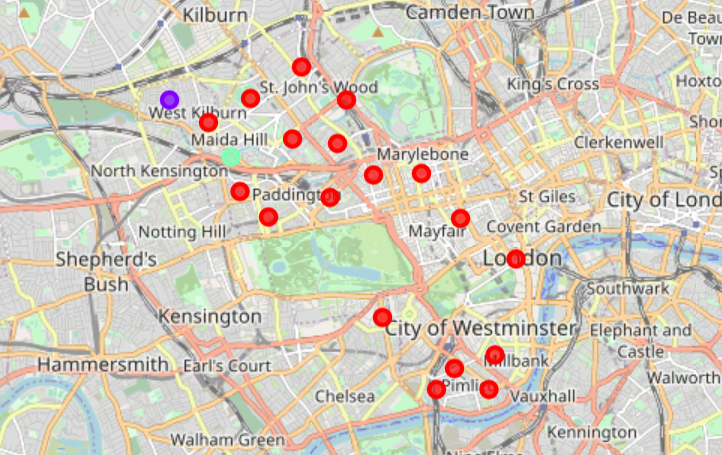
4 Gastropub 0.06

## Modeling

Using K-mean to clustering neighborhood due to theirs similarities then visualize the resulting clusters.



Plot clusters in map



# Results

To find the places for hotel, I find the neighborhood with no hotel in top 5 categories. The following are candidate places.

----Westbourne----

venue freq

0 Pub 0.25

1 Gym / Fitness Center 0.12

2 Bus Station 0.06

3 Convenience Store 0.06

4 Gastropub 0.06

----Church Street----

venue freq

0 Coffee Shop 0.09

1 Theater 0.06

2 Grocery Store 0.06

3 Sandwich Place 0.06

4 Yoga Studio 0.03

----Harrow Road----

venue freq

0 Grocery Store 0.11

1 Coffee Shop 0.09

2 Park 0.09

3 Café 0.06

4 Indian Restaurant 0.06

----Little Venice----

venue freq

0 Pub 0.15

1 Café 0.13

2 Coffee Shop 0.05

3 Lebanese Restaurant 0.05

4 Pizza Place 0.03

----Maida Vale----

venue freq

0 Deli / Bodega 0.12

1 Thai Restaurant 0.12

2 Coffee Shop 0.12

3 Japanese Restaurant 0.06

4 Wine Shop 0.06

----West End----

venue freq

0 Clothing Store 0.12

1 Boutique 0.05

2 Cosmetics Shop 0.04

3 Indian Restaurant 0.03

4 Coffee Shop 0.03

# Discussion

Base on characteristics of each neighborhood we can detail what type of guests the hotel will focus. For example, Maida Vale is suitable to attract Asian guest with many Thai and Japanese restaurants, while West End is suitable for guests with shopping favorite.

# Conclusion

This project illustrates the data science approach to support a systematic analysis of neighborhoods with the help of FourSquare API data sources and analysis methods. This analysis will help Accomp Corp to choose the right location for future hotels in the London City and surrounding areas based on venue categories around a location.

Thank you!.