Capstone Program
Wendao Lu

**Introduction**
London is the capital of the United Kingdom, with the largest municipal population in the European Union. It has a diverse range of people and cultures, and more than 300 languages are spoken in the region. Most of Greater London south of the Thames belongs to the historic county of Surrey, while most of Greater London north of the Thames belongs historically to the county of Middlesex. Area Greater London, 607 square miles (1,572 square km). Pop. (2001) Greater London, 7,172,091; (2011 prelim.) Greater London, 8,173,941.
Its estimated mid-2016 municipal population (corresponding to Greater London) was 8,787,892, the most populous of any city in the European Union and accounting for 13.4% of the UK population. London's urban area is the second most populous in the EU, after Paris, with 9,787,426 inhabitants at the 2011 census. The population density is 14,500/sq mi.
London is a city with a high population and population density. As from Real Estate investor point of view we want to invest in such places were the housing prices are low and the facilities (shops, restaurants, parks, Hotels, etc.) and social venues are nearby. Keeping above things in mind it is very difficult for an individual to find such place in such big city and gather this much information. In the Capstone Project, I want to create a map and information chart where the real estate index is placed on London and each district is clustered according to the venue density.

**Data Sources**
List of areas of London with its boroughs and postcodes
https://en.wikipedia.org/wiki/List_of_areas_of_London
housing prices with postcodes
https://propertydata.co.uk/cities/london
Forsquare API
https://developer.foursquare.com/

**Data Process**
First of all, the data scraped from Wikipedia has to be clean.
I removed all the hyperlinks and there are more than one Postal code for some Locations so I kept only one Postal code. Firstly, I removed all null values and then get rid of unwanted columns and only kept 'Area' and 'Avg price' columns. Then 'Avg Price' columns contain string so I processed it to make integer by removing pound sign and comma.
After cleaning two tables I performed inner join and merge two table and from resulting table I dropped 'Dial Code' and 'OS grid ref' columns as they were of no use. Then by using geocoder library I find the Longitudes and Latitudes of the Location and add a column of each in my dataframe.
I utilized the Foursquare API to explore the boroughs and segment them. I designed the limit as 100 venue and the radius 1400 meter for each borough from their given latitude and longitude information. Here is a head of the list Venues name, category, latitude and longitude information from Foursquare API. Finally, by using the Foursquare API in conjunction with the

created datasets, a table of most common visited venues in London neighborhoods is generated.

**Outcomes**
I came to the result that the house prices in the downtown and with Hotels and Social venues nearby are very high you can clearly visualize in the map above while in the suburbs and the neighborhoods away from the city center have low prices but the facilities are also good. Almost all low-price neighborhoods are close to restaurants, pubs, sports facilities etc. Some Boroughs such as Westminster, and Kensington and Chelsea have very high house prices. Bexley, Croydon, and Sutton Boroughs have very low house prices but have good venues to visit nearby.

**Conclusion**
As people are turning to big cities to start a business or work. For this reason, people can easily interpret where to live with all facilities and cheaply. Not only for investors but also city managers can manage the city more regularly by using similar data analysis types or platforms.