## Capstone Project - The Battle of Neighborhoods

**Author: Sahand Azad**

# Problem Statement

London is a lucrative market for real-estate and has been an investment heaven to both local and foreign property investors. Due to this reality buying a home in London is quite challenging and requires lots of research. Thankfully by using Land Registry data and some data science magic we can create a simple street recommender for specific price points and help potential clients make inform decision.

**Target audience:**

Potential clients looks to buy suitable property in London but are skeptical due to lack of knowledge and volatile market conditions.

**Stakeholders:**

1. London municipality
2. Sellers
3. Buyers
4. Real estate agents

# Data Section

Following sources of data are used while executing the Capstone Project: -

* **Data title: -**

Open Data published by Government of UK under the section **HM Land Registry: Price Paid Data**

**Type of data: -**

Dataset in form of CSV file

**Duration: -**

August 2018 data

**Description of the dataset: -**

Price Paid Data includes information on all property sales in England and Wales that are sold for full market value and are lodged with us for registration.

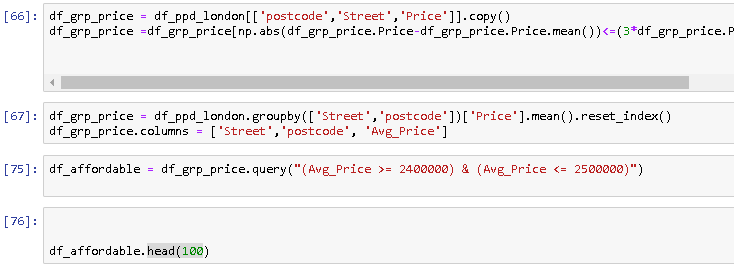
The dataset includes the transactions received at HM Land Registry in the period from the first to the last day of August 2018.

**Methodology:**

Price Paid Dataset is a list of sold properties in England provided by HM Land Registry for free and updated on monthly basis. The current data set used is from the year 2018.

The script used in this project does the following:

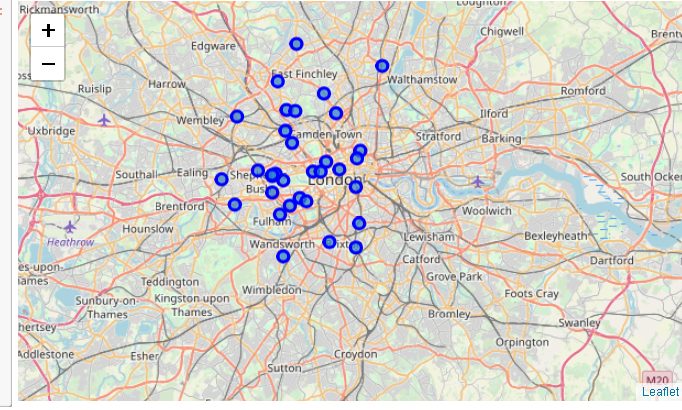
* Parse the data from HM Land Registry dataset for year 2018
* Data set is cleansed and the location is condensed by selecting for only city of London
* Street names with only single real-estate transaction are filtered out
* Prices farther than 3 standard deviation are filtered out
* Location coordinate (latitude, longitude) of these street names are fetched by making API calls to Google Maps.
* The properties are grouped by street and average price
* Depending on the client budget the transcript recommendations are plotted on the map of London. The locations are labelled with the respective street names and their average property price.
* The recommended locations are fed into the Foursquare to explore the neighborhood and find venues and their close proximity to the recommended location.
* Important facilities such as Hospitals and grocery stores, school are searched in vicinity of each location and then reported in a tabular format.



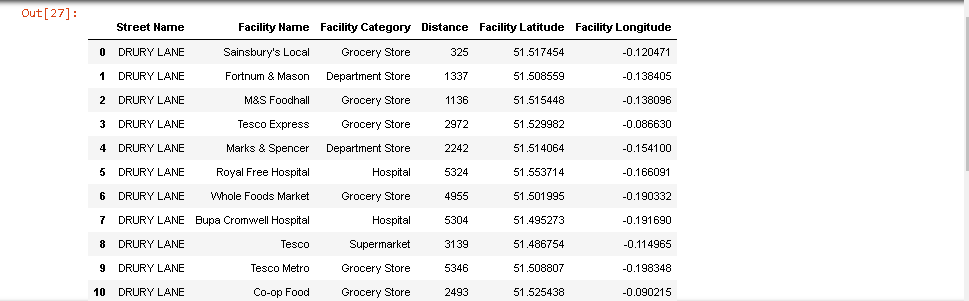
**Results**

Using exploratory data analysis for the city of London the machine learning algorithm recommends 76 streets for houses in price range of 2.4 to 2.5 million pound.



****

Facilities close to each street in the selected price range are listed with their relative distance which can be an important factor for a future buyer.

****

**Discussion:**

An alternative way to approach this problem is to use a multi regression model to predict prices and market trends in the future and advice buyers on whether current market conditions favors buyers or seller and what are price projection in the next few months.

The current model has many short comings as it only indicates which streets has houses sold in selected price range and it does not inform the buyer whether facilities are in approximity to the street or the property they want. The reason for this short coming is sparsity and quality of the free data, I would have created a more detailed model with more predictive elements however the data provided by HMS does not allow that.

For example, if the data had floor square footage, number bedrooms, and house condition of each property listed I could have run a multi regression model to learn and approximate the market behavior and give better suggestion to the potential buyer.

**Conclusion:**

Buyers who look to invest in real-estate or buy a home for their family may have different desires for the property. Maybe to a family approximate to schools or school catchments, or other facilities such as hospitals, daycares are the highest priority. While proximity to the financial district or having views are maybe more important to a real-estate investor.

This tool is not complete as the data is lacking. In its current form its capable of indicating which streets have house in price range of a potential buyer and can show different facilities and their distance to that street. However a more robust predictive model with more details is definitely needed. I will try to upgrade this farther if I find a new source of free data.