**Capstone Project - The Battle of the Neighborhoods (Week 1)**

**Business problem:**

A potential client, who is aspiring to buy a suitable property would like to become knowledgeable about the ongoing pricing to make a conscious decision. Further, he/ she would like to consider several factors like proximity to schools, medical care, restaurants to accommodate his/ her familial needs.

People of all kinds from around the world flock to London, UK with some of them aspiring to make this remarkable place a home of their own. With varying budgets and needs, people find it very hard to find a suitable place and neighborhood to accommodate them and their families. Due to high cost of living and other multiple issues, London housing has been struggling. With government provided authentic data on London properties coupled with data science techniques, one can make derive the useful information about current pricing in different localities of London while considering other factors of his choice. This would help the potential client to make an informed decision about buying a suitable property.

Target audience: Potential clients looking to buy a suitable property in London but are skeptical due to lack of knowledge and volatile market conditions.

**Data:**

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

**Data title** Price Paid Data (Open source data published by Government of UK under the section HM Land Registry)

**Data type:** 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.

**Source:** <http://landregistry.data.gov.uk/>

**Data title:** Google Maps Geocoding API

**Data type:** JSON

**Description of the data:** Location coordinates obtained by G-maps API calls.

Location Information obtained from Price Paid Dataset is used to obtain the location coordinates from Google Maps. A separate Python script has been developed to extract the unique street names, district names from the Price Paid Dataset and embed those in the G-Maps API calls to obtain the required information.

**Source:** Google Cloud Platform/ Google Maps

**Data title:** Foursquare location data

**Data type:** JSON

**Description of the data:** Location coordinates obtained by Foursquare API calls.

To determine the proximity of various amenities as per the client’s requirement, Foursquare location data is used.

**Source:** <https://foursquare.com/>