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

## **Real Estate Analysis: Safest Neighborhood in London Borough**

#### **Author: Jeetendra Swami**

#### **Date: January 07, 2020**

### **1. Introduction**

#### **1.1 Background**

Moving When you have a family can be Hectic, especially when you’re moving to a city as diverse as London . So as part of a change, when I was considering setting up my new home in London, even though it was exciting, it was not an easy task. Currently settled in a peaceful locality in the suburbs of Michigan, the bustling life and current political scenario of London was daunting to say the least. Upon some research, I could find that the slump in London’s housing market may be about to end. And this is the right time to invest in a property.

However, before making an investment, I decided to do a little research on the best neighborhood.

#### **1.2 Problem**

Data that might contribute to determining the best neighborhood include how safe the neighborhood is, affordability, accessibility and facilities like parks, restaurants etc. This project aims to predict the best neighborhood based on the above facts.

#### **1.3 Interest**

Obviously, expats who are considering relocating to London as part of career change could utilise this report for finding a safe borough and district in London for buying/renting a house. People who are currently in London, but have not invested in a house yet due to the current political/economical scenario, could also use this report.

### **2. Data acquisition and cleaning**

#### **2.1 Data sources**

The first step in the analysis was to find all the boroughs in London. This data could be obtained from a wiki page, <https://en.wikipedia.org/wiki/List_of_London_boroughs>.

The latest crime data records could not be obtained for the boroughs. However, the data pertaining to 2016, is there as a kaggle dataset, <https://www.kaggle.com/jboysen/london-crime>.

Next, we can check the affordability of the houses using a dataset from Office for National Statistics. <https://webarchive.nationalarchives.gov.uk/20171102125110/https://visual.ons.gov.uk/wp-content/uploads/2017/10/map.csv>'

During the research, I also came across a survey by itv, which records the happiness index of the people in the uk. The link for the report is <https://www.itv.com/news/london/2016-08-16/londons-happiest-boroughs-revealed-as-richmond-hits-the-top-spot/>. The data from this web page can be scrapped and the best boroughs can be found out.

Finally to find the districts in Sutton, which is the best borough for buying a house, we scarp the wiki page: <https://en.wikipedia.org/wiki/London_Borough_of_Sutton#Districts>

#### 

#### **2.2 Data Cleaning**

**To find the London borough names**, we use **BeautifulSoup** to scrap the wikipage. After that string manipulation is done using regular expression and the exact name of the boroughs are extracted.

**For the crime rate analysis**, the data set from Kaggle is downloaded. The dataset consist of

1. lsoa\_code: code for Lower Super Output Area in Greater London
2. borough: Common name for London borough.
3. major\_category: High level categorization of crime
4. minor\_category: Low level categorization of crime within major category.
5. value: monthly reported count of categorical crime in given borough
6. year: Year of reported counts, 2008-2016
7. month: Month of reported counts, 1-12

The csv file is read using pandas. Only the data from the year 2016 is extracted. If the 'value' column is 0, then is filtered out. Finally the dataset can be grouped together on borough name to get the count/crime rate.

**For the analysis of price per sq.m of houses in the boroughs**, the dataset from <https://www.ons.gov.uk/> is extracted. This contains:

1. local authority code
2. local authority name
3. year
4. price per m2 Only local authority name which was extracted from the wiki page is filtered and taken into a dataset. Next, the local authority code and year can be dropped. This will give the dataset with the borough name and price/m2. The dataset when sorted in the ascending order will give the most affordable areas.

**For the happiness index**, the itv page which contains the happiness index analysis is scrapped and the table with the values are extracted. Next string manipulation is done to get only the borough names in the decreasing order of happiness value. These datas are also converted into a dataframe

From the analysis, we conclude the best borough for real estate in London to be Sutton. Next we need to build a dataset with the neighborhoods in Sutton. For that, the wiki page <https://en.wikipedia.org/wiki/London_Borough_of_Sutton#Districts> is scrapped and the district names are obtained. Borough name will be Sutton. **The geographical cordinated are obtained using geopy client**.

Finally **Using Foursquare Location Data**, the 100 most popular venues in a radius of 500m for each district is Sutton is obtained. The data obtained is a JSON file, and we need to turn that into a data-frame. This final dataset will contain:

1. District
2. District Latitude
3. District Longitude
4. Venue
5. Venue Latitude
6. Venue Longitude
7. Venue Category

In [ ]: