Project Report

London Property Recommendation

1. Business Problem

People of all kinds from around the world flock to London, UK with some of these 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 the inevitable Brexit, the problem has further compounded. A potential client aspiring to buy a suitable property would like to become knowledgeable about the ongoing pricing to make a concious decision. Further, he/ she would like to consider several factors like proximity to schools, medical care, restaurants to accommodate his/ her familial needs.

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

Stakeholders: -

1. Government of UK

2. Sellers

3. Buyers

4. Real estate agents

2. 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.

This dataset was downloaded and later hosted on https://labs.cognitiveclass.ai/ for ease of use.

Source: -

http://landregistry.data.gov.uk/

Data title: -

Google Maps Geocoding API

Type of data: -

JSON

Duration: -

N/A

Description of the data: -

Location coordinates obtained by Gmaps 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 GMaps API calls to obtain the required information.

Source: -

Google Cloud Platform/ Google Maps

Data title: -

Foursquare location data

Type of data: -

JSON

Duration: -

N/A

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/

3. Methodology

Price Paid Dataset contains the sale prices of properties in England and Wales submitted to HM Land Registry for registration. This is an open dataset which is hosted on http://landregistry.data.gov.uk/. This data is updated monthly and is made available from 1995.

This project is focused on investigating the most recent market prices of Property in the city of London and recommend various locations where the prospective client can buy a property based upon his/ her budget.

The automated script developed as a part of this project does the following: -

1) Parse the necessary data from the price paid dataset which includes the transactions received at HM Land Registry in the period from the first to the last day of August 2018.
2) The data is cleansed and any data of sales agreements which predates 2016 is deleted from the dataset.
3) The data is further condensed by selecting it only for the city of London which is area of choice in this project.
4) Unique "Street names" in the city of London where recent transactions for sale of property were done are filtered from the dataset.
5) Location coordinates (latitude, longitude) of these street names are fetched by making API calls to Google Maps. A separate one-time Python script was developed to fetch this data and store it in a CSV file.
6) The average price of property on each of these streets is determined by taking a mean on recent transactions of sale of property on respective streets.
7) Based upon the budget of the client, the current average prices are compared and all recommendations for the locations are made by plotting them on map of London. The locations popups are labelled with the respective street names and their average property price.
8) The recommended locations are further fed into Foursquare API calls to determine various venues in proximity to them. All reported venues are then tabulated and presented to the user.
9) Important facilities like Hospitals, Grocery stores, Elementary schools, High Schools are searched in vicinity of each location and then reported in a tabular form to the user.
10) In order to conduct a similar analysis for any other city in UK or Wales, the automated script has been written to accommodate a change in:
    - City/ Town
    - Budget of the client

Such changes can be made with minimal effort and would generate the recommended locations to buy a property in the city of choice.

4. Results

Upon running the exploratory data analysis for city of London with a hypothetical budget of GBP 2.2 Million – GBP 2.5 Million, the machine learning algorithm recommends 39 streets in London where the prospective client can choose to buy the property as per the current market prices.

A list of such locations is presented to the user with location coordinates and most recent average prices.

Street	Latitude	Longitude	Avg_Price
DULWICH WOOD AVENUE	51.425586	-0.082416	2.297000e+06
SOUTH HILL PARK	51.557134	-0.164343	2.466667e+06
TEIGNMOUTH ROAD	51.550139	-0.214496	2.295000e+06
BURNSALL STREET	51.489042	-0.166883	2.286500e+06
FORDWYCH ROAD	51.551511	-0.206736	2.290000e+06
PORTEN ROAD	51.498603	-0.214120	2.200000e+06
ALBERT BRIDGE ROAD	51.477861	-0.164743	2.383333e+06
EDITH VILLAS	51.491665	-0.206556	2.402500e+06
WESTBOURNE GROVE	51.514797	-0.197071	2.300000e+06
LADBROKE ROAD	51.508776	-0.203410	2.261250e+06

These recommended street names are plotted on the map of London with the average market prices.
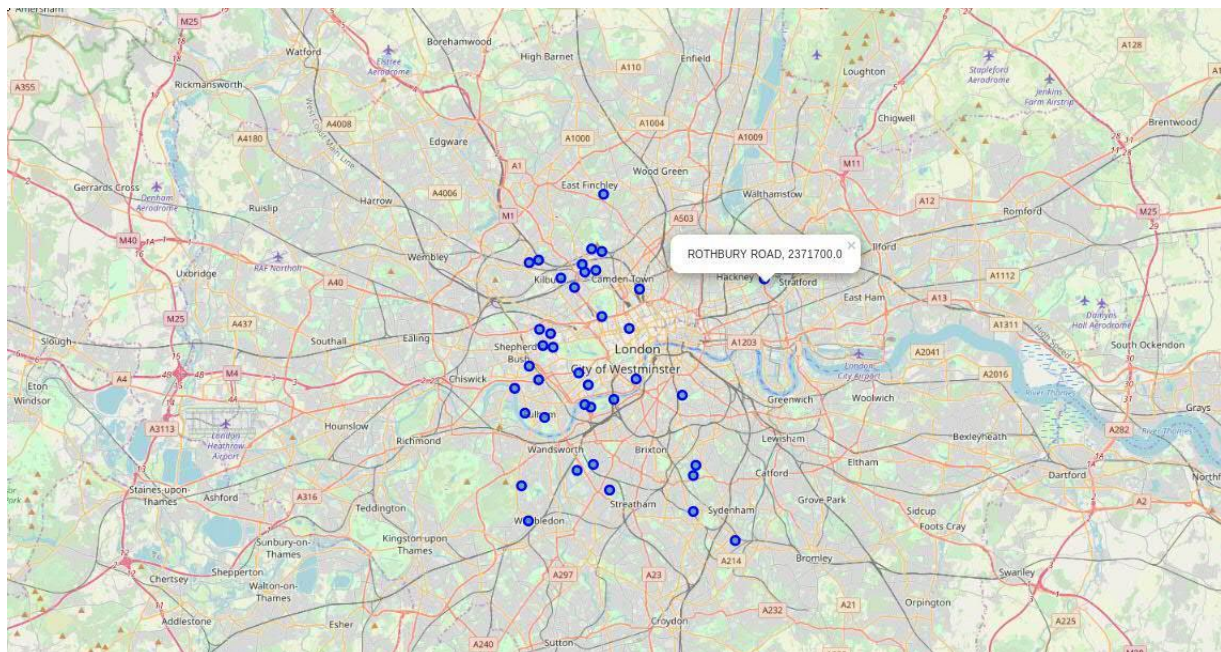
Further, following venues are enlisted for the user to make an informed decision while choosing a location.

	Street	Street Latitude	Street Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	DULWICH WOOD AVENUE	51.425586	-0.082416	Beer Rebellion	51.424580	-0.083425	Bar
1	DULWICH WOOD AVENUE	51.425586	-0.082416	The Indian Dining Club	51.427795	-0.086488	Indian Restaurant
2	DULWICH WOOD AVENUE	51.425586	-0.082416	The Paxton	51.427880	-0.086168	Pub
3	DULWICH WOOD AVENUE	51.425586	-0.082416	Gipsy Hill Railway Station (GIP)	51.424530	-0.083959	Train Station
4	DULWICH WOOD AVENUE	51.425586	-0.082416	Brown & Green @ The Station	51.424425	-0.083836	Breakfast Spot
5	DULWICH WOOD AVENUE	51.425586	-0.082416	Manuel's Restaurant and Bar	51.427591	-0.086131	Italian Restaurant
6	SOUTH HILL PARK	51.557134	-0.164343	Daunt Books Hampstead	51.555513	-0.166277	Bookstore
7	SOUTH HILL PARK	51.557134	-0.164343	Hampstead Heath Ponds	51.559300	-0.165973	Lake
8	SOUTH HILL PARK	51.557134	-0.164343	Hampstead Heath	51.559622	-0.164921	Park
9	SOUTH HILL PARK	51.557134	-0.164343	Paradise	51.555476	-0.166312	Indian Restaurant
10	SOUTH HILL PARK	51.557134	-0.164343	Keats House	51.555745	-0.167975	Museum
11	SOUTH HILL PARK	51.557134	-0.164343	karma bread	51.554494	-0.165586	Bakery
12	SOUTH HILL PARK	51.557134	-0.164343	The Garden Gate	51.554733	-0.165697	Pub
13	SOUTH HILL PARK	51.557134	-0.164343	The Little Thai	51.554115	-0.164737	Thai Restaurant
14	SOUTH HILL PARK	51.557134	-0.164343	Parliament Hill	51.559661	-0.159639	Scenic Lookout
15	SOUTH HILL PARK	51.557134	-0.164343	Silverberry Deli & Kitchen	51.554174	-0.164975	Cafe
16	SOUTH HILL PARK	51.557134	-0.164343	The Freemasons Arms	51.556968	-0.168606	Pub
17	SOUTH HILL PARK	51.557134	-0.164343	Zara	51.554423	-0.165561	Greek Restaurant
18	SOUTH HILL PARK	51.557134	-0.164343	The Stag	51.553420	-0.161576	Gastropub
19	SOUTH HILL PARK	51.557134	-0.164343	M&S Simply Food	51.553956	-0.165119	Grocery Store
20	SOUTH HILL PARK	51.557134	-0.164343	Mimmo la Bufala	51.555340	-0.166230	Italian Restaurant

5. Discussion

Based upon the findings in the results section, the user can take a conscious decision about choosing a street/ location based upon his/ her requirements.

The results section enlists 39 places where a prospective client can buy a property based upon his needs and choices. Such choices would be affected by the venues and facilities which are close to the property which match against his familial needs.

Few possible cases are: -

1. A prospective client with elders in the family would be inclined to choose a location where hospitals and grocery stores are located in close proximity.

2. A prospective client with kids in the family would choose a location where elementary and high schools are close-by. He would also like to choose a place with parks and other venues to accommodate his family are in the close vicinity.

3. A bachelor would be inclined to choose a property which has pubs, bars, entertainment places close to the property.

6. Conclusion

The decision of a buyer is influenced by the familial needs, personal biases. So, based upon the findings summarized in the results and discussion sections, following conclusions can be made: -

1. While making recommendations to a prospective client, it is imperative to know his/ her immediate needs and requirements besides the budget. This would help to catch his/ her attention.

2. Knowledge about the most recent market prices can be very helpful for the client and can help him take a decision.