VINDHYA R

vindhyamvr@gmail.com

Abstract

A client wants to buy a property in England and asked our help , we have analyzed ther neighborhoods in England and given them suggestion

Neighborhood Analysis Of England

**Neighborhood Analysis Of England**

***Introduction***

* 1. Problem Description

The housing market in the UK has always been a topic of national attention, with news sites dedicating sections that report on key news that can potentially affect housing prices and the trends in recent months. Not only are the trends in housing market of concern to buyers and owners, they reflect the current economic situation and social sentiments in the country. For many people, buying a property is one of the most important decision and purchase in life. Besides the affordability of a house, other factors such as the desirability of the location and the long-term investment prospects also affect the decision-making process. However, many reports on housing prices are averaged across regions in the UK, where England is considered one region. England homebuyers who want to find out more about housing trends in various districts would therefore have little recourse. For example, housing prices in Barking & Dagenham are likely to differ from those in Kensington & Chelsea, but it is difficult to find sources that provide such information. If there were location-specific predictions on housing prices within England, making informed decisions over which areas to consider can be greatly facilitated.

In the last decade, the housing market in England has been rapidly growing, with average housing prices increasing by more than 10% yearly on most years . Together with stronger economic growth and increasing price expectations, this presents good news to current homeowners and potential Besides overall trends, the price of a house in England can vary greatly depending on the location. Properties in London fetch much higher prices compared to properties in other places.

1.2 Business Problem

My client, a resident in UK is looking to settle in England and wants to buy a property. They want a green area, that has good schools, easy access to hospitals and is children friendly.

Since the England demography is so big, my client needs deeper insight from available data in other to decide where to buy a property. My client is asking a help from us to analyse demography of England and give some insight about the Neighbourhoods.

1.3 Target Audience

Considering the diversity of UK, there is a high multicultural sense. England is a place where different shades live. As such, in the search for an property, there is a high demand. The target audience is broad, it ranges from people of England, and others who wants to settle in England.

## ***Data***

2.1 Description of Data

This project will rely on public data from UK government dataset and Foursquare.

#### 2.2 Dataset:

Within the England Area, there are lot of Boroughs. The focus of this project will be the neighbourhoods. The England Area consists of 380 Boroughs.

We extract venues from all neighbourhoods within a certain radius and pick top 5 venues in each neighbourhood and analyse which neighbourhoods are good for property investment depending on client’s requirements.

2.3 Cleaning Data

1) I have downloaded the data from the UK government website. The file is a json format. I have retrieved the data from the file and converted into a data-frame. The data obtained are shown below:

![A screenshot of a cell phone

Description automatically generated]()

Fig 1. Dataframe obtained from json file

2) In this step I have removed the unwanted columns/attributes from the data-frame for ease of analysis. I have renamed few other columns / attributes to give some meaningful names according to the context. The snapshot of resultant data-frame is shown below:

![A screenshot of a cell phone

Description automatically generated]()

Fig 2: Data frame after cleaning

3) Check the data frame for missing values in the data retrieved. If there are any missing values, we will fill it with either mean of all values or with the previous values. When I checked, there was no missing data. The result shown below.

![A screenshot of a cell phone

Description automatically generated]()

Fig 3) Count of missing values

1. ***Methodology***

3.1 Getting venues from all neighbourhoods

As said before my client wants to buy a property and settle in England. I have analysed the neighbourhood of England. First, have extracted the latitude and longitude of England. The values are shown below:

![A picture containing text, map

Description automatically generated]()

Fig 4. Plot of England

Extract the first neighbourhood city in England from the data frame and get the geographical coordinates.

The url for foursquare has been generated and the limit for the number of venues to be retrieved and radius within which the venues are to be searched are specified.

There are 28 venues in Hartelpool that are retrieved within radius of 5000. All venues along with the geographical coordinates and the category of the venue are put into a dataframe as shown below in fig 5

![A screenshot of a cell phone

Description automatically generated]()

A function has been created to get the nearby locations from all the neighbourhoods in the data-frame extracted from json file. Once all venues are extracted, they are stored in the data frame along with their latitude, longitude and category of the venue as shown in fig 6 below

![A screenshot of a cell phone

Description automatically generated]()

Fig 6: Venues in each neighborhood

The count of venues extracted are counted in each neighbourhood and extract unique categories of venue that may help my client to decide the city to settle in as shown below.

![A screenshot of a cell phone

Description automatically generated]()

Fig 7: Count of venues in each neighbourhood

Analyse venue in each neighbourhood and extract top 10 venues in each neighbourhood.

3.2 Modelling

We use K-means clustering to analyse the neighbourhoods.

K-means clustering is one of the simplest and popular unsupervised machine learning algorithms.

Typically, unsupervised algorithms make inferences from datasets using only input vectors without referring to known, or labelled, outcomes.

“the objective of K-means is simple: group similar data points together and discover underlying patterns. To achieve this objective, K-means looks for a fixed number (k) of clusters in a dataset.”

A cluster refers to a collection of data points aggregated together because of certain similarities.

You’ll define a target number k, which refers to the number of centroids you need in the dataset. A centroid is the imaginary or real location representing the centre of the cluster.

Every data point is allocated to each of the clusters through reducing the in-cluster sum of squares.

In other words, the K-means algorithm identifies k number of centroids, and then allocates every data point to the nearest cluster, while keeping the centroids as small as possible.

The ‘means’ in the K-means refers to averaging of the data; that is, finding the centroid.

We have used this algorithm to cluster the neighbourhood. The number of clusters used are 5. Advantage of using this algorithm is it is fast and efficient in terms of computational cost, is highly flexible to account for mutations in real estate market in England and is accurate.

The results are shown below.

![A screenshot of a cell phone

Description automatically generated]()

Fig 8: Merged table after clusters are formed

![A picture containing text, map

Description automatically generated]()

Fig 9: The map plotted after clustering has been done is shown below:

The neighbourhoods are all divided into 5 different clusters and analysed.

1. ***Results***

The neighbourhood clustering results as 5 different clusters

First Cluster

![A close up of a piece of paper

Description automatically generated]()

Analysis: As we can see here in first cluster there are lot of pubs, coffee shops, restaurants and cafes. This may not be ideal for a family with children like my client. This is good for people who hang out often and enjoy partying.

Second Cluster

![A close up of a piece of paper

Description automatically generated]()

Analysis: Here there are pubs and café but compared to first cluster we see gardens, grocery stores, park, pharmacy, Zoo, bookstores. This is ideal for a family.

Third Cluster

![A close up of a piece of paper

Description automatically generated]()

Analysis: There are lot of markets, Historic places, Train station, park so this is a good choice for a family with children.

Fourth Cluster

![A close up of a piece of paper

Description automatically generated]()

Analysis: As we can see here in first cluster there are lot of pubs, coffee shops, restaurants and cafes. This may not be ideal for a family with children like my client. This is good for people who hang out often and enjoy partying.

Fifth Cluster

![A screenshot of a cell phone

Description automatically generated]()

Analysis: There are family friendly places here. As we see Chesterfield, Doncaster and others have good greenery, supermarkets.

***6 Discussion***

First of all, even though the England Housing Market is always in demand in people who want to settle in United Kingdom.

We may discuss our results under two main perspectives.

First, we may examine them according to neighbourhoods. It is interesting to note that, although Notting Hill, Kensington, Chelsea, Marylebone and might be considered highly profitable venues to purchase a real estate according to amenities and essential facilities surrounding such venues i.e. elementary schools, high schools, hospitals & grocery stores, Basingstoke, Kent, Chesterfield, Doncaster are arising as next future elite venues with a wide range of amenities and facilities. Accordingly, one might target under-priced real estates in these areas of England in order to decide on settling down.

Second, we may analyse our results according to the five clusters we have produced. Even though, all clusters could praise an optimal range of facilities and amenities, we have found two main patterns. The first pattern we are referring to, i.e. Clusters 0, 3 and 4, may target home buyers prone to live with children areas with supermarkets, parks, waterfronts. Instead, the second pattern we are referring to, i.e. Clusters 1 may target individuals who love pubs, theatres and soccer.

As my client is an Asian, I would suggest areas around Birmingham, Reading, Kent where more of Asian population are present with more Asian Restaurants and other cultural amenities.

***7 Conclusion***

UK housing prices is facing several different headwinds, including the prospect of higher taxes and a warning from the Bank of England that U.K. home values could fall as much as 30 percent in the event of a disorderly exit from the European Union. In this scenario, it is urgent to adopt machine learning tools in order to assist homebuyers in England to make wise and effective decisions. As a result, the business problem we were posing was : Helping one of my client to select a place to settle with family and children.

To solve this business problem, we clustered England neighbourhoods in order to recommend venues where homebuyers can make a real estate investment. We recommended profitable venues according to amenities and essential facilities surrounding such venues i.e. grocery stores, parks, restaurants.

First, we gathered data on England neighbourhood from the UK govt website. Moreover, to explore and target recommended locations across different venues according to the presence of amenities and essential facilities, we accessed data through Four Square API interface and arranged them as a data frame for visualization. By merging data on Neighbourhoods and data on amenities and essential facilities surrounding such properties from Four Square API interface, we were able to recommend profitable real estate investments.

Second, The Methodology section comprised four stages:

1. Collect Inspection Data

2. Explore and Understand Data

3. Data preparation and pre-processing

4. Modelling. In the modelling section, we used the k-means clustering technique as it is fast and efficient in terms of computational cost, is highly flexible to account for mutations in real estate market in England and is accurate.

We drew conclusion that England Housing Market is always in demand in people who want to settle in United Kingdom.

We may discuss our results under two main perspectives.

First, we may examine them according to neighbourhoods. It is interesting to note that, although Notting Hill, Kensington, Chelsea, Marylebone and might be considered highly profitable venues to purchase a real estate according to amenities and essential facilities surrounding such venues i.e. elementary schools, high schools, hospitals & grocery stores, Basingstoke, Kent, Chesterfield, Doncaster are arising as next future elite venues with a wide range of amenities and facilities. Accordingly, one might target under-priced real estates in these areas of England in order to decide on settling down.

Second, we may analyze our results according to the five clusters we have produced. Even though, all clusters could praise an optimal range of facilities and amenities, we have found two main patterns. The first pattern we are referring to, i.e. Clusters 1, 2 and 4, may target home buyers prone to live with children areas with supermarkets, parks, waterfronts. Instead, the second pattern we are referring to, i.e. Clusters 0,3 may target individuals who love pubs, theatres and soccer.

As my client is an Asian, I would suggest areas around Birmingham, Reading, Kent where more of Asian population are present with more Asian Restaurants and other cultural amenities.