BATTLE OF NEIGHBORHOODS

The United Kingdom

1. Introduction/Background

Sometimes when people need to relocate, they would like to explore and try to gather as much information as possible about the new place to make a decent choice. It can be the neighborhood, locality, market, price of the place and many more factors including neighborhood analysis. This is can be termed as request for a search algorithm which usually returns the requested features such as population rate, median house price, school ratings, crime rates, weather conditions, recreational facilities etc...

It would be beneficial and convenient to have an application which could make easy by considering a comparative analysis between the neighborhood with provided factors.

This project therefore will help people who are looking for moving to and buying property in the United Kingdom. If they are looking to move to the UK they can see:

- Which district has cheaper house price or,

- They can choose to live in residential or commercial areas and can see for example which residential districts is best.

2. Data Sets and APIs

This project will analyze data sets collected from Office for National Statistic of the UK, website: https://www.ons.gov.uk. Data includes statistics on towns and cities in England and Wale with a focus on housing and deprivation in 2015 (such as median house price, crime rank, living environment rank, ect...). After statistically analyzing and recommending best towns for living based on different standards (cheap town/ safe town/ best living environment...), this project would compare 2 randomly picked neighborhoods and analyses the top 10 most common venues in each of those two neighborhoods based on the number of visits by people in each of those places. Also, this project uses K-mean clustering unsupervised machine learning algorithm to cluster the venues based on the place category such as restaurants, park, coffee shop, gym, clubs etc using Foursquare APIs. This would give a better understanding of the similarities and dissimilarities between the two chosen neighborhoods to retrieve more insights and to conclude with ease which neighborhood wins over other.

Foursquare API

This API has a database of more than 105 million places. This would use Four-square API as its prime data gathering source. Many organizations are using to geo-tag their photos with detailed info about destination, while also serving up contextually relevant locations for those who are searching for a place to eat, drink or explore. This API provides the ability to perform location search, location sharing and details about a business. Foursquare users can also use photos, tips and review in many productive ways to add value to the result.

Work Flow

HTTP requests would be made to this Foursquare API server using name districts of England and Wales to pull the location information (Latitude and Longitude).

Foursquare API search feature would be enable to collect the nearby places of the neighborhoods. Due to http request limitations the number of places per neighborhoods parameter would reasonably be set to 100 and the radius parameter would be set to 500.

Folium - Python visualization library would be used to visualize the neighborhoods cluster distribution of the UK over an interactive leaflet map.

Unsupervised machine learning algorithm K-mean clustering would be applied to form the clusters of different categories of places residing in and around the neighborhoods.

Python packages and Dependencies:

- Pandas - Library for Data Analysis
- NumPy - Library to handle data in a vectorized manner
- JSON - Library to handle JSON files
- Geopy - To retrieve Location Data
- Request - Library to handle http requests
- Matplotlib - Python Plotting Module
- Sklearn - Python machine learning Library
- Folium - Map rendering Library

3. Methodology

The workflow of the project starts with data wrangling. Data file was collected from https://www.ons.gov.uk including 108 districts in the United King (England and Wales) and which followed with respective information about Median House Price and IMD rank (2015). Median House Price reflects the midway point of all the houses/units sold at price over a set period (yearly in this project) while the Index of Multiple Deprivation (IMD rank) is the official measure of relative deprivation for small areas in England (usually from Income, Employment, Education, Health, Crime, Housing barrier and Living environment) . If a small area's rank is closer to 1 than that of another area, it is more deprived.

Second step, we will have a closer look at Median House Price and IMD rank of 108 districts in England and figure out which districts are the most affordable and which districts have higher IMD rank, as well as figure out there is any relationship between Median House Price and IMD rank by plotting scatter plot using Matplotlib.

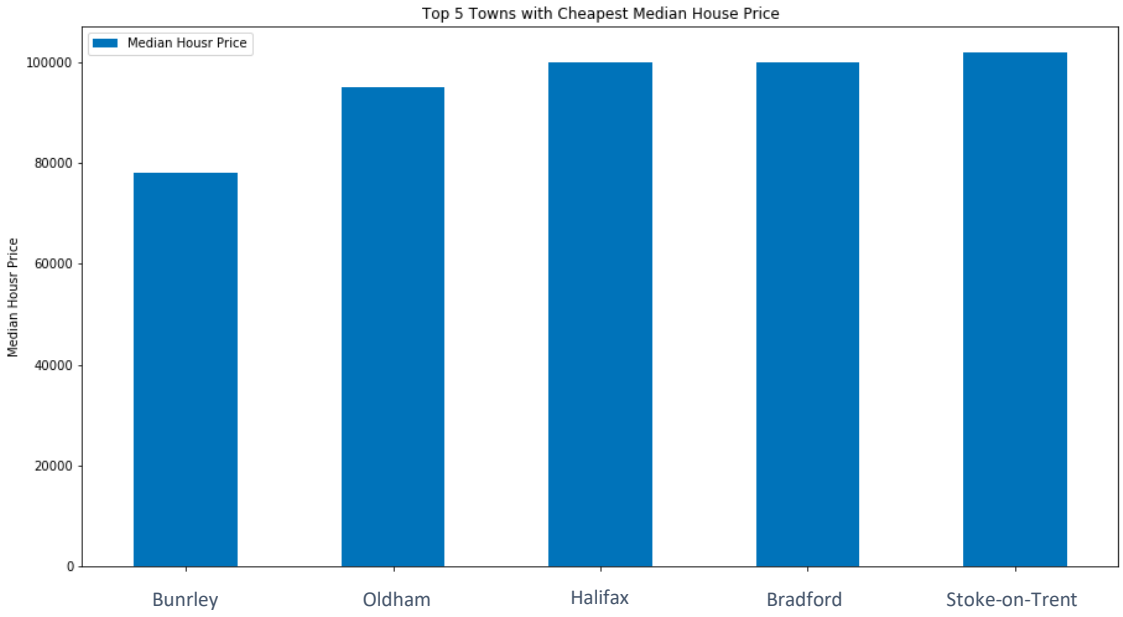
Next we will process to derive the latitude and longitude of those districts, and visualize them on map with the Folium Map

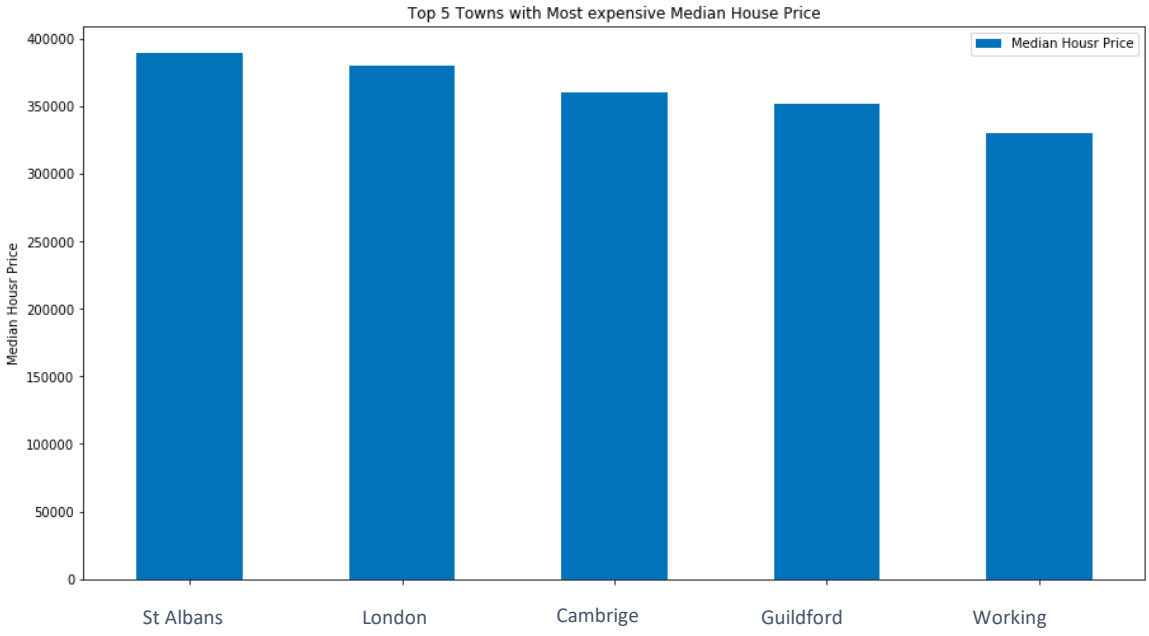
Then, Foursquare API and K-mean clustering methods are used to retrieve the to trend venues of the UK neighborhoods.

Finally, depends on one's budget, expectation and living habit, best towns will be recommended using the results of cluster neighborhoods, Median House Price and IMD rank analysis.

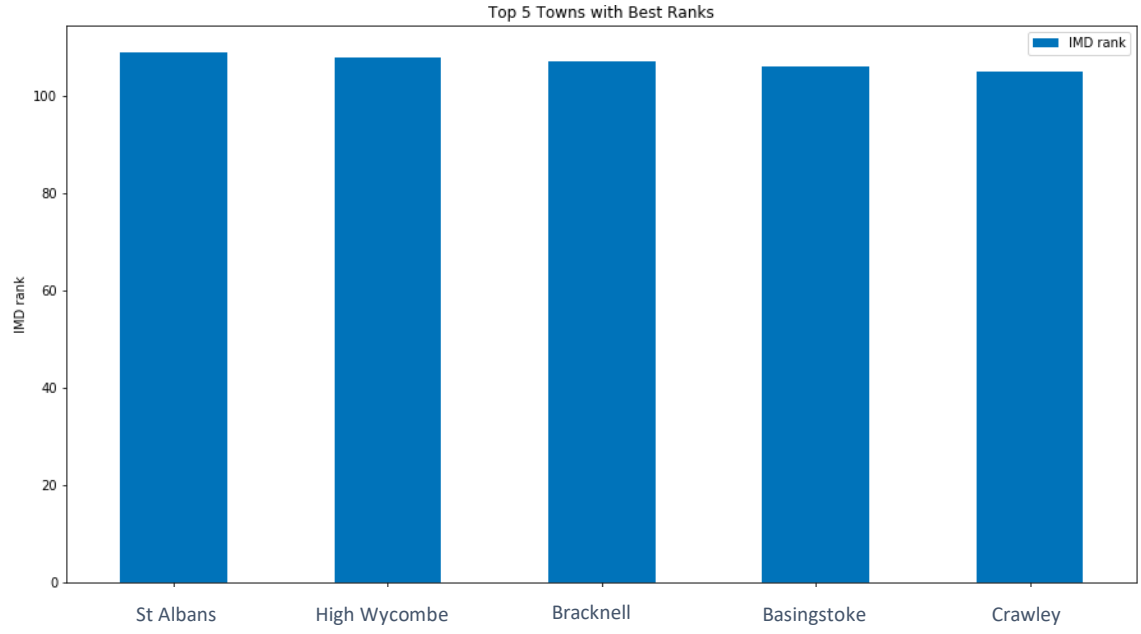
4. Results

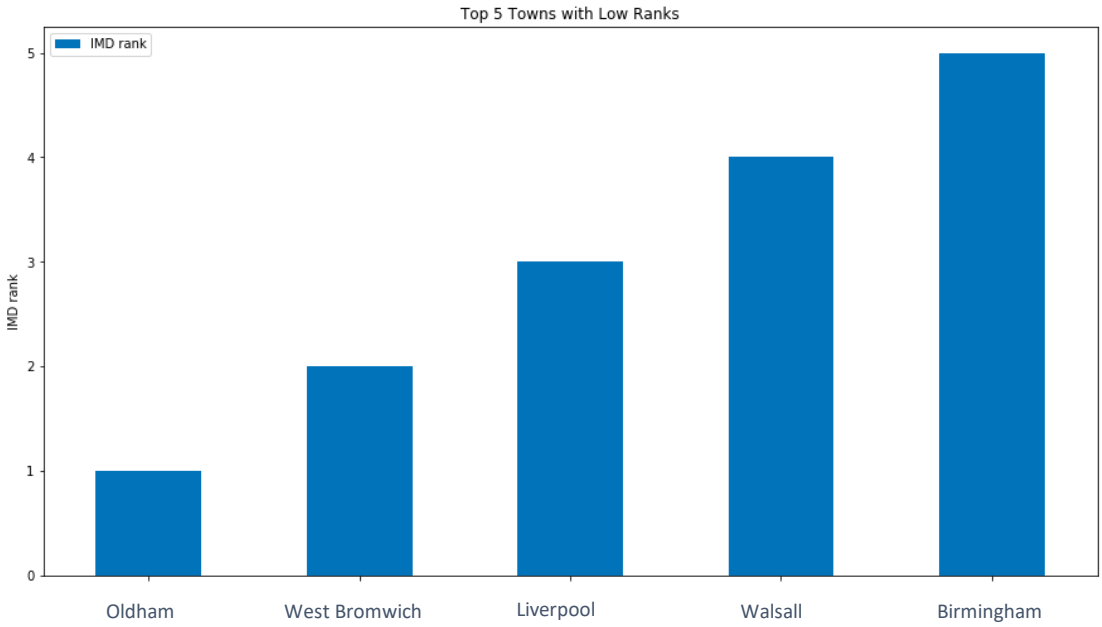
Top 5 Towns with the most affordable house price and top 5 Towns with most expensive house price

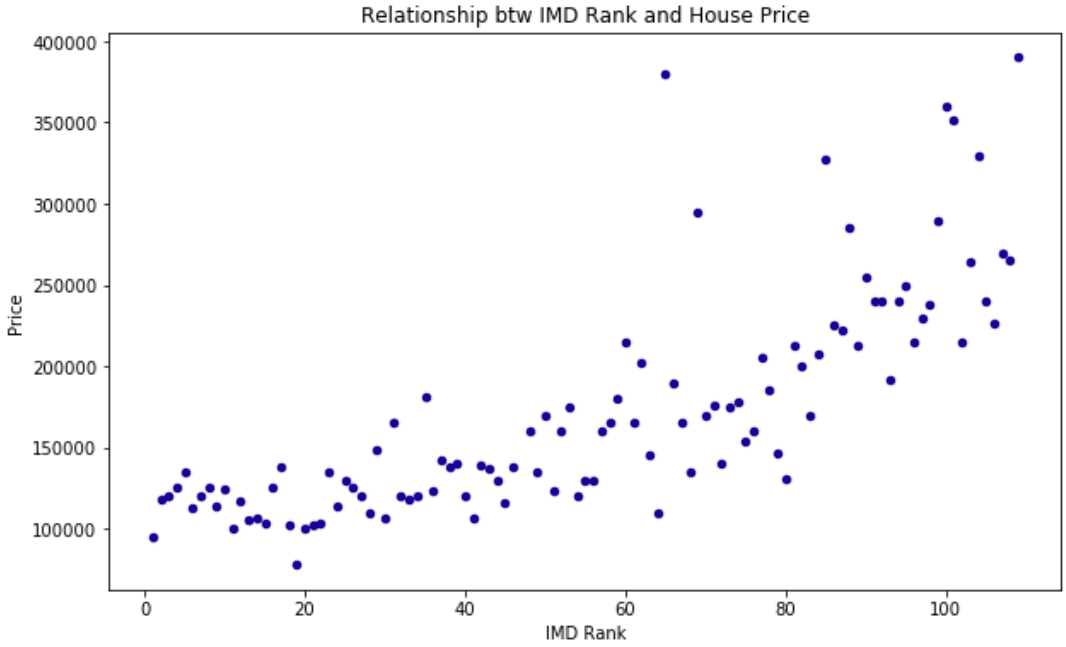




Top 5 Towns with low IMD rank and top 5 Towns with best IMD rank



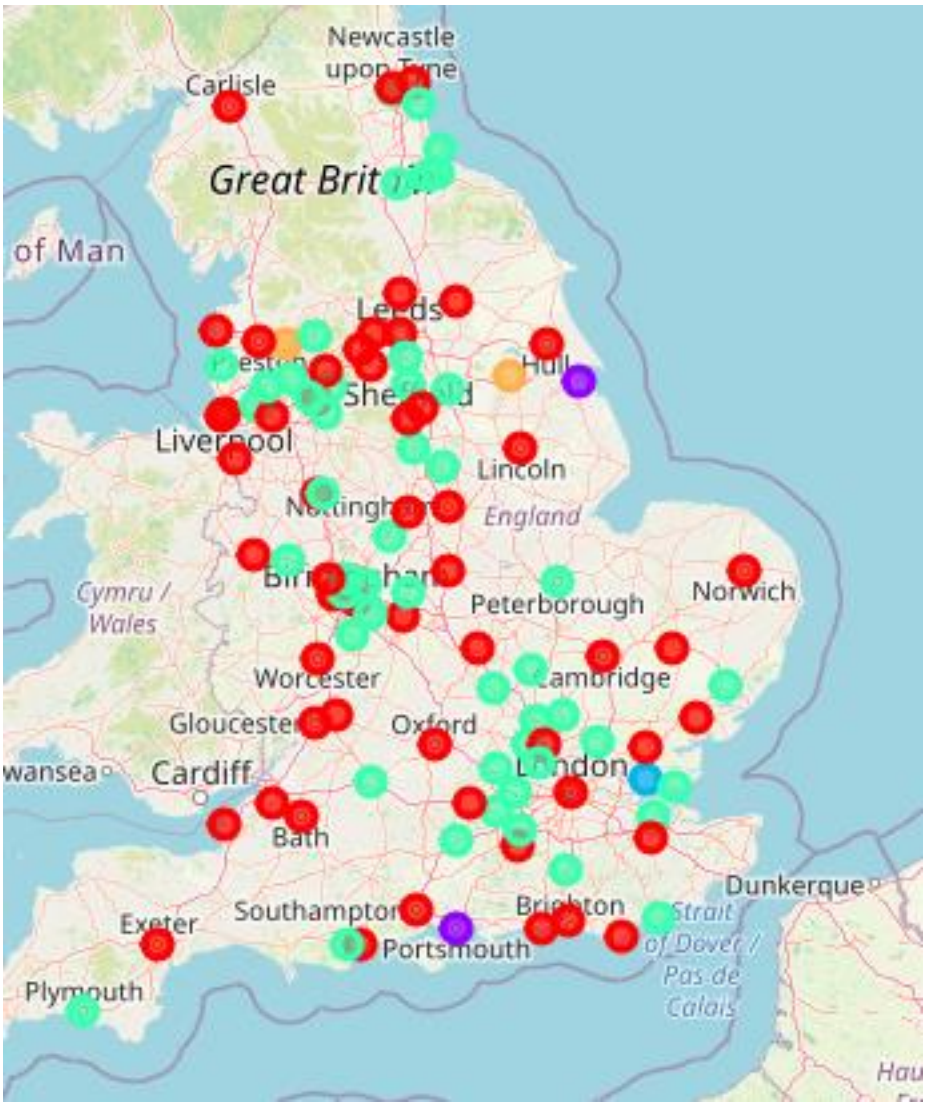




Plotting relationship between Median House Price and IMD rank: scatter plot shows that there's a negative correlation between Median House Price and IMD rank: when a town has lower median house price, it seems that that town is more deprived.

Clustering Neighborhoods

There are 5 clusters with their top 10 common venues as shown in the map below, in which there're 2 main clusters inclusding commericial areas (with top 10 common venues such as pubs, Coffee shops, restaurants, bas, shopping malls…) and local areas which seem to be quieter (with top 10 common venues like supermarkets, bookstores, local restaurants, clothing stores…)



5 Recommendation

Depend on user's budget and personal reference, this project can help them to determine which is the best option for them to settle down. For example, with a budget about 150.000 – 200.000GBP, if One would like to live in a commercial area with IMD is not less than 80, there are two best options for he/she: York and Shrewsbury. York has IMD rank of 93, which is a great rank and median house price falls around 191.150 GBP; with a lot of Pubs, Bars, Cafes, Plazas and Historic sites. Shrewsbury offers a more affordable house price of 169.950GBP, but lower IMD rank of 83, also has a bunch of Pubs, Bars, Cafes and Clothing stores…