Capstone Project

Business Problem

- The objective of this project is to explore Bristol city, segment and cluster its various post code areas and identify a suitable place to open a Hotel Business.
- Bristol is one of the UK's most popular tourist destinations. World famous Festivals like Bristol balloon festival, Bristol Harbour Festival draw lot of tourists all year
- According to the British Hospitality Association (BHA), total combined turnover for the hotel industry is estimated to exceed £40bn – a significant portion of the £127bn tourist economy. By 2025 this figure is forecast by accountancy firm Deloitte to rise to £257bn.
- Based on the background, it looks like opening a hotel in this city is a good business proposition

Data acquisition

To identify various post codes and cluster them we will be using the following data:

- List of Post codes and their coverage. This will define the area/city in scope for this project. Wikipedia will be used to source this data.
- Latitude and Longitude co-ordinates of these post codes which will be used to plot the map and also to get the popular venues list. This can either be calculated using Python's Geocoder package or pulled from a compiled list readily available on various sites.
- Venue data will be used to perform clustering of the post code areas. Explore function of Foursquare API will be used to get the most common venue categories in each post code, and then use this feature to group the various post codes of Bristol into clusters.

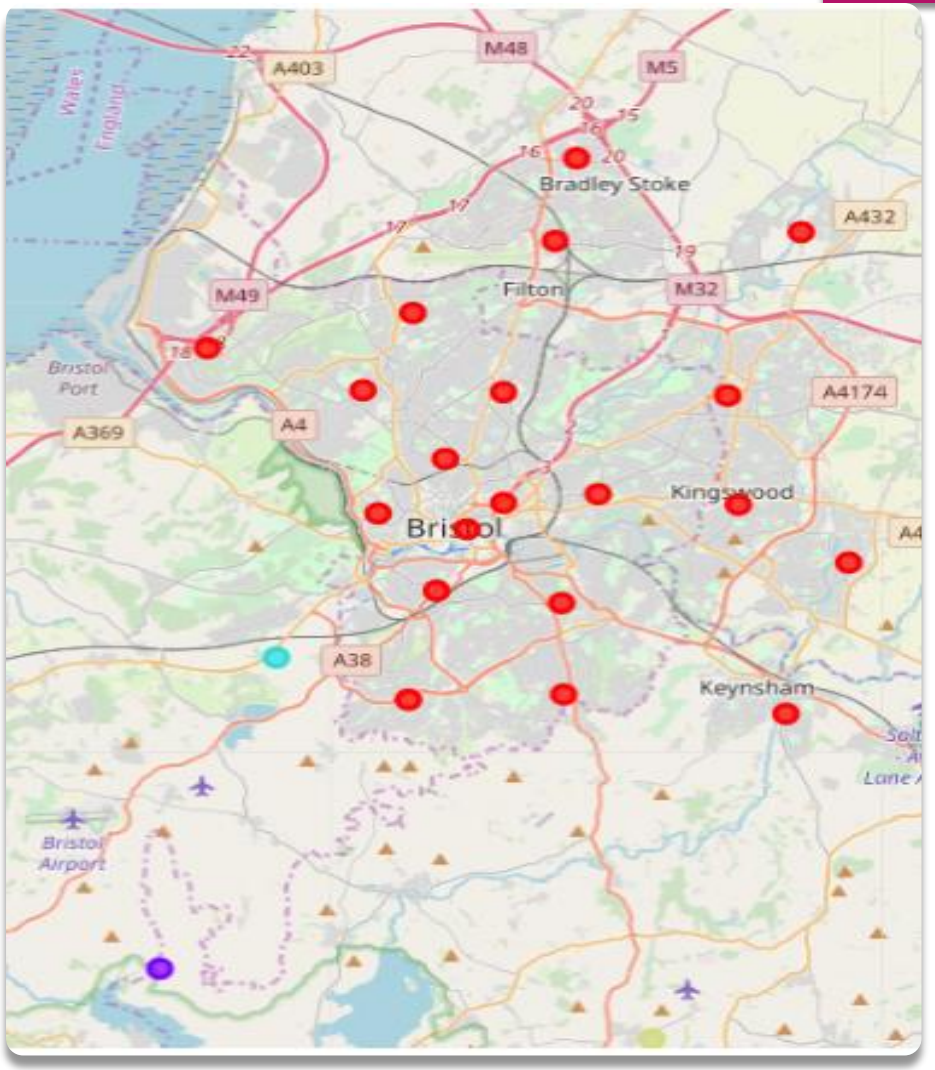
Methodology

- Firstly Wikipedia page will be scraped to get all the Post code areas and their Coverage areas
- Using the spreadsheet downloaded from site, identify Latitude and Longitude details of each of the post code areas
- Use Foursquare API to get venues data around each of these post code areas.
- Analyse this list to understand number of venues by each post code and also number of hotels within each area
- Use K-means clustering to perform clustering on the venue data using best k value
- Then finally display the clusters by overlaying them on Bristol Map using Folium and examine the clusters

Result

Clustering the Area into 4 clusters gave below insights

- Cluster 1 is busy city centre area/similar areas as it is evident from the high venue_counts in these rows. These are also the areas that have lots of cafe's, restaurants, tourist places etc
- Clusters 2 3 and 4 are not so busy and are outside the main city region with a mix of Golf courses, pub, restaurants etc.



Discussion and Recommendations

- Cluster 1 is mainly busy city centre area where there are lot of different venues for tourists. This could be a good location to setup new hotel or acquire an existing hotel. There are already hotels in this area, but looking at the demand we can view this as a good option.
- Clusters 2, 3 and 4 are similar in that they have a mix of Golf course, pubs, restaurants etc. It does not look as busy as Cluster 1 but looks like a good option if a client wants to build a hotel in quieter areas.

Conclusion

- The main purpose of the project was to help identify suitable areas in Bristol where a client can open a Hotel. To achieve this:
  - We have used foursquare api to identify venues in each of the areas
  - We identified the best k that can be used to cluster the venue data
  - Applied the algorithm to form 4 clusters
  - Superimposed the clusters on the map and examined the clusters to gain insight
- This was an attempt to collect readily available information and come up with best available options. This project can be enhanced by analysing additional points like:
  - Number of hotels in each area
  - Distance between each of them
  - Average price of rooms in a given area etc