Introduction

Business Problem:

Find best location for Home in London

London is the capital of and largest city in England and the United Kingdom, with the largest municipal population in the European Union. Standing on the River Thames in the southeast of England, at the head of its 50-mile (80 km) estuary leading to the North Sea, London has been a major settlement for two millennia.

London is considered to be one of the world's most important global cities and has been termed the world's most powerful, most desirable, most influential, most visited, most expensive, innovative, sustainable, most investment friendly, most popular for work, and the most vegetarian-friendly city in the world.

Finding best location where the price is low and a location near to facilities like supermarkets, hospitals and schools is very difficult with our proper data.

Data:

List of locations with postal code

Housing Prices with postal codes

Common venues

Geo locations file of London

Sources of data:

https://en.wikipedia.org/wiki/List_of_areas_of_London

https://propertydata.co.uk/cities/london

https://developer.foursquare.com/

Data Processing:

Cleaning & Structuring the Data

Typo

Out-of-Range Values

Missing Values

Organize data in a table, in required columns and rows

Methodology:

● **Univariate Analysis:** Analyze/explore variables one by one

- Continuous variable: explore central tendency and spread of the values

- Summary statistics

- mean, median, min, max

- IQR, standard deviation, variance, quartile

- Visualize Histograms, Boxplots

● Bivariate Analysis: Explore relationship between variables

● Coverage, missing values: treating unknown values

● Outliers: detect and treat values that are distant from other observations

Machine Learning Algorithms:

Clustering : K-Means Clustering

Find the clusters by minimizing distances of cluster centers to data.

Algorithm:

• Instantiate k distinct random guesses μ- of the cluster centers

• Each data point classifies itself as the μ- it is closest to it

• Each μ- finds the centroid of the points that were closest to it and jumps there

Tools:

● Python: NumPy, Pandas, SciPy, matplotlib

Results:

Identified Top 5 venues which are near to Supermarket, Coffee Shop, Convenience Store and Train Station.

Discussion:

Home prices near the social venues and hotels located in the middle of city are very high, compare to the neighborhoods which are awa y from the city center and also these neighborhoods also near to soc ial venues and hotels.

Conclusion:

Find the Best location where the home price is low and a location near to facilities like supermarkets, hospitals and schools.