Recommendation for suitable neighbourhood in London for Expat Indians

IBM Data Science Professional Certificate Capstone Project

Abhishek Agarwal

Introduction

Indian expats prefer to stay closer to their community to avoid cultural shock and for emotional support, getting along, food and culture.

Their Criteria of convenient neighbourhoods are:

- Indian or Asian dominated Boroughs

- Boroughs with least reported crimes of any type

- Boroughs with large number of Indian restaurants and stores

- Boroughs popular for Indian restaurants and stores

Challenge

- Have to rely on unauthenticated information present on portals

- Have to rely on other people's experience and recommendation

- Difficult to analyse and interpret data from multiple source

Solution

- Data driven solution that can collate and analyse multiple data to provide insights and recommendation

Data Acquisition, Feature selection & Preparation - 1

- Ethnic groups in London – to identify the concentration of Indian or Asian population (https://en.wikipedia.org/wiki/Ethnic_groups_in_London)

- Population in London Boroughs – As a denominator to calculate the % of ethnic groups (https://data.london.gov.uk/download/2011-census-demography/62f62c4d-eb60-4846-9efd-1e1373641452/london-unrounded-data.xls)

- Crime Data (Latest 24 month) by Borough (https://data.london.gov.uk/download/recorded_crime_summary/d2e9ccfc-a054-41e3-89fb-53c2bc3ed87a/MPS%20Borough%20Level%20Crime%20%28most%20recent%2024%20months%29.csv)

- FourSquare – To calculate most popular nearest venues data for the boroughs
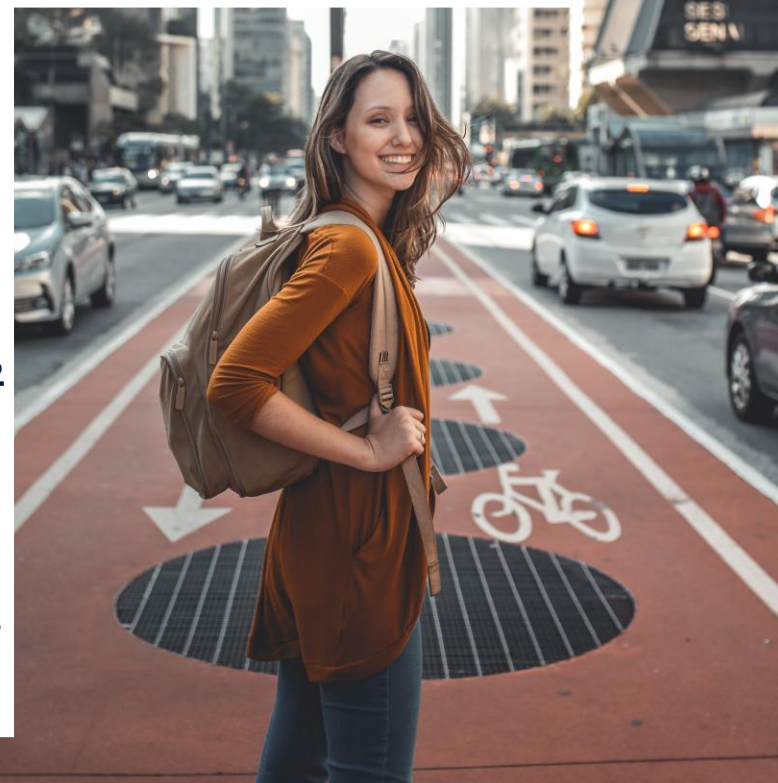
Data Acquisition, Feature selection & Preparation - 2

Concentration of Ethnic groups in London

	Borough	IndianPercent
4	Harrow	26.37
6	Hounslow	18.96
2	Brent	18.64
1	Redbridge	16.37
5	Ealing	14.25
0	Newham	13.79
7	Hillingdon	13.43
8	Barnet	7.83
9	Croydon	6.79
11	Merton	4.06
17	Barking and Dagenham	4.00
13	Enfield	3.73
10	Waltham Forest	3.54
15	Westminster	3.29
16	Greenwich	3.08

Crime Data by London Borough

Borough	TotalCrime
Barking and Dagenham	38231
Barnet	59112
Brent	60983
Camden	74864
Croydon	64392
Ealing	59413
Enfield	57762
Greenwich	54167
Harrow	31820
Hillingdon	52096
Hounslow	51863
Merton	28389
Newham	72057

Most popular venues in the Boroughs

Borough	Venue
Barking and Dagenham	38
Barnet	88
Ealing	100
Harrow	85
Wandsworth	100

Final Datasets after data preparation, cleaning and modelling

Exploratory Data Analysis - 1

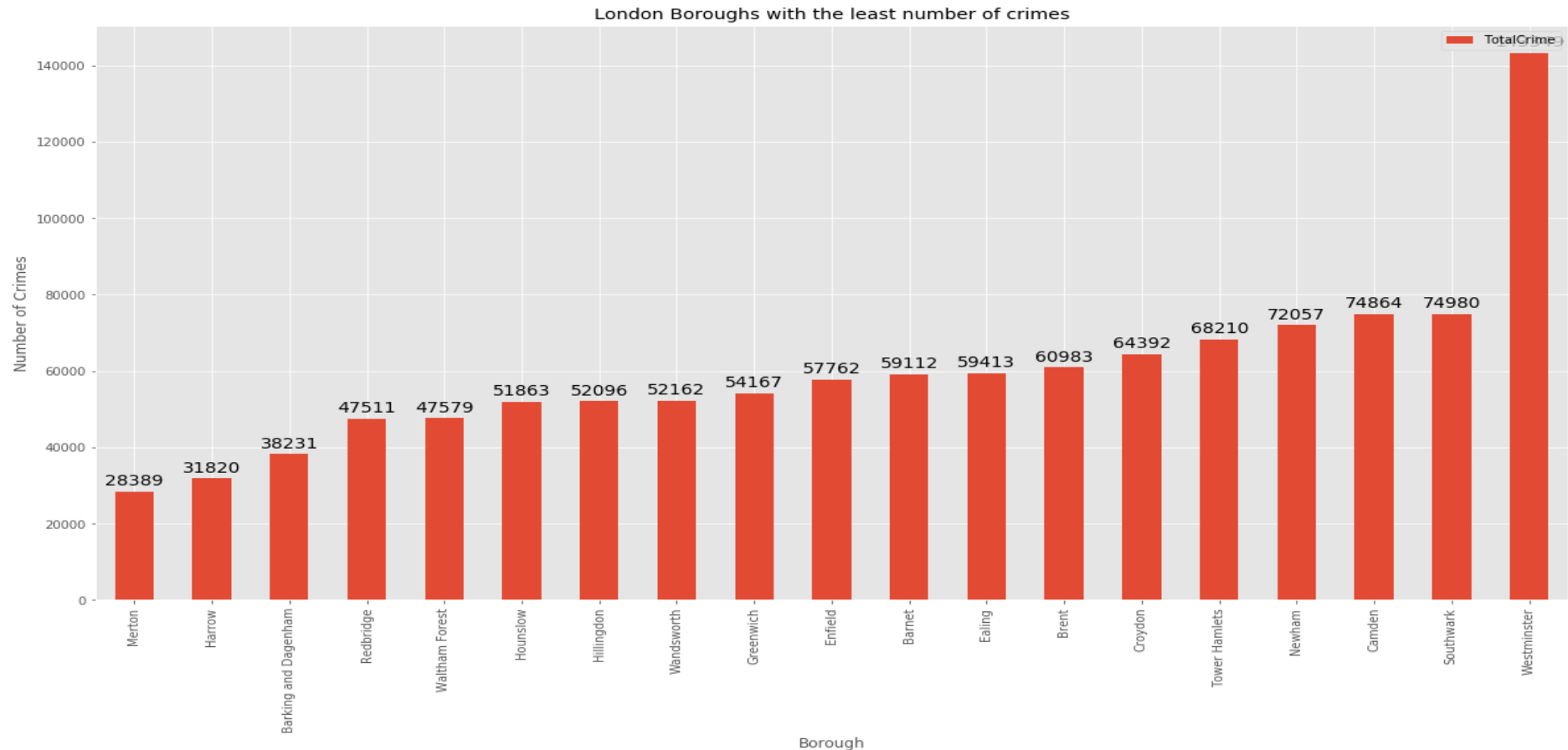## + Ethnicity

% Indian of Total Borough Population
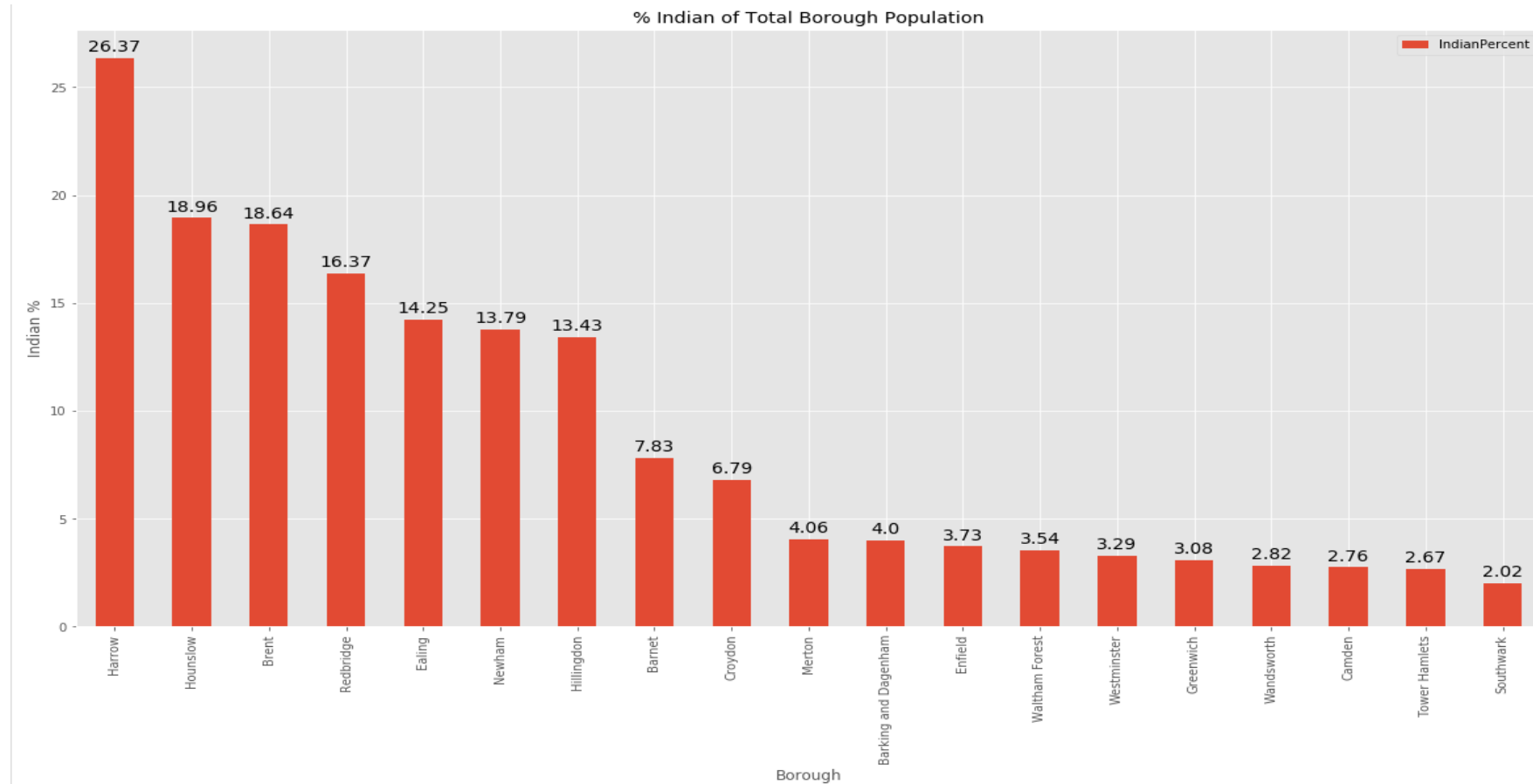
Harrow, Brent, Hounslow, Ealing, Barnet, Barking & Dagenham seems good target Boroughs based on high (%) Indian population concentration

Exploratory Data Analysis - 2

+ Crime Rates

London Boroughs with the least number of crimes

Merton, Harrow, Barking & Dagenham, Barnet seems like good target Boroughs based on low crime rates

Finding the right Boroughs for further deep dive

London Boroughs

- **High Indian Population**

- **Low crime rates**

Top 5 Selected Boroughs

Barking and Dagenham

Barnet

Ealing

Harrow

Wandsworth

Unsupervised Machine Learning Model (K-means clustering)

	Borough	Art Gallery	Arts & Crafts Store	Asian Restaurant	Bagel Shop	Bakery	Bar	Beer Store	Bike Shop	Bistro	...	Sushi Restaurant	Tennis Court
0	Barking and Dagenham	0.000000	0.00	0.000000	0.00	0.000000	0.026316	0.00	0.00	0.000000	...	0.00	0.000000
1	Barnet	0.000000	0.00	0.011364	0.00	0.011364	0.000000	0.00	0.00	0.011364	...	0.00	0.022727
2	Ealing	0.010000	0.00	0.010000	0.01	0.020000	0.010000	0.00	0.00	0.010000	...	0.02	0.000000
3	Harrow	0.011765	0.00	0.000000	0.00	0.011765	0.023529	0.00	0.00	0.000000	...	0.00	0.000000
4	Wandsworth	0.000000	0.01	0.010000	0.00	0.030000	0.010000	0.01	0.01	0.000000	...	0.01	0.000000

- One-hot encoding

- K-means Clustering

	Borough	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue
4	Harrow	51.596769	-0.337275	1	Coffee Shop	Indian Restaurant	Park	Sandwich Place	Pub	Fast Food Restaurant	Grocery Store
5	Ealing	51.512655	-0.305195	0	Pub	Coffee Shop	Park	Hotel	Pizza Place	Italian Restaurant	Café
8	Barnet	51.648784	-0.172913	0	Coffee Shop	Pub	Italian Restaurant	Grocery Store	Café	Turkish Restaurant	Park
14	Wandsworth	51.457027	-0.193261	0	Coffee Shop	Pub	Park	Café	Pizza Place	Bakery	French Restaura
17	Barking and Dagenham	51.554117	0.150504	2	Grocery Store	Supermarket	Gas Station	Park	Pizza Place	Soccer Stadium	Pub

Cluster Analysis & Discussion

- Cluster 1 presents 3 borough options with a mix of moderate crime rates, Indian concentration and other amenities. This would be interest to somebody who enjoys different cuisine (Thai, Chinese, Italian).

Cluster 1

```
df_LondonBorough_clusters.loc[df_LondonBorough_clusters['Cluster Labels'] == 0, df_LondonBorough_clusters.
columns[[0] + list(range(4, df_LondonBorough_clusters.shape[1]))]]
```

	Borough	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th M Comm Venue
5	Ealing	Pub	Coffee Shop	Park	Hotel	Pizza Place	Italian Restaurant	Café	Persian Restaurant	Sandwich Place	Groce Store
8	Barnet	Coffee Shop	Pub	Italian Restaurant	Grocery Store	Café	Turkish Restaurant	Park	Fish & Chips Shop	Pizza Place	Super
14	Wandsworth	Coffee Shop	Pub	Park	Café	Pizza Place	Bakery	French Restaurant	Thai Restaurant	Supermarket	Groce Store

Cluster 2

```
df_LondonBorough_clusters.loc[df_LondonBorough_clusters['Cluster Labels'] == 1, df_LondonBorough_clusters.
columns[[0] + list(range(4, df_LondonBorough_clusters.shape[1]))]]
```

	Borough	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
4	Harrow	Coffee Shop	Indian Restaurant	Park	Sandwich Place	Pub	Fast Food Restaurant	Grocery Store	Supermarket	Gym / Fitness Center	Chinese Restaurant

- Cluster 2 - If an individual prefers great Indian restaurant, along with other features like gym, park, super market, then cluster 2 would be ideal for him. Borough in this cluster has the highest Indian concentration and is the safest.
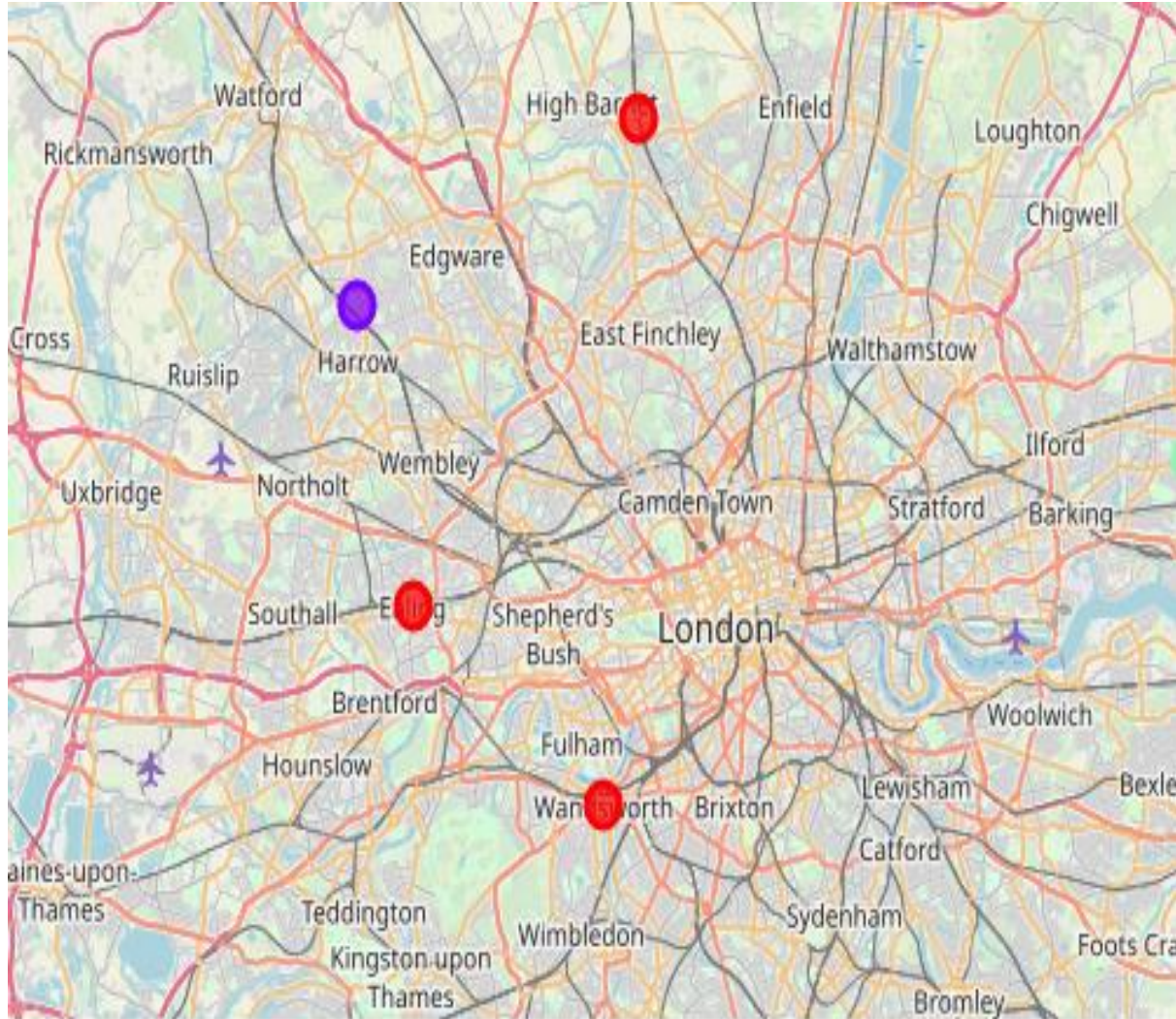
- Cluster 3 - If train commute is a deciding factor for an individual then cluster 3 option would be a good option. Also it has 3rd lowest crime rate, modest Indian concentration and good selection of other amenities such as grocery stores.

Cluster 3

```
df_LondonBorough_clusters.loc[df_LondonBorough_clusters['Cluster Labels'] == 2, df_LondonBorough_clusters.
columns[[0] + list(range(4, df_LondonBorough_clusters.shape[1]))]]
```

	Borough	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
17	Barking and Dagenham	Grocery Store	Supermarket	Gas Station	Park	Pizza Place	Soccer Stadium	Pub	Racetrack	Metro Station	Restaurant

Conclusion

- With Park, Gym and Grocery store, coupled with high Indian density, lowest crime rates – it looks like Harrow is a good suggestion according to the criteria provided by a short-term Indian expat.

- Barking & Dagenham is a good option which provides good transport option with a metro station and other amenities and satisfying the selection criteria of our clients.

- Until now many Indian expats have relied on word of mouth and basic internet search as an alternative approach to identify safe areas to consider but that approach was always very subjective to the liking and bias of the person recommending it. This project provides a data driven approach to recommendations and insights to make an informed decision.

Future work

It would be nice to add following datasets to provide more robust recommendations:

1. Travel time to work place

2. Housing availability (short term house listings)

3. Housing prices

4. Cost of living index

THANK YOU