The Battle of the Neighborhoods:
LONDON

Victoria Ahmadi
Coursera Capstone

Navigating London’s Real Estate Market

- London is a dynamic and transient city that attracts many expatriates
- London Real Estate is DAUNTING due to: size, diversity, and competitiveness
- Determining best neighborhood varies for everyone
- For a young professional, three things matter: safety, affordability, and vibe

Data Acquisition and Cleaning

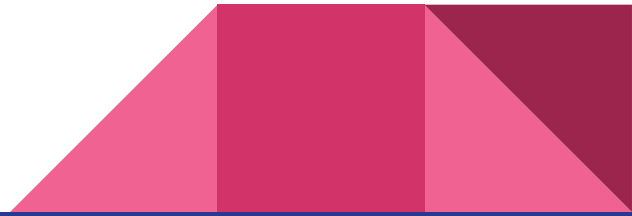
- Data Sources:
  - Listing of London boroughs from Wikipedia
  - Official crime records from Kaggle
  - Housing prices (price per square meter) from UK Office of National Statistics
  - Happiness index
- Data cleaning
  - Boroughs: official names taken from Wiki page
  - Crime rate: utilized most recent year in dataset - 2016; removed rows with missing data
  - Housing prices: kept two columns only - borough name and price per meter squared

Methodology

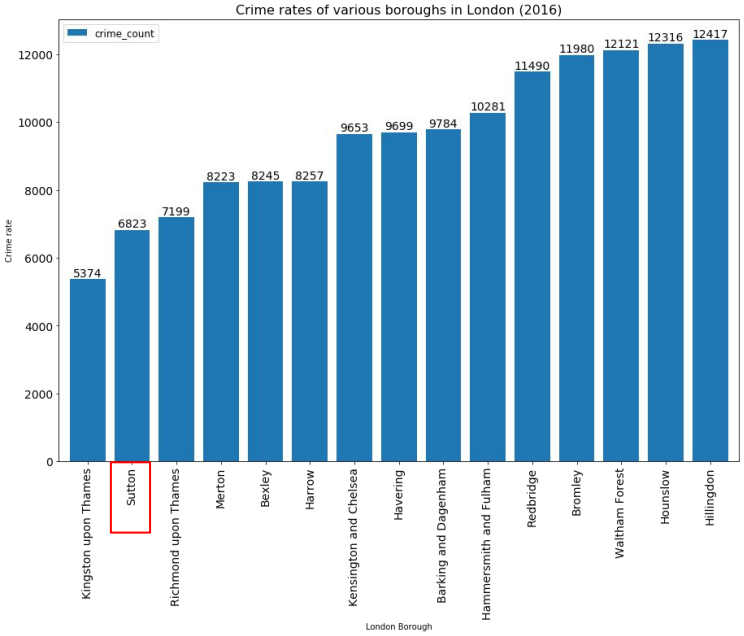
Two steps:

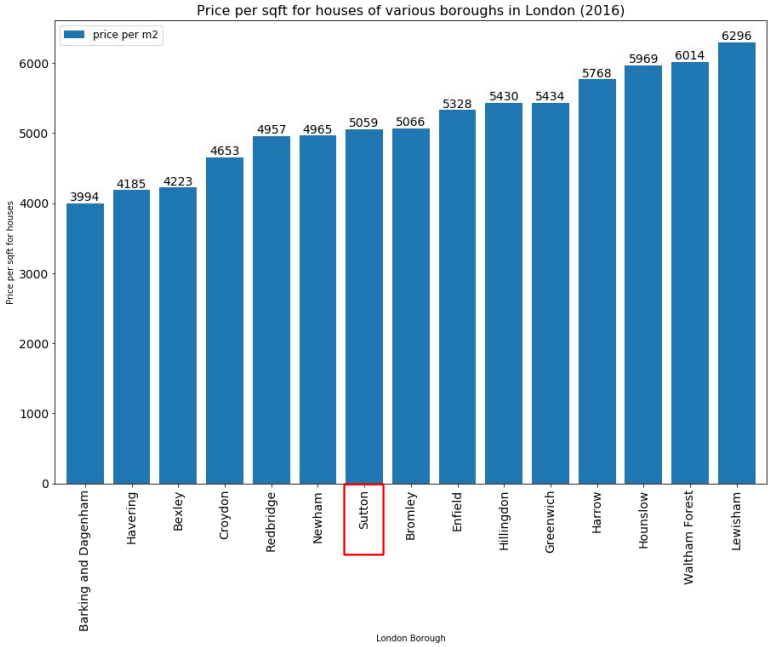
1. Exploratory Data Analysis: to visualize safest, most affordable, and happiest places in London; then select borough to further analyze

2. Modeling: further delve into the specifics of the selected borough; analyze most popular venues



Exploratory Data Analysis





	London_Borough
0	Richmond Upon Thames
1	Kingston upon Thames
2	Bromley
3	Sutton
4	Wandsworth
5	Camden
6	Barnet
7	Ealing
8	Greenwich
9	Havering
10	Hackney
11	Waltham Forest
12	Merton
13	Kensington & Chelsea
14	Hammersmith & Fulham

Crime Rate

Price per square meter

Happiness index

Modeling

1. First, we obtain an official list of districts within **Sutton** - from Wikipedia.

2. Then, using Foursquare Location data, we obtained the 100 most popular venues within a 500m radius for each district.

3. Use one-hot encoding to transform categorical variables to binary format.

4. Perform k-means clustering

Cluster 1 - Residential

Most common venues include parks, restaurants, and hardware stores. Note that there is a lack of TRANSPORT venues, which suggest that this area is quieter, more removed, and perhaps residential.

	District	Borough	Latitude	Longitude	Cluster_FINAL	Cluster_Label	Cluster Label	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Co Ver
0	Bandon Hill	Sutton	51.364777	-0.134833	0	1	0	1	Park	Pub	Co Sto
1	Beddington	Sutton	51.371988	-0.132393	0	1	0	1	Indian Restaurant	Hardware Store	Par
14	South Beddington	Sutton	51.371988	-0.132393	0	1	0	1	Indian Restaurant	Hardware Store	Par

Cluster 2 - Busy, Happening, Central

	District	Borough	Latitude	Longitude	Cluster_FINAL	Cluster_Label	Cluster Label	Cluster Labels	1st Most Common Venue	2nd Most Common Venue
3	Belmont	Sutton	51.343785	-0.201152	1	2	1	2	Train Station	Asian Restaurant
4	Benhilton	Sutton	51.371642	-0.191571	1	2	1	2	Indian Restaurant	Gym / Fitness Center
5	Carshalton	Sutton	51.365788	-0.161086	1	2	1	2	Grocery Store	Park
6	Carshalton Beeches	Sutton	51.357196	-0.169351	1	0	1	0	Train Station	Italian Restaurant
8	Cheam	Sutton	51.357616	-0.216241	1	2	1	2	Pub	Coffee Shop
9	Hackbridge	Sutton	51.379613	-0.156754	1	2	1	2	River	Train Station
10	Little Woodcote	Sutton	51.346076	-0.145932	1	5	3	5	Park	Sports Club
11	North Cheam	Sutton	51.371578	-0.220225	1	2	1	2	Coffee Shop	Gym / Fitness Center
12	Rosehill	Sutton	51.012505	-0.140639	1	2	1	2	Restaurant	Grocery Store
13	St. Helier	Sutton	51.386695	-0.180057	1	3	1	3	Pharmacy	Breakfast Spot
15	Sutton	Sutton	51.357511	-0.173640	1	0	1	0	Train Station	Italian Restaurant
16	Sutton Common	Sutton	51.375373	-0.196032	1	2	1	2	Train Station	Gym / Fitness Center

This cluster is by far the largest, with a myriad of diverse venues, most notably: train stations, pubs, grocery stores, coffee shops, and gyms. Based off this data, this cluster is perhaps located in central Sutton, which is busier, more happening, and accessible to other parts of London.

Cluster 3 - Business District

The last cluster mainly consists of business services, racetracks, and vets. The lack of essential services - grocery stores, transport - suggest that this location is perhaps located in a remote and quieter part of Sutton.

	District	Borough	Latitude	Longitude	Cluster_FINAL	Cluster_Label	Cluster Label	Cluster Labels	1st Most Common Venue	2nd Most Common Venue
2	Beddington Corner	Sutton	51.386942	-0.149532	2	4	2	4	Business Service	Racetrack

Sutton for a Young Professional

- Sutton -
  - Second safest
  - Seventh most cheapest
  - Fourth happiest

- Cluster 2 -
  - Active - parks, diversity in types of venues
  - Accessible - in numerous districts within cluster 2, transport was noted as the most common venue, making other parts of London easily accessible from Sutton
  - Social - pubs, pubs, pubs

Conclusion and future directions

- Obtained and cleaned real data
- Built useful model using k-means algorithm
- Identified a safe, affordable, and happy location to live in London for a young professional