THE BATTLE OF NEIGHBORHOOD

Vrinda Sharda

Business Problem Section

- London is the largest city and the capital of United Kingdom. The city covers a total area of 607 sq. miles.
- In a country with such a huge area and high real estate prices, it becomes very tough to find a property with its appropriate value, keeping in mind all the amenities and essential facilities surrounding such venues i.e. elementary schools, high schools, hospitals & grocery stores.
- As a result, the business problem we are currently posing is: How could we recommend profitable venues to support homebuyers according to amenities and essential facilities surrounding such venues?
- The real estate investors would be very interested in improving their business and investing at a place which provides them maximum profit.

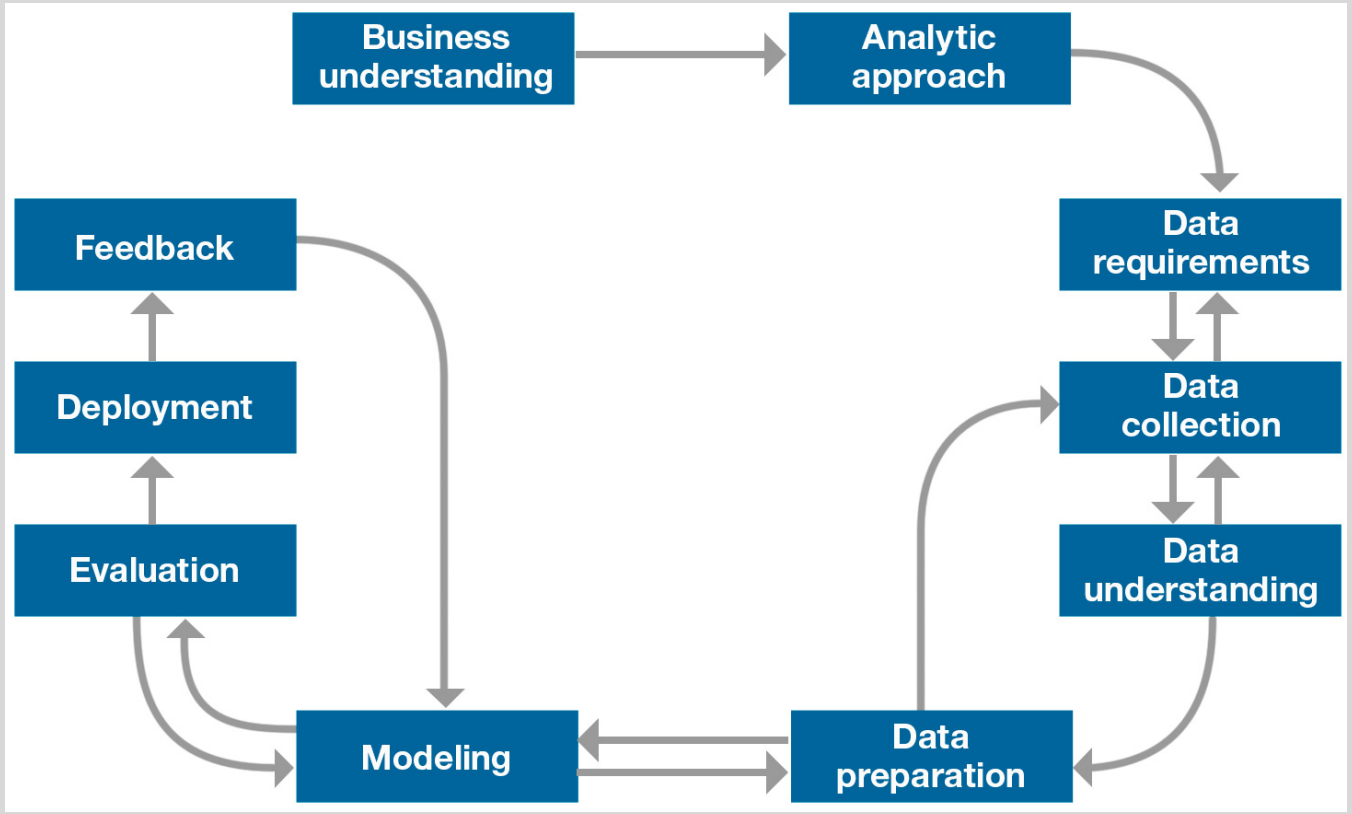
Data Section

- Data on London properties and the relative price paid data were extracted from the HM Land Registry (http://landregistry.data.gov.uk/).
- To explore and target recommended locations across different venues according to the presence of amenities and essential facilities, we will access data through FourSquare API interface and arrange them as a data frame for visualization.
- By merging data on London properties and the relative price paid data from the HM Land Registry and data on amenities and essential facilities surrounding such properties from FourSquare API interface, we will be able to recommend profitable real estate investments.

Methodology

Data Science Methodology consists of the following steps:

- Business understanding
- Analytic approach
- Data Requirement
- Data Collection
- Data Understanding
- Data Preparation
- Modeling
- Evaluation
- Deployment
- Feedback.



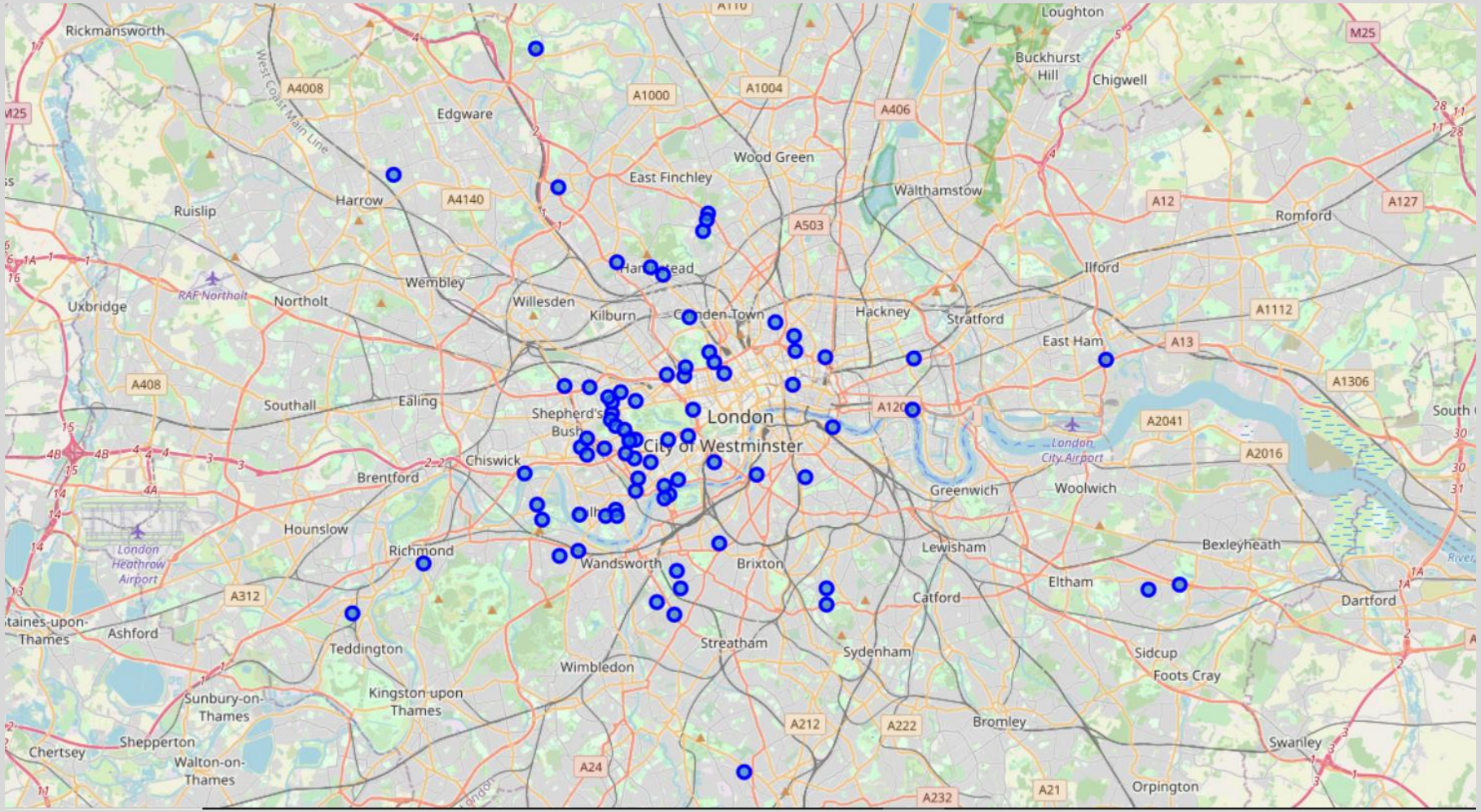
Data Preparation

At this stage, we convert our data from raw initial form into another pre processed from in order to prepare the data for further analysis. Therefore we perform the following steps:

- Renaming the columns
- Formatting the date column and sorting the data by the date of sale
- Making a list of street names in London.
- Calculating the streetwise average price and the property
- Mapping the locations

	Street	Avg_Price
196	ALBION SQUARE	2.450000e+06
390	ANHALT ROAD	2.435000e+06
405	ANSDELL TERRACE	2.250000e+06
422	APPLEGARTH ROAD	2.400000e+06
855	BARONSMEAD ROAD	2.375000e+06
981	BEAUCLERC ROAD	2.480000e+06
1102	BELVEDERE DRIVE	2.340000e+06
1215	BICKENHALL STREET	2.208500e+06
1253	BIRCHLANDS AVENUE	2.217000e+06
1553	BRAMPTON GROVE	2.456875e+06
1632	BRIARDALE GARDENS	2.397132e+06
1797	BROOKWAY	2.400000e+06
1914	BURBAGE ROAD	2.445000e+06
1980	BURY WALK	2.492500e+06

Affordable locations



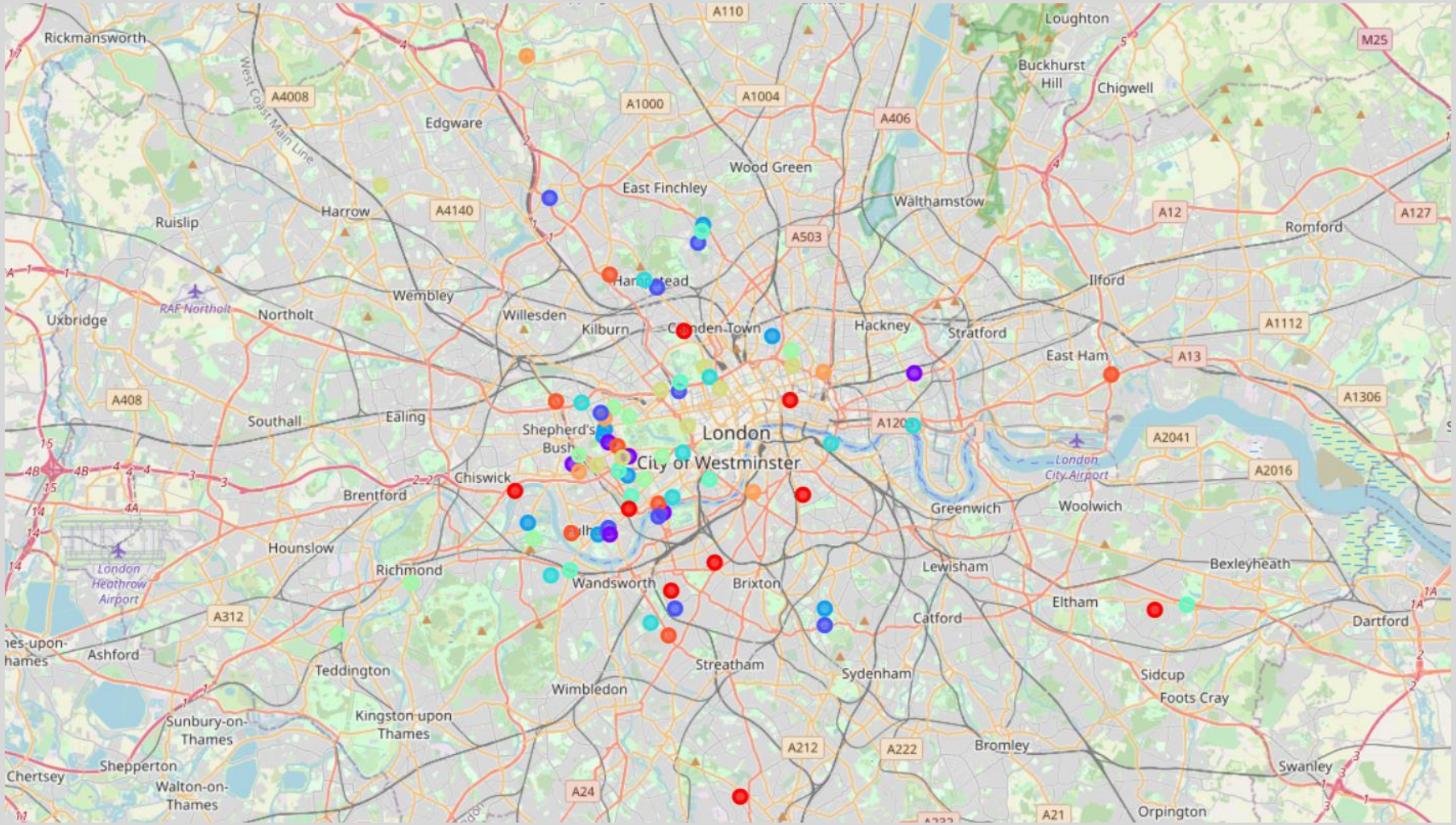
Modeling

- We have a list of affordable streets in London. Now, out of these 162 available options, which one should the investor invest in?
- To help our client, we shall now use a machine learning algorithm.
- Since we need to group the object into similar objects and dissimilar objects to data points in other groups, we will use clustering.
- We will use the K-Means Clustering technique as it is the fastest and efficient in terms of computational cost, is highly flexible to account for mutations in real estate market in London.
- We start with the FourSquare API and collect all the venues int the neighborhood of our filtered streets.
- Since we need decimal values in our algorithm, we use one hot encoding and then get its mean.

- After sorting them according to the most visited venues, we make a data frame having the street name and top 10 most common venues. We now have a table of shape(149, 11)
- We will be distributing the area into 10 clusters and then we merge our two tables to get a final one which includes street, average price, latitude, longitude, cluster labels, and 10 most common venues.

	Street	Avg_Price	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	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
196	ALBION SQUARE	2.450000e+06	-41.273758	173.289393	7	Café	Indian Restaurant	Pub	Coffee Shop	Restaurant	Bar	Burger Joint	New American Restaurant	Seafood Restaurant	Museum
390	ANHALT ROAD	2.435000e+06	51.480316	-0.166801	1	Pub	French Restaurant	Grocery Store	Diner	Plaza	Japanese Restaurant	Gym / Fitness Center	Cocktail Bar	English Restaurant	Garden
405	ANSDELL TERRACE	2.250000e+06	51.499890	-0.189103	8	Juice Bar	Hotel	Pub	Restaurant	Indian Restaurant	Clothing Store	Italian Restaurant	Café	Supermarket	Mediterranean Restaurant
422	APPLEGARTH ROAD	2.400000e+06	53.748654	-0.326670	9	Pub	Nightclub	Bar	Casino	Food & Drink Shop	Food	Food Service	Exhibit	Factory	Falafel Restaurant
855	BARONSMEAD ROAD	2.375000e+06	51.477315	-0.239457	3	Pub	Thai Restaurant	Pizza Place	Community Center	Restaurant	Coffee Shop	Park	Farmers Market	Café	Nature Preserve
981	BEAUCLERC ROAD	2.480000e+06	30.211452	-81.617981	4	Spa	Automotive Shop	Harbor / Marina	Pizza Place	Zoo Exhibit	Filipino Restaurant	Event Space	Exhibit	Factory	Falafel Restaurant
1102	BELVEDERE DRIVE	2.340000e+06	44.707562	-63.545599	5	Campground	Zoo Exhibit	Film Studio	Exhibit	Factory	Falafel Restaurant	Farm	Farmers Market	Fast Food Restaurant	Filipino Restaurant
1215	BICKENHALL STREET	2.208500e+06	51.521201	-0.158908	2	Gastropub	Restaurant	Coffee Shop	Pizza Place	Hotel	Italian Restaurant	Garden	Bakery	Greek Restaurant	Bar
1253	BIRCHLANDS AVENUE	2.217000e+06	51.448394	-0.160468	2	Pub	Breakfast Spot	Coffee Shop	Brewery	French Restaurant	Chinese Restaurant	Lake	Train Station	Bakery	Pizza Place
1553	BRAMPTON GROVE	2.456875e+06	51.589961	-0.318525	7	Food Service	Home Service	Zoo Exhibit	Film Studio	Event Space	Exhibit	Factory	Falafel Restaurant	Farm	Farmers Market
1632	BRIARDALE GARDENS	2.397132e+06	51.560175	-0.195431	9	Indian Restaurant	Chinese Restaurant	Health & Beauty Service	Coffee Shop	Grocery Store	Fast Food Restaurant	Ethiopian Restaurant	Event Space	Exhibit	Factory
1797	BROOKWAY	2.400000e+06	45.432185	-122.802812	9	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
1914	BURBAGE ROAD	2.445000e+06	52.538507	-1.353674	7	Construction & Landscaping	Bar	Grocery Store	Dance Studio	Athletics & Sports	Film Studio	Exhibit	Factory	Falafel Restaurant	Farm
1980	BURY WALK	2.492500e+06	52.145529	-0.423593	4	Supermarket	English Restaurant	Rental Car Location	Gym	Hardware Store	Fast Food Restaurant	Coffee Shop	Café	Dry Cleaner	Park

- We now head towards our final step, where we map the cluster into the map of London and compare.



Results and Discussion

- We examined them according to neighbourhoods/London areas. although West London (Notting Hill, Kensington, Chelsea, Marylebone) and North-West London (Hampsted) might be considered highly profitable venues to purchase a real estate according to amenities and essential facilities surrounding such venues i.e. elementary schools, high schools, hospitals & grocery stores, South-West London (Wandsworth, Balham) and North-West London (Islington) are arising as next future elite venues with a wide range of amenities and facilities. Accordingly, one might target under-priced real estates in these areas of London in order to make a business affair.
- we divided them into 10 clusters. Clusters 0, 1 and 3 are for the ones who enjoy spending weekends in clubs and pubs, while cluster 5 and 6 are for theatre lovers. Investing in clusters 4, 8 and 9 would be the best as they provide the maximum facilities in their neighbourhoods. For the home buyers who are prone to live in 'green' areas should go for areas in cluster 2, 5, 6 and 7 as they have parks, gardens, farms and waterfronts.