Matching Homes and Homebuyers using Machine Learning

The Great Dilemma: Where to Buy A House

Some Factors that Affect Decision

- Facilities: shops, restaurants, parks/recreation facilities, cinemas, parking, etc
- Space: when buying a home for the long run, buyers look for spacious settings where they can possibly grow families
- Noise: pubs, clubs, football stadiums
- Transport: distance to tube or rail station

Data Acquisition and Cleaning

Two datasets were used for the analysis

- The first was sourced directly from HM land registry and the relevant csv.file was downloaded. This publicly accessible data contains information that was collected as part of the land registration process
- The second dataset was collected from doogal.co.uk which contained all postcodes in London.
- **Both datasets were combined and used for the analysis**

Hypothetical Situation:

Jane (the client) is interested in finding a flat in London to lease under with sale value of under 400,000 pounds. How do we match her to the right home?

Table 1: Summary of Housing Top Ten Cheapest Districts Across London

District	Price	Latitude	Longitude
BARKING AND DAGENHAM	215287.3936	51.546836	0.120551
BEXLEY	232622.1975	51.459085	0.137659
HAVERING	235475.6915	51.574386	0.202485
CROYDON	260489.1525	51.377398	-0.090463
GREENWICH	261457.0153	51.476143	0.049714
ENFIELD	267108.1709	51.643028	-0.077646
HILLINGDON	267435.8518	51.539299	-0.440874
SUTTON	269644.5216	51.364263	-0.181839
REDBRIDGE	270996.4715	51.582171	0.065969
HOUNSLOW	272027.224	51.468124	-0.360679

Given that the are Barking and Dagenham, and Bexley. *Cluster analysis* or *clustering* was used to compare both areas to find Jane a suitable flat.

Table 2: Top 1st and 2nd Most Common Venue Category Per Cluster in Barking and Dagenham

Cluster	First Most Common Venue	Second Most Common Venue
1	Hotel	Indian Restaurant
2	Grocery Store	Pharmacy
3	Warehouse	Flea Market
4	Cafe	Park
5	Park	Construction and Landscaping
6	Cafe	Fish and Chips Shop
7	Cafe	Dry Cleaner
8	Hardware Store	Grocery Store
9	Grocery Store	Grocery Store
10	Platform	Bar

Map Showing Barking and Dagenham Clusters

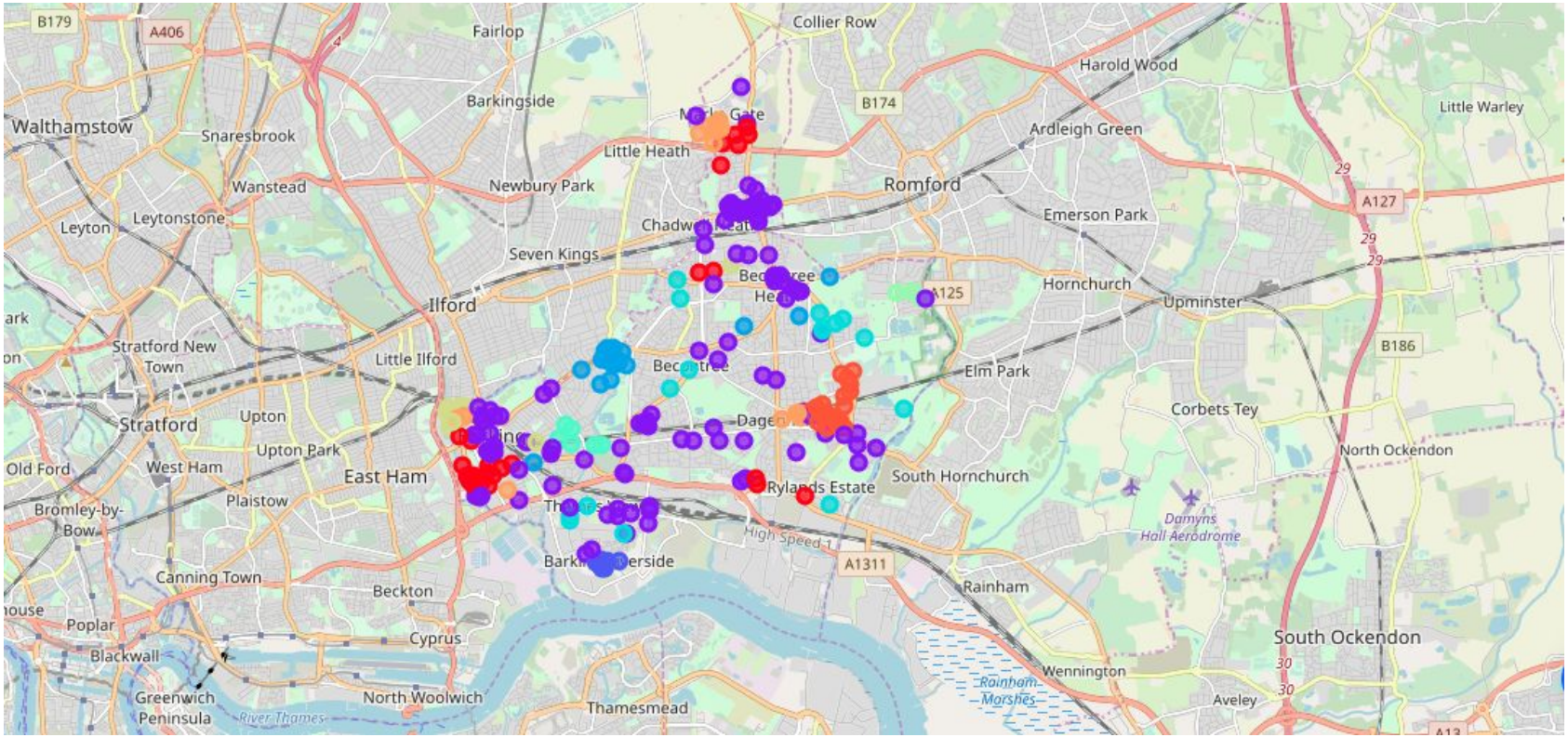
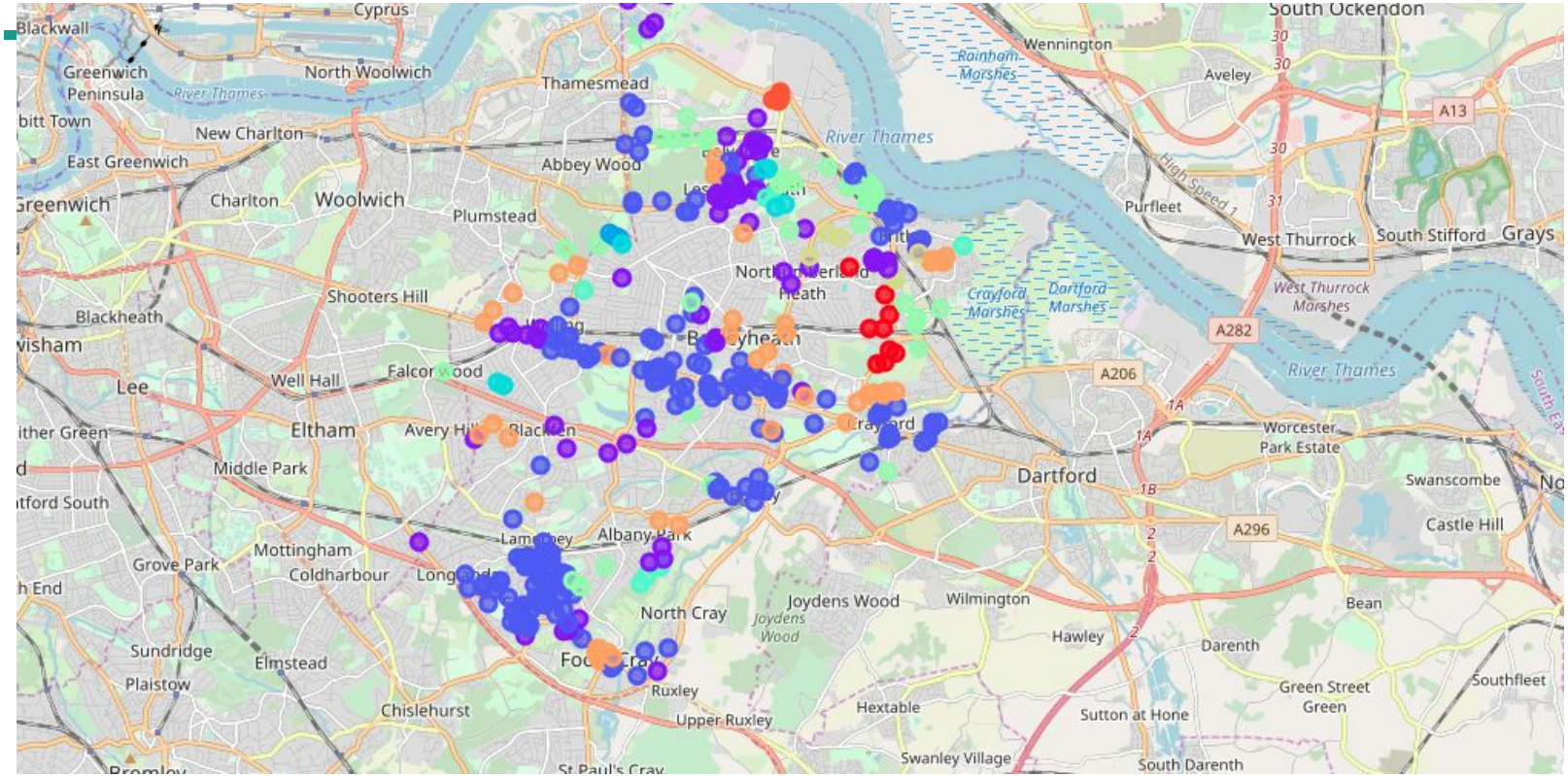


Table 3: Top 1st and 2nd Most Common Venue Category Per Cluster in Bexley

Cluster	First Most Common Venue	Second Most Common Venue
1	Fish and Chips Shop	Fish and Chips Shop
2	Grocery Store	Fast Food Restaurant
3	Coffee Shop	Coffee Shop
4	Forest	Zoe Exhibit
5	Indian Restaurant	Zoe Exhibit
6	Construction and Landscaping	Zoe Exhibit
7	Pizza Place	Park
8	Grocery Store	Zoe Exhibit
9	Pub	Pub
10	English Restaurant	Coffee Shop

Map Showing Bexley Clusters



Barking and Dagenham vs. Brexley

Barking and Dagenham (Winner!!)	Brexley
More recreational areas such as parks and playgrounds.	More postcodes/flats that are in areas with accessible food places (grocery stores, cafes, pubs, restaurants, etc)
More variety in the popular venues	More clusters that would generally be attractive to a home buyer (As seen in Table 2 and 3),
Flat prices are on average lower in Barking and Dagenham compared to Brexley (the decider!)	

Jane is now happy!

Conclusion

Going forward, this type of research can be improved by combined the Clustering Machine Learning used in the project with other machine learning techniques to maximize solutions the “affordability and suitability” dilimea.