Finding the Optimal Location for Opening a New Restaurant in South West London

IBM Data Science Capstone Project

Ken Young, August 2020

Location is a key factor in determining profitability for a new restaurant venture

- Our client operates restaurants in Central London.
- Reputation for excellent service with innovative menus.
- Blended cuisine from East and South East Asia.
- Our client would like to expand into South West London.

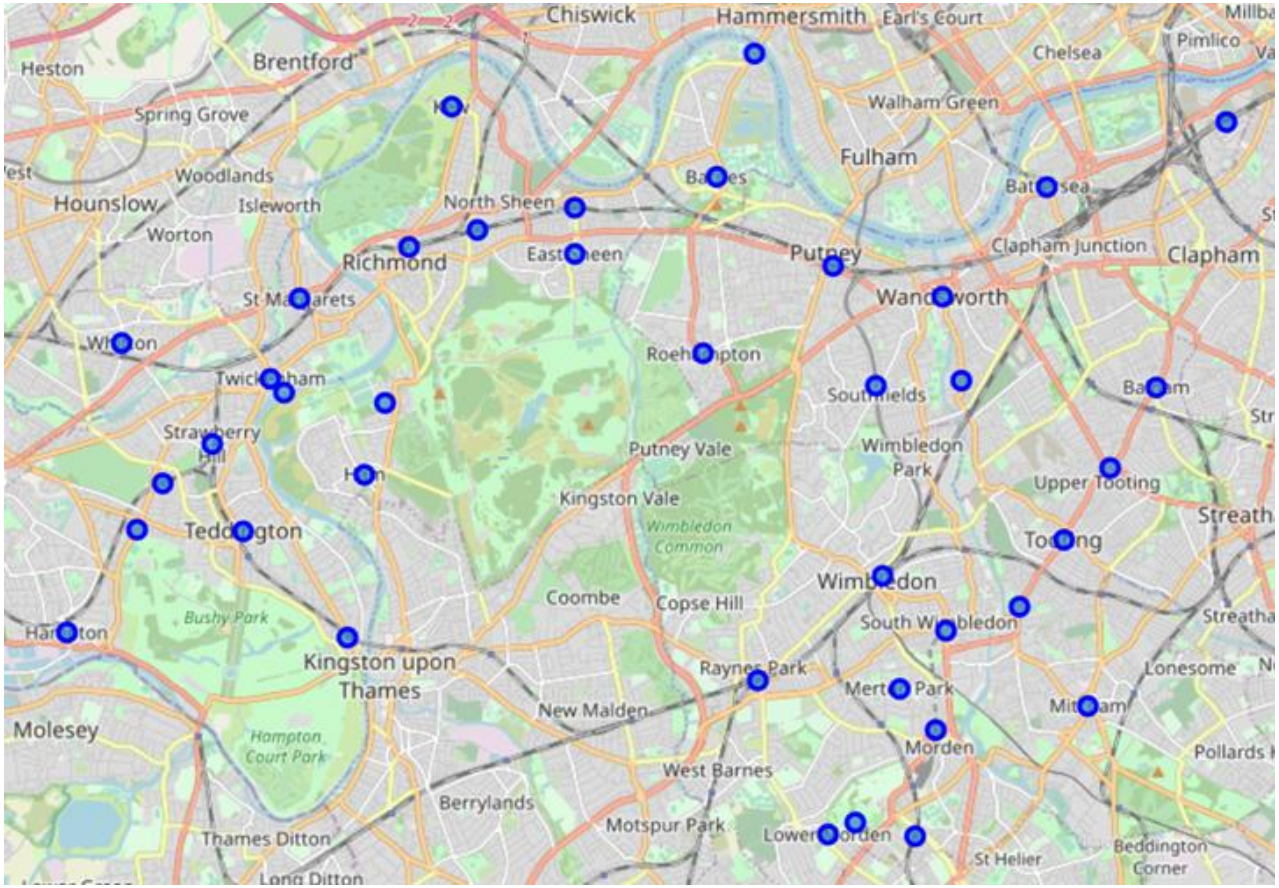
Our challenge:-

- To find potential optimum locations for a new restaurant.
- We will use data science techniques to assist in our search.
- We will consider the following factors in our analysis:-
  - Potential competitor restaurants already established.
  - Spending power of local population.
  - Proximity to transportation hubs.
  - Proximity to entertainment venues.

Data acquisition and cleaning

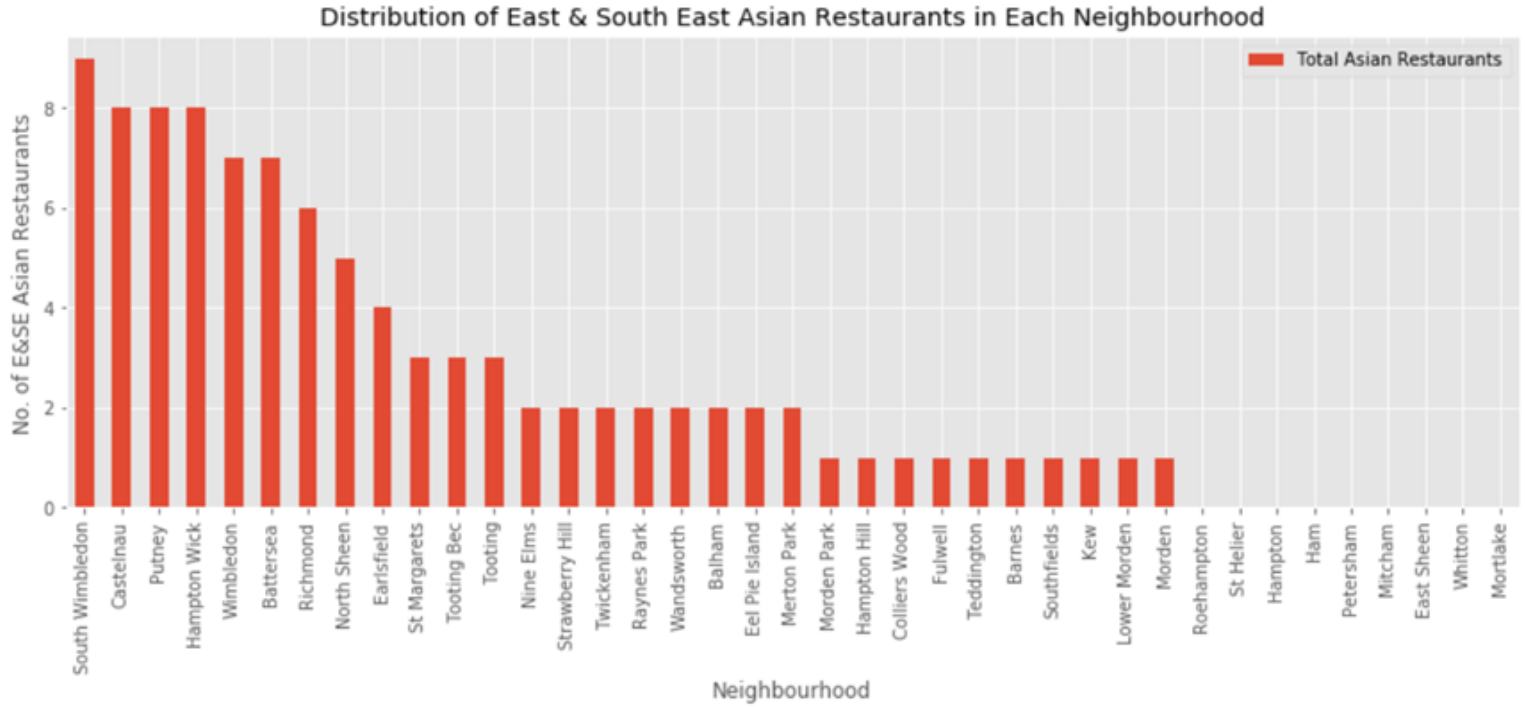
- List of neighbourhoods for our three target boroughs (Merton, Richmond, Wandsworth) scraped from https://en.wikipedia.org/wiki/List_of_areas_of_London
- Median household income for latest year available downloaded from https://data.london.gov.uk/dataset/household-income-estimates-small-areas
- Latitude and longitude coordinates extracted from Python Geocoder.
- Location information for competitor restaurants, train stations, entertainment venues extracted from Foursquare.
- In total we have 39 neighbourhoods in our dataset.

Neighbourhoods in South West London



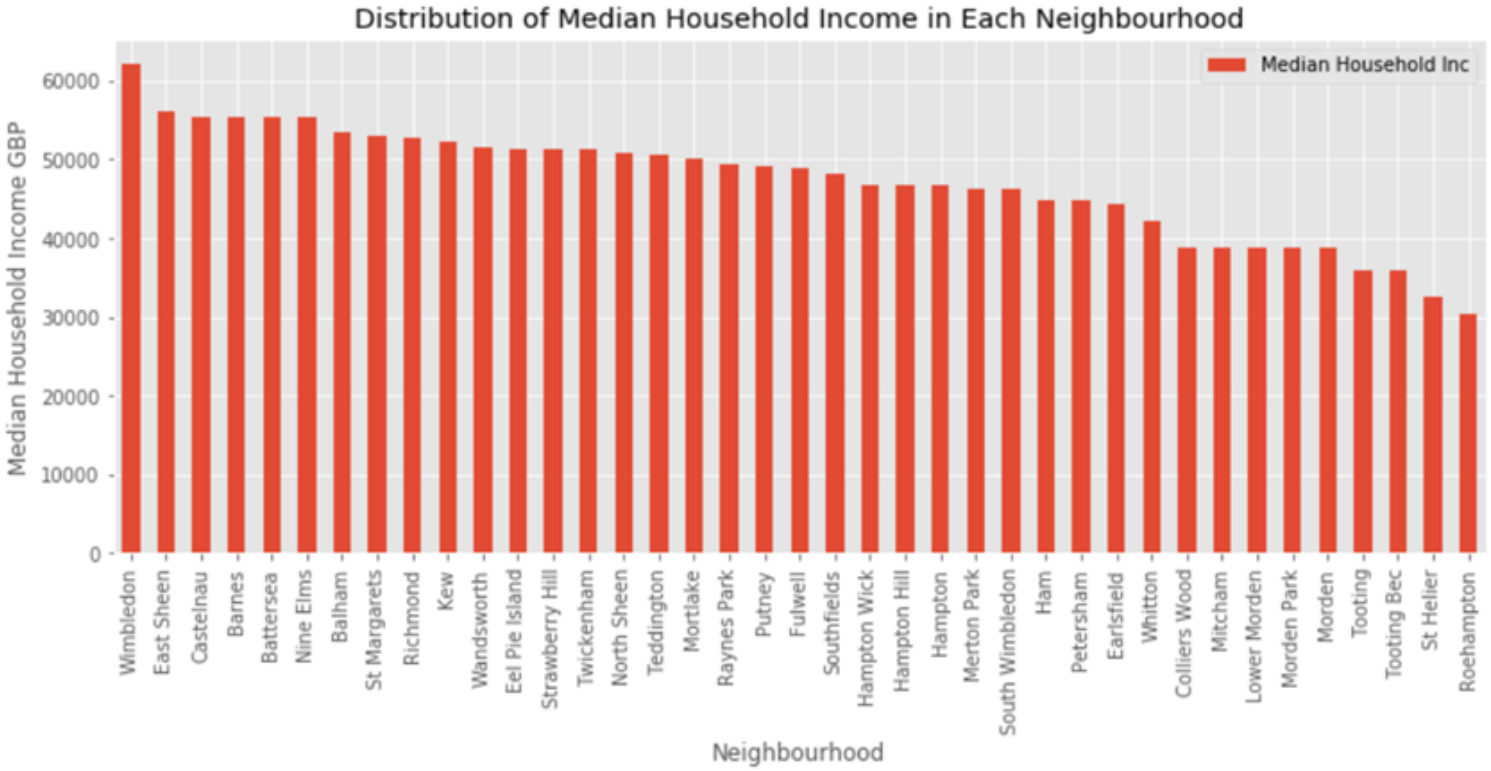
Competitor restaurants already established

- Number of venues operating within 1km of each neighbourhood centre.
- Totals include individual country cuisines as well as pan East & South East Asian.
- Our assumption is too many competitors is a crowded market and results in barriers to entry; whereas no competitors could mean no demand.



Median household income: Spending power in each neighbourhood

- Our client operates in the casual dining market with most customers in the higher income brackets.
- Our assumption is the higher the median household income, the higher the disposable income available to spend on casual dining.



Proximity to transportation hubs and entertainment venues

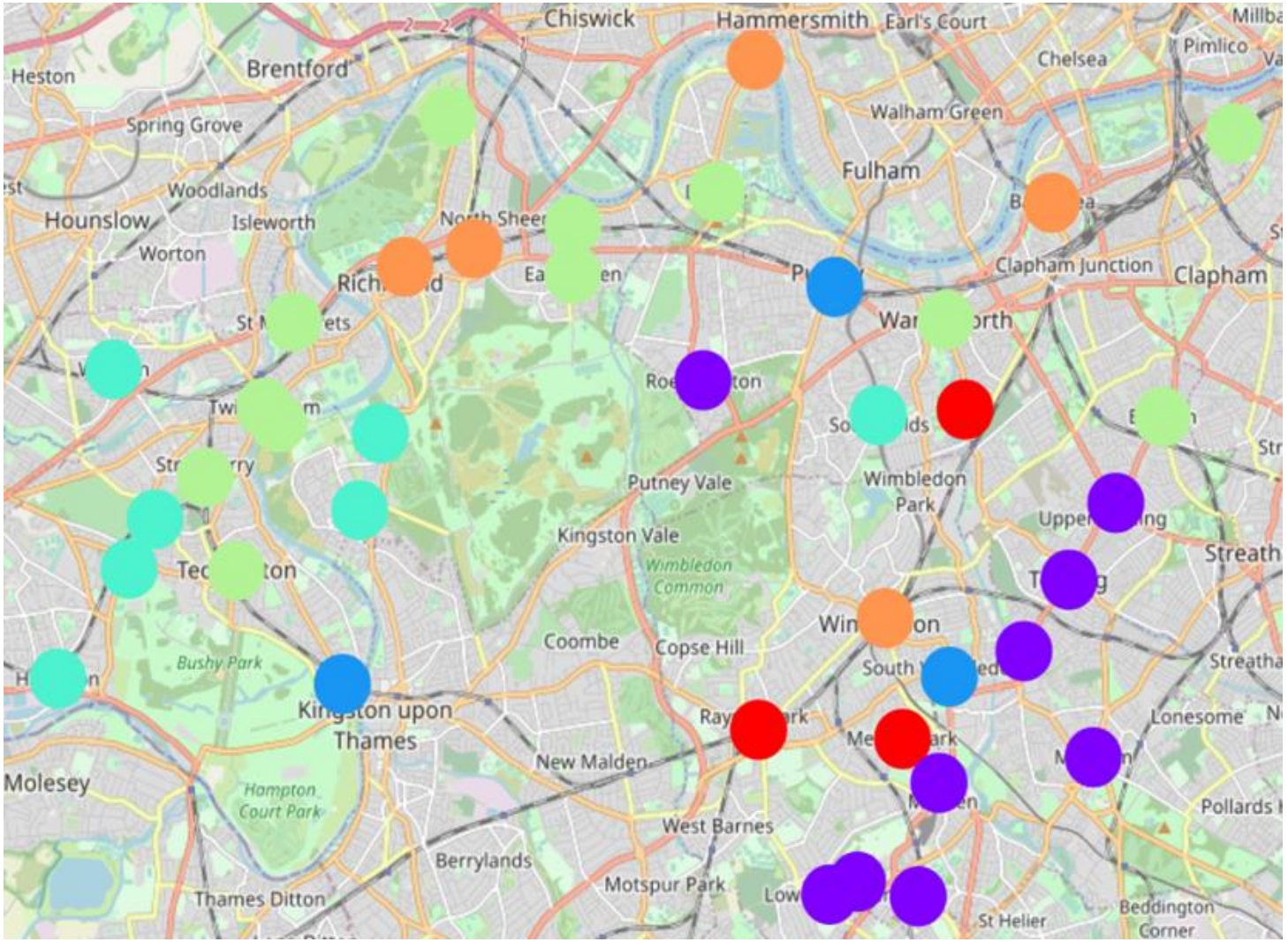
- Most neighbourhoods in our study are well served by at least one train station, with frequent services to Central London.
- South West London in general is suburban with commercial centres in each neighbourhood offering various mixes of supermarkets, shops, pubs, cafes and restaurants.
- Some neighbourhoods have entertainment venues (mostly cinemas) or major tourist attractions (e.g. Kew Gardens).
- Our assumption is there is little value in conducting data analysis on these factors at this stage; however these factors should certainly be considered once we have a list of preferred neighbourhoods.

Methodology

- We will compare neighbourhoods using the criteria of competitor Asian restaurants and median household income.
- We will use k-means clustering to generate groups of neighbourhoods that have similar relationships to the above criteria.
- K-means clustering is relatively straightforward and easy to interpret.
- Our aim is to obtain a list of neighbourhoods with the optimum blend of higher household income and a mid level of competitors.
- Our process includes normalising our dataset using Standard Scaler, then using the Squared Error method to assess the optimal number of clusters.
- Number of clusters to be used in the k-means model is established at k=6.

Results

- We plot the results of our k-means clustering algorithm on a map.



Results (2)

- Green circle cluster: Our optimum cluster

	Cluster Label	Neighbourhood	Borough	Latitude	Longitude	Median Household Inc	No. of Asian Restaurants
0	4	Balham	Wandsworth	51.445645	-0.150364	53420.0	-0.199482
1	4	Barnes	Richmond upon Thames	51.471896	-0.238744	55450.0	-0.588472
6	4	East Sheen	Richmond upon Thames	51.462371	-0.267094	56090.0	-0.977462
7	4	Eel Pie Island	Richmond upon Thames	51.445126	-0.325408	51380.0	-0.199482
13	4	Kew	Richmond upon Thames	51.480663	-0.291929	52130.0	-0.588472
19	4	Mortlake	Richmond upon Thames	51.468069	-0.266940	50100.0	-0.588472
20	4	Nine Elms	Wandsworth	51.478743	-0.136263	55380.0	-0.199482
30	4	St Margarets	Richmond upon Thames	51.456709	-0.322412	52870.0	0.189508
31	4	Strawberry Hill	Richmond upon Thames	51.438592	-0.339937	51380.0	-0.199482
32	4	Teddington	Richmond upon Thames	51.427784	-0.333653	50600.0	-0.588472
35	4	Twickenham	Richmond upon Thames	51.446744	-0.328189	51350.0	-0.199482
36	4	Wandsworth	Wandsworth	51.457027	-0.193261	51420.0	-0.199482

- All neighbourhoods in this cluster have median household income above £50,000.
- In this cluster, no neighbourhood has more than 3 Asian restaurants within 1km radius; the average for this cluster is 1.5.

Results (3)

- The remaining clusters offer sub-optimal outcomes.
- Orange circle cluster
  Higher income, higher number of competitors
- Dark blue cluster
  Mid income, higher number of competitors
- Red cluster
  Mid income, mid number of competitors
- Light blue cluster
  Mid income, lower number of competitors
- Purple cluster
  Lower income, lower number of competitors

Discussion

- We now have our list of neighbourhoods that offer the optimum blend of household income and number of competitors.
- Our preferred list includes neighbourhoods such as Balham and Wandsworth (town centre). These are well connected to Central London by train or underground and are popular with young professionals.
- Our list also includes neighbourhoods such as Twickenham and Teddington. They are further out from Central London but are still well connected by train and are popular with commuters and families.
- The neighbourhood Nine Elms is also in our list. This is a large redevelopment zone with new apartments, commerce and a soon to be completed London Underground station.
- All of the above neighbourhoods will likely have a different mix of potential customers.

Conclusion and next steps

- In this study, we've grouped neighbourhoods in our target area, South West London, according to what we believe are the key characteristics household income and number of competitor restaurants.
- Using k-means clustering, we've identified a list of preferred locations for opening a restaurant specialising in East & South East Asian cuisine.
- These preferred locations have higher household incomes meaning larger potential base of customers with money to spend on casual dining.
- These preferred locations may already have some competitors, indicating some potential for increased demand but not yet in a crowded market.

<u>Next steps</u>

- Our analysis could be refined further to demonstrate how varying the radius around each neighbourhood affects the results. Or possibly comparing the ratio of Asian restaurants to all restaurants in each neighbourhood
- As these preferred locations will all have their own potential customer mix, we recommend further demographic analysis.
- It is also recommended to conduct further analysis making an assessment of potential footfall in these locations. Close proximity to train stations, entertainment venues and tourist attractions can attract casual diners.
- Other factors such as rental prices, property availability, business rates all need to be considered as well.