Capstone project - Capstone London Restaurant

Problem & Background

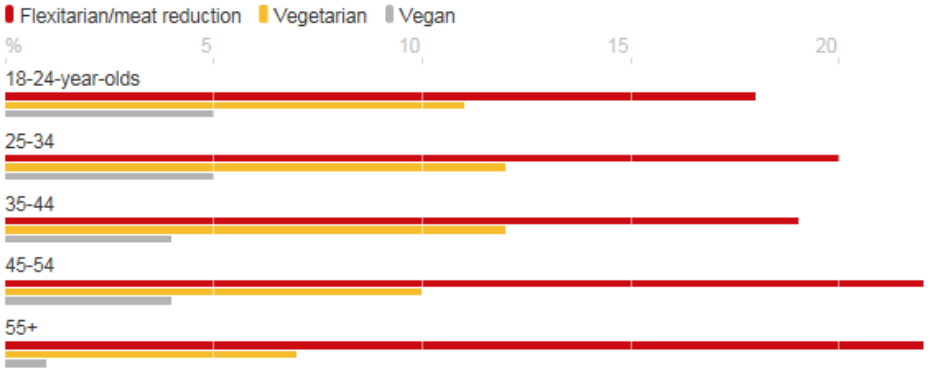
Define a problem where you need to leverage the Foursquare location data. Describe the audience and why they would care about the problem.

My client is looking to establish a new vegan restaurant business in London. The business will prepare vegan meals and deliver to homes, primarily within a 2km radius. The location of the business is key to its success, given the local nature of the business.

The client has chosen London as their target city, due to its young, multicultural and fast growing population, and wishes to identify the best area to set up, based on:

- Density of young residents;
- Propensity of local residents to spend on take-away food; and
- Economic growth prospects of the area.

The vegan restaurant sector is a fast growing market, and is particularly in demand with the younger demographics (18-34 year olds), as evidenced by research from the high-end supermarket chain Waitrose:



Guardian Graphic | Source: Waitrose food and drink report 2018-19

The client has requested an analysis of London to help them identify potential areas to establish the business. The analysis will be based on public data to identify postcodes with a large youth population, high consumer expenditure on food and takeaways and high growth projections. Foursquare will then be used to assess existing competition in these areas so that the client can judge where they are likely to be most successful.

For the purpose of this analysis, we will first identify London boroughs which best meet the three criteria set out by the client, and then look at a postcode level to assess which areas have the highest expenditure on take-away food. We can then select from these postal areas based on the Foursquare competitor data for 'Vegetarian & Vegan Restaurants'.

The analysis could later be extended to other cities or looking at the three criteria at a more granular postal level. This approach could also be used for other clients looking at different types of restaurant.

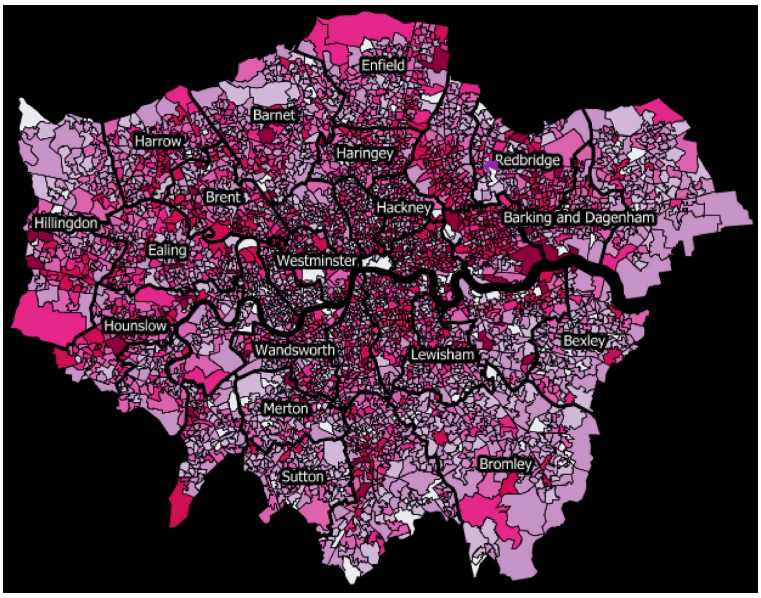
Data and how it will be used to solve the problem

Describe the data you will use to solve the problem. Need to include Foursquare data. Explain and discuss, with examples, data you will use.

There are numerous sources of data on London demographics, economics and geography.

For this project, I will need data at a UK Borough level on population by age group, household expenditure on food and the local economic growth. I will then need data at a more granular postal code level showing spending on takeaway food. For both boroughs and postal areas, I will need data to map to longitude/latitude coordinates, in order to plot maps of the analysis.

Mango Map provides useful demographic data on London (https://mangomap.com/demographics/maps) sourced from the Greater London Authority's Map Portal. The London Demographics Map Portal is a free and open data-sharing portal for demographic data relating to the UK capital.



The Portal provides a csv table with the following fields:

lsoa11cd	object	hholds	int64	a_h_w_chil	int64
lsoa11nm	object	avhholdsz	float64	a_h_wo_chi	int64
msoa11cd	object	a_pop	int64	a_sing_par	int64
msoa11nm	object	a_p0_15	int64	a_h_w_one	int64
lad11cd	object	a_p16_29	int64	a_h_other	int64
lad11nm	object	a_p30_44	int64	a_h_w_ch_1	float64
rgn11cd	object	a_p45_64	int64	a_h_wo_c_1	float64
rgn11nm	object	a_p65o	int64	a_sin_parp	float64
usualres	int64	a_working_	int64	a_h_w_onep	float64
hholdres	int64	a_area_h	float64	a_h_other_	float64
comestres	int64	a_peo_per_	int64		
popden	float64	a_househ	int64		

The fields we need are lad11nm (Local Authority District 2011 name) and a_p16_29 (Population of 16-29 year olds). We will use this age category as a close proxy for the desired 18-34 category.

For the economics of each borough, the UK Office for National Statistics (ONS) publishes lots of relevant data.

Office for National Statistics

In particular, it provides comprehensive country-wide data on gross disposable household income (GDHI) per head for NUTS3 local areas (boroughs) at https://www.ons.gov.uk/economy/regionalaccounts/grossdisposablehouseholdincome/bulletins/regionalgrossdisposablehouseholdincomegdhi/1997to2016 .

The ONS site provides a CSV table with 'NUTS3' local area name and gross disposable household income for each year from 1997 to 2017 (with a column for each year). We will just use the AREANM and the 2017 columns.

The third key source will be the London Government DataStore.

LONDON DATASTORE

The London DataStore has current and projected data on expenditure by consumers in each London postal area on a range of categories, including Food, Restaurants and Takeaway food at https://data.london.gov.uk/dataset/london-consumer-expenditure-estimates-2011-2036. This will tell us both which areas spend the most on take-away food and which areas have the greatest growth prospects.

The site has a number of databases on London Consumer Expenditure projections from 2011 to 2036, each with a different level of granularity. The data was produced in 2011 so is a bit old, but the provision of projections gives us a proxy for 2019 and an approximation of 2019-2029 10-year growth prospects.

The data is provided as a spreadsheet with a worksheet for each city ('Greater London') for the regional level data and a worksheet for each Borough for the Postal Base level database. For the regional level database, the sheet has borough, sector and a column for expenditure projected for each year. The sectors covered are:

- Food
- Non-alcoholic beverages
- Alcoholic beverages
- Tobacco
- Clothing and footwear
- Actual rentals for housing
- Imputed rentals for housing
- Maintenance and repair of the dwelling
- Water supply and miscellaneous services relating to the
- Electricity, gas & other fuels
- Furniture & Textiles
- Household Goods and Services
- Medical Products
- Medical Services
- Purchase of vehicles
- Operation of personal transport equipment
- Transport services
- Postal services
- Telecommunications Services
- Audio-visual
- Other major durables for recreation and culture
- Other recreational items and equipment

For the Postal Base data, each borough worksheet lists the abbreviated postcode (postcode area plus the first number of the postcode district), sector and a column for expenditure projected for each year. Sectors include:

- Convenience
- Comparison – Bulky
- Comparison - Not Bulky
- DIY
- Gardening
- Accommodation Services
- Restaurants and Cafes
- Takeaway / Snack Spending
- On Licence (i.e. Pubs & Wine Bars)
- Leisure
- Other Goods and Services
- Other Spending (Mostly Household related, Health and Education)

Finally, we will use the ONS Postcode Directory from Open Postcode Geo, provided on the 'get the data' site (https://www.getthedata.com/open-postcode-geo), to get data to map boroughs and postal areas to longitude/latitude coordinates. The table is very big and lists every postcode in the UK. We will need to map our abbreviated postcodes to this data.

Foursquare will then provide the data we need to establish what other Vegetarian & Vegan restaurants exist in each area. Foursquare includes a category for Vegetarian & Vegan Restaurant venues (category ID: 4bf58dd8d48988d1d3941735). We will look for any such venues within 2km of our selected postal areas, as this is the typical delivery range for such a business.