

2. Data

Datasets used in this analysis are from three sources: Foursquare location data, Wikipedia and The London Datastore. The data extracted from these sources will become the inputs of the machine learning model to classify the ratings of grocery stores.

2.1 Foursquare API

Foursquare API is the main source of data in this report. It is a popular location data vendor that provides venues locations, user ratings and tips, etc. It is utilised to find the coordinates of London's Tesco stores, explore the neighbourhood of these stores and obtain the ratings of these stores. Here is an example of Foursquare data:

id	name	categories	referralId	hasPerk	location.address
4b8d4de5f964a52038f332e3	Tesco Metro	[{'id': '4bf58dd...	v-1595789668	False	17-25 Regent St
4bbe6094006dc9b64512fb3f	Tesco Express	[{'id': '4bf58dd...	v-1595789668	False	1-4 Charing Cross
4acf3e3ef964a520dcd220e3	Tesco Metro	[{'id': '4bf58dd...	v-1595789668	False	22-25 Bedford St, Covent Garden

2.2 Wikipedia

The “List of areas of London” (https://en.wikipedia.org/wiki/List_of_areas_of_London) is used. It provides the areas and boroughs of London and the corresponding post codes, which will be used for the locations of the grocery stores in the model. Here is an example of the Wikipedia data:

Location	London borough	Post town	Postcode district
Abbey Wood	Bexley, Greenwich [7]	LONDON	SE2
Acton	Ealing, Hammersmith and Fulham[8]	LONDON	W3, W4
Addington	Croydon[8]	CROYDON	CR0
Addiscombe	Croydon[8]	CROYDON	CR0

2.3 The London Datastore

The London Datastore is a free and open data-sharing portal where anyone can access data relating to the capital. The median house prices of each areas in London is downloaded from this website (<https://data.london.gov.uk/dataset/average-house-prices>). Here is an example of The London Datastore data:

New code	Ward name	Borough name	Year ending Dec 1995	Year ending Mar 1996	Year ending Jun 1996	Year ending Sep 1996	Year ending Dec 1996	Year ending Mar 1997	Year ending Jun 1997
E09000001	City of London	City of London	-	-	-	-	-	-	-
E05000026	Abbey	Barking and Dagenham	53,000	50,000	50,000	53,500	49,998	46,995	45,000
E05000027	Alibon	Barking and Dagenham	45,000	45,000	45,500	45,998	45,968	46,148	48,000
E05000028	Becontree	Barking and Dagenham	49,000	50,000	50,000	50,000	51,000	53,995	56,250
E05000029	Chadwell Heath	Barking and Dagenham	59,000	57,000	56,000	58,000	58,000	58,250	59,738