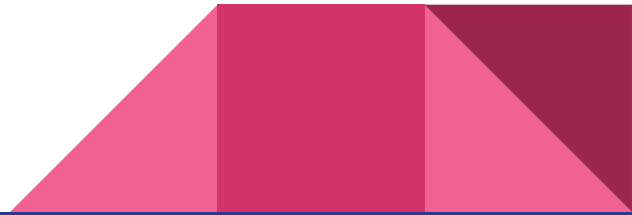
Walking Areas of London for Foodies on a Budget

Applied Data Science Capstone Project

Author: Mita Naicker

Introduction & Scenario

- London is a tourist hotspot with many restaurants & cuisines to choose from
- Choosing which restaurants to visit can be challenging when on a budget
- Scenario: tourist has put 3 days aside for “gastronomic tourism”, with the aim of visiting at least 3 restaurants in each area per day
- 3 main criteria per area: high rated restaurants, low price level, walking distance between restaurants
- Anyone interested in eating great food in London is interested in results



Data Sources & Acquisition

1. Google Places API: place search performed for restaurants in 3 km radius around London and also 3 km radius around 4 points N,E,S,W of London
2. Foursquare Places API: venue search performed for restaurants in 3 km radius around London and also 3 km radius around 4 points N,E,S,W of London; then venue details retrieved for each venue_id

Fields in Resulting DataFrames:

	Name	Rating	Price_Level	Latitude	Longitude	Address	Source
185	222 Vegan Cuisine	8.8	2	51.486023	-0.202981	222 North End Rd, Hammersmith, London W14 9NU	googlemaps
20	34 Mayfair	8.8	4	51.510302	-0.152263	34 Grosvenor Square Entrance on, S Audley St, ...	googlemaps

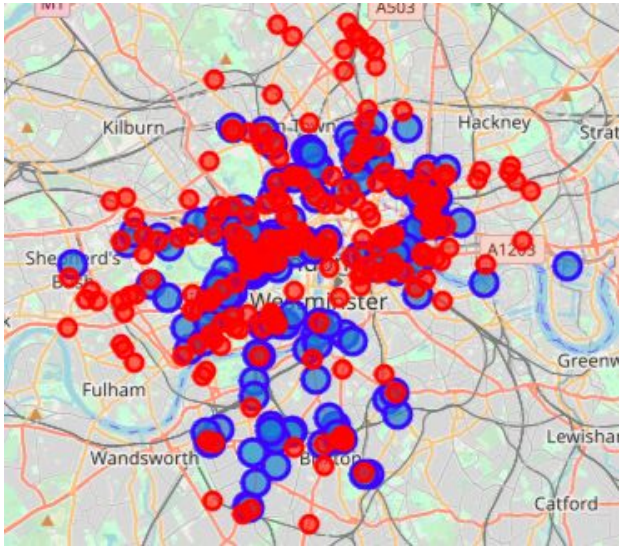
Data Preparation & Cleaning

- Drop duplicates: some overlap with 3km radiuses so duplicate restaurants needed to be dropped
- Price_Level field cast from object to numeric
- Adjust Ratings values: Google ratings out of 5.0 and Foursquare out of 10.0 so Google values multiplied by 2
- 2 dataframes were concatenated together after cleaning, for a total of 396 rows

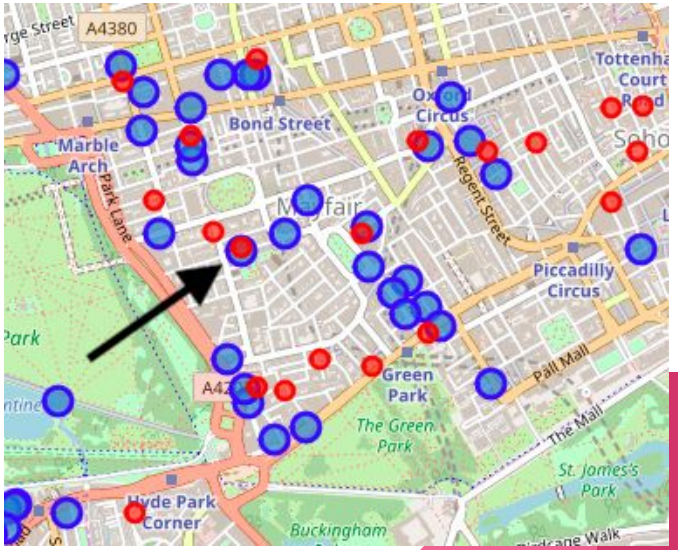
Exploratory Data Analysis: Data Visualization

- Very few duplicates visible between Google & Foursquare results; duplicates kept as they show variance in the data

All Data (Google in red; FS in blue)



Zoomed in: Duplicates found



Exploratory Data Analysis: describe()

- describe() results: lower ratings contained in Foursquare data

Foursquare

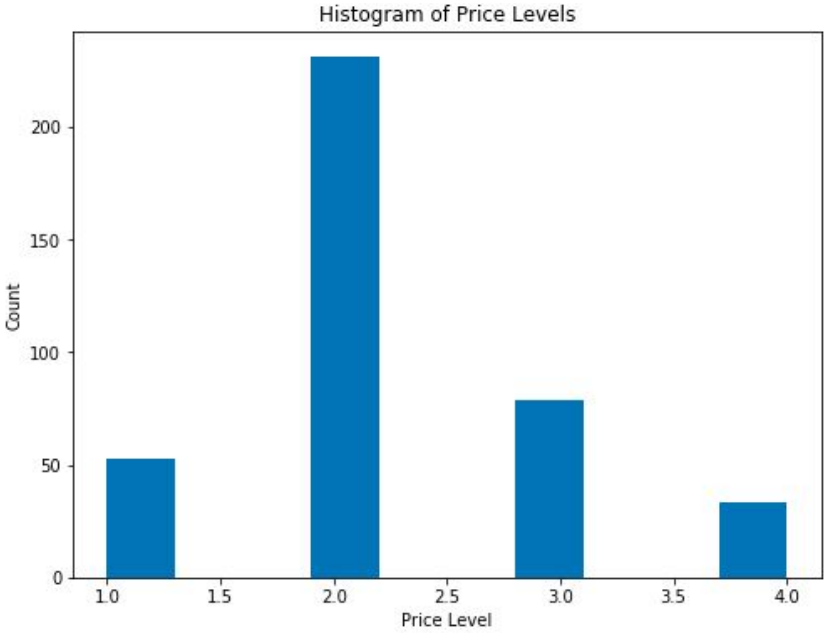
	Rating	Price_Level
count	196.000000	196.000000
mean	8.319898	2.178571
std	0.822418	0.739889
min	5.800000	1.000000
25%	7.800000	2.000000
50%	8.500000	2.000000
75%	9.000000	3.000000
max	9.600000	4.000000

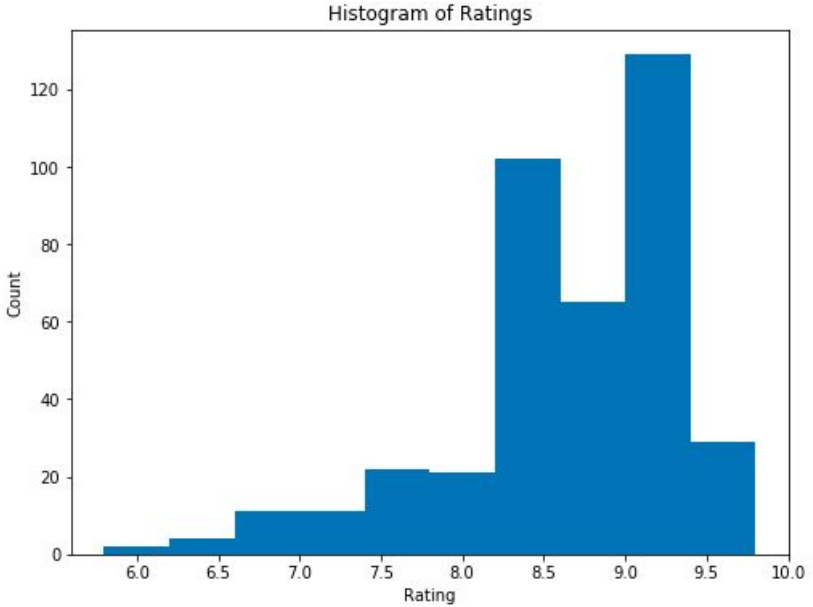
Google

	Rating	Price_Level
count	200.000000	200.000000
mean	8.893000	2.285000
std	0.339599	0.822898
min	8.000000	1.000000
25%	8.600000	2.000000
50%	9.000000	2.000000
75%	9.200000	3.000000
max	9.800000	4.000000

Inferential Statistical Testing: Histograms

- Good number of lower price levels and high ratings across all data





Inferential Statistical Testing: Scatterplot

- Highest rated restaurants actually fall in lowest price levels

Machine Learning: DBScan

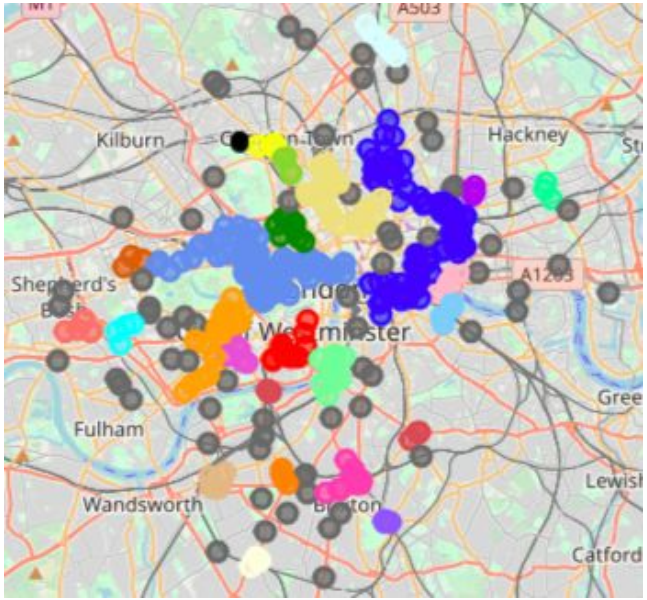
- Walking distance an important criteria for results
- DBScan clusters based on physical (geodetic) distance (eps) between each point and minimum cluster size (min_samples)
- Other clustering algorithms such as k-means minimize variance, not geodetic distance
- DBScan parameters chosen:
  - db = DBSCAN(eps=0.15, min_samples=3, metric ='euclidean').fit(Clus_dataSet)

DBScan Results

- 329 of 396 points were clustered (67 categorized as noise/outliers)
- 26 clusters created from 329 points
- 3 clusters met criteria of high overall rating >= 9.0 and low price <= 2.0

DBScan Clusters on Map

All clusters



Clusters Meeting Criteria

Results: Recommended Restaurants/Areas

- Hammersmith, Finsbury Park and Camden Town are chosen areas

	Name	Rating	Price_Level	Latitude	Longitude	Address	Source	Clus_Db
37	Chez Abir Lebanese Restaurant	9.0	2	51.497064	-0.212609	34 Blythe Rd, Hammersmith, London W14 0HA	googlemaps	13
108	Los Molinos	8.8	2	51.497394	-0.222459	127 Shepherds Bush Rd, Hammersmith, London W6 7LP	googlemaps	13
140	Pentolina	9.6	2	51.497612	-0.217451	71 Blythe Rd, Hammersmith, London W14 0HP	googlemaps	13
173	The Bird in Hand	8.8	2	51.499733	-0.215419	88 Masbro Road, Masbro Rd, London, Brook Green...	googlemaps	13
55	Dotori	9.2	2	51.564676	-0.105179	3 Stroud Green Rd, Finsbury Park, London N4 2DQ	googlemaps	14
56	E-Mono	9.2	1	51.565849	-0.106959	13 Stroud Green Rd, Finsbury Park, London N4 2AL	googlemaps	14
90	Il Cavaliere Italian Restaurant	9.4	1	51.562691	-0.100947	81 Blackstock Rd, Finsbury Park, London N4 2JW	googlemaps	14
141	Petek Restaurant	9.2	2	51.568405	-0.110101	96 Stroud Green Rd, Stroud Green, London N4 3EN	googlemaps	14
199	Yildiz Restaurant	9.0	2	51.561209	-0.098796	163 Blackstock Rd, Finsbury Park, London N4 2JS	googlemaps	14
13	Asakusa	8.8	2	51.534021	-0.138289	265 Eversholt St, Kings Cross, London NW1 1BA	googlemaps	25
50	Daphne	9.6	1	51.537759	-0.140337	83 Bayham St, Camden Town, London NW1 0AG	googlemaps	25
353	The Blues Kitchen	8.6	2	51.537182	-0.141099	111-113 Camden High St, Camden Town, Greater L...	foursquare	25

Conclusion and Future Direction

- Hammersmith, Finsbury Park and Camden Town areas of Greater London offer up some fantastic cheap eats, all within walking distance of each other
- Would be interesting to see how k-means results would compare
- Could ‘metric’ parameter be set to an orthodromic distance? Results might be better.
- Adding data from another large source of user-rated restaurants, such as TripAdvisor, would add more credence to results