Predicting best London borough for Gordon Ramsay's new upscale restaurant

Optimal Italian restaurant placement

Introduction

- Gordon Ramsay is starting a new fancy Italian restaurant in London, but where?

- Restaurant competition in London is high and we need to ensure plenty of potential customers – with thick wallets

- Data science can be used to compare Londons 32 boroughs based on:
  - Competitiveness (number of Italian restaurants per square km)
  - Customer spending proclivity (aggregated income per square km)

Datasets & APIs

- Wikipedia page: List of London Neighborhoods with postal codes
  - https://en.wikipedia.org/wiki/List_of_areas_of_London

- Greater London Authority: Average Income of Taxpayers, Borough and Population Density, Borough, datasets
  - https://data.london.gov.uk/dataset/average-income-tax-payers-borough
  - https://data.london.gov.uk/dataset/land-area-and-population-density-ward-and-borough

- Geocoder API: Called to get coordinates for adresses
  - https://geocoder.readthedocs.io/index.html

- Foursquare API: Called to get list of venues within specified range of a coordinate
  - https://developer.foursquare.com/

Methodology – Data Aquisition

- Income of Taxpayers and Population Density were downloaded straight into Pandas dataframes

- List of areas of London was scraped from Wikipedia using BeautifulSoup package

- Coordinates added by calling Geocoder for each neighborhood and borough

Methodology – Data Cleaning I

- Income of Taxpayers and Population Density datasets were filtered to only keep the following data for the London boroughs:
  - Median Household Income 2018, Population per km^2, Square kilometers
- Some neighborhoods were assigned to multiple boroughs
  - A function was written to assign venues to the closest of the boroughs which its neighborhood was assigned to, based on venue and borough coordinates
- Foursquare was called for each neighborhood coordinates, up to 500 results within 2 km
  - Duplicates were removed
  - Resulting dataframe had 28989 venues with 322 different categories

Methodology – Data Cleaning II

- Venues dataframe filtered for 'Italian' and 'Pizza' categories
  - Results 545 pizza places and 477 Italian restaurants

- Venues were grouped by borough
  - No results returned for four boroughs – these were removed

- Number of restaurants per square km and aggregated taxpayer income per square km calculated (Venue Density & Income Density) for each borough
 - Cleaned and ready dataframe header below:

	Borough	Count	Square kilometres	Population per km^2	Median Household Income 2018	Latitude	Longitude	Venue Density	Income Density
0	Barnet	103.0	86.7483	4577.02	28700	51.627300	-0.253760	1.18734	1.31361e+08
1	Bexley	3.0	60.5807	4126.71	26900	51.452078	0.069931	0.0495207	1.11009e+08
2	Brent	60.0	43.2326	7791.78	24700	51.609783	-0.194672	1.38784	1.92457e+08
3	Bromley	11.0	150.135	2216.23	32000	51.601511	-0.066365	0.0732674	7.09193e+07
4	Camden	76.0	21.7893	11594.5	37300	51.591180	-0.165040	3.48795	4.32477e+08

Results I

- Exploratory analysis by ranking boroughs on Venue Density (lower density is better) and Income Density (higher is better)

- Plotted on Folium heat map by venue rank (border color) and income rank (fill color) of markers

- Boroughs with border and fill on blue spectrum are promising (high on both rankings)
  - Ealing
  - Hackney
 - Lambeth
  - Waltham Forest
  - Westminster

	Borough	Venue Rank	Income Rank
0	Barnet	14.0	23.0
1	Bexley	4.0	27.0
2	Brent	16.0	14.0
3	Bromley	5.0	29.0
4	Camden	28.0	6.0
5	City of London	29.0	19.0
6	Croydon	6.0	26.0
7	Ealing	8.0	16.0
8	Enfield	7.0	28.0
9	Greenwich	12.0	18.0
10	Hackney	19.0	5.0
11	Hammersmith and Fulham	23.0	8.0
12	Haringey	24.0	11.0
13	Harrow	3.0	22.0

	Borough	Venue Rank	Income Rank
14	Hounslow	2.0	24.0
15	Islington	25.0	1.0
16	Kensington and Chelsea	27.0	2.0
17	Kingston upon Thames	1.0	20.0
18	Lambeth	17.0	7.0
19	Lewisham	20.0	12.0
20	Merton	11.0	17.0
21	Newham	18.0	13.0
22	Redbridge	9.0	21.0
23	Richmond upon Thames	10.0	25.0
24	Southwark	22.0	9.0
25	Tower Hamlets	26.0	3.0
26	Waltham Forest	13.0	15.0
27	Wandsworth	21.0	10.0
28	Westminster	15.0	4.0
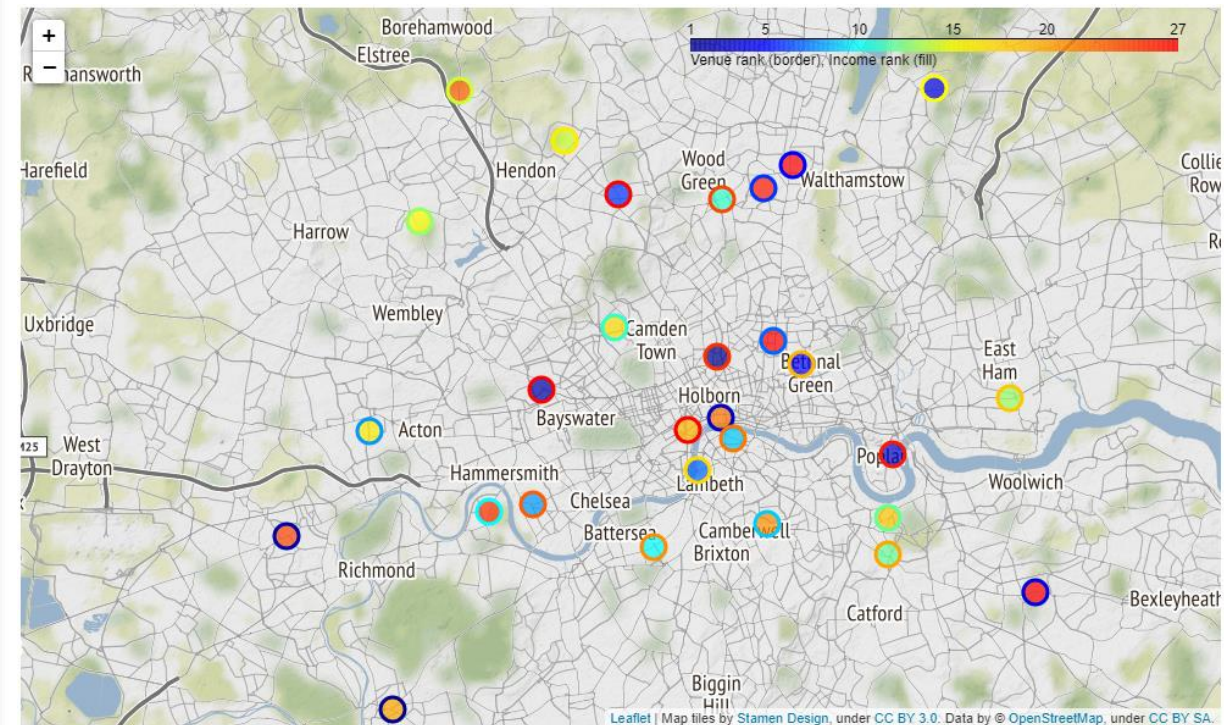
Results II

- Rankings were complemented by plooing the relationship between venue density and income density

- Linear regression trendline added
  - Best boroughs will be below line and far to the right: Few restaurants and high income compared to other boroughs

- Westminster, followed by Hackney and Lambeth are clear best candidates

- Islington possible candidate – high income

- Waltham Forest in bottom left has low income – not a candidate for high end restaurant

Scatter plot of italian restaurant competition versus customer spending proclivity
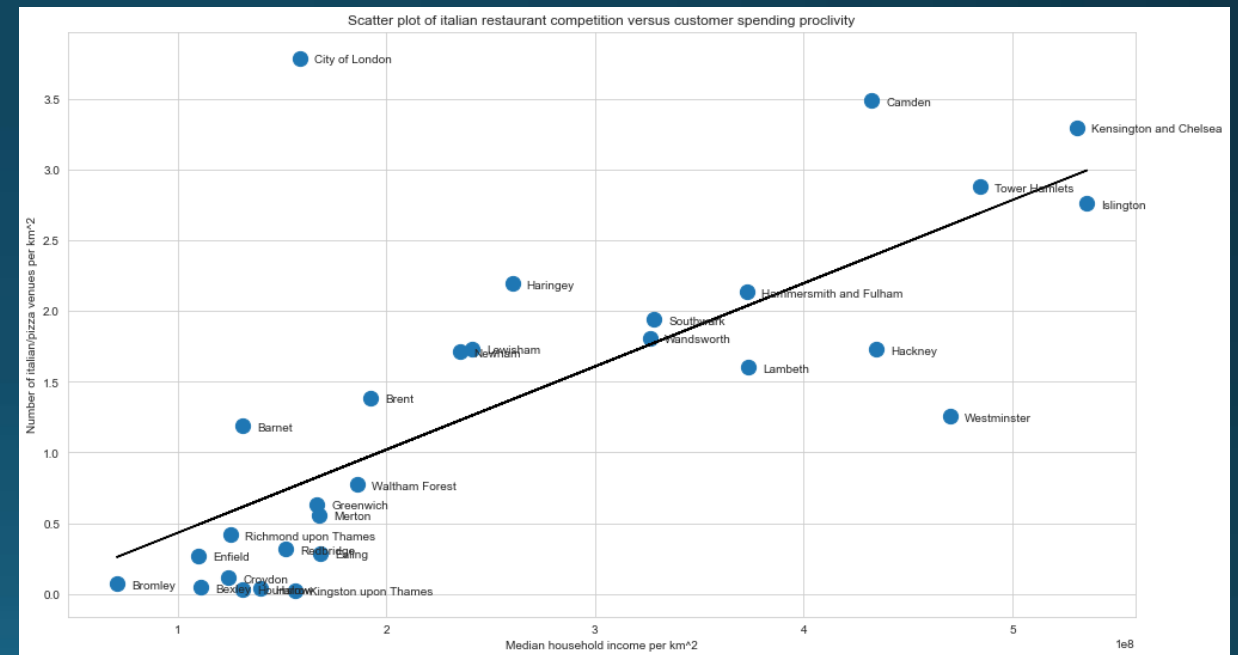
Conclusion

- Open data from the Greater London Authority, Wikipedia, Foursquare and Geocoder was used to find the optimal placement of a high-end italian restaurant in London

- Parameters used were number of competing Italian restaurants per square km and aggregated income per square km

- Analysis indicated Westminster being the best candidate, with Lambeth, Hackney and Islington as viable options

- Further analysis should be conducted on the neighborhood level (as opposed to borough) and using a greater range of parameters (e.g property cost, competing non-Italian restaurants)