A Gastronomic Exploration

Maria Walton Campagna
DSI11 - Capstone Project

Contents

“We all eat, and it would be a sad waste of opportunity to eat badly.”

—Anna Thomas

“One cannot think well, love well, sleep well, if one has not dined well”

—Virginia Woolf

Project Aims & Approach:

London Restaurant Reviews
Bias Analysis

Create a model that predicts ratings bias
(measured as critic vs user scores)

and identifies the key features influencing bias
(hypothesised to be influenced by location)

Take a different angle to this classic problem, building my own data set.

Approach

Individual user review score prediction from reviews is a classic data science project. I wanted to look for a new angle…

Scrape data from the web - primarily TimeOut and Google

Exploration of APIs and existing datasets e.g. Yelp, Foursquare

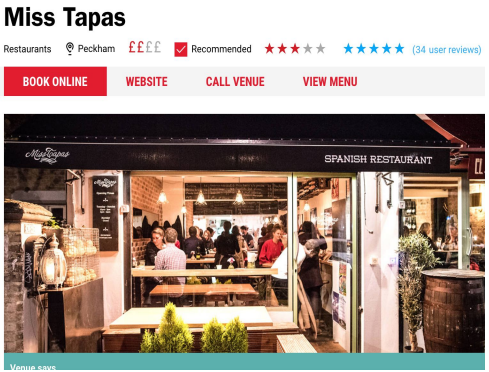
Identify additional location data to create new features

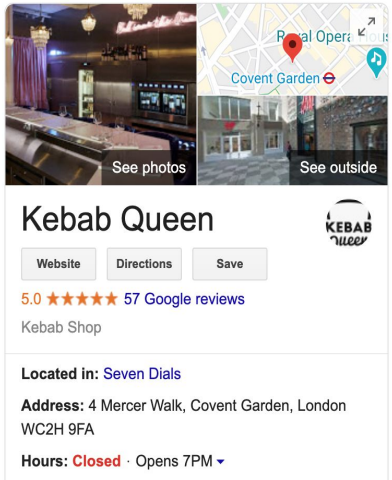
Explore the data and test the location bias hypothesis

Clean and process the data and then assess the best modelling approach

Data and Modelling

Data Collection - Web Scraping





Data

TimeOut

Core data with list of c.4,000 restaurants
1,200 critic scores
3171 restaurants with user scores

Google

Average User reviews & number of ratings for c3,700 restaurants
Category & budget info

ONS Data

2011 Census
Regional Classifications
Energy consumption

Other Location Data

Tube Stations
Tourist Attractions
Online postcode converter tool

Other Data

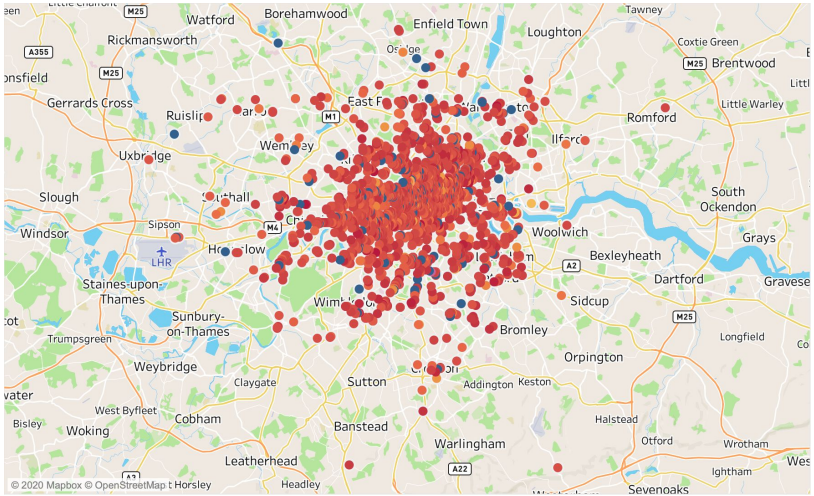
Foursquare
FSA
Yelp

Feature Engineering

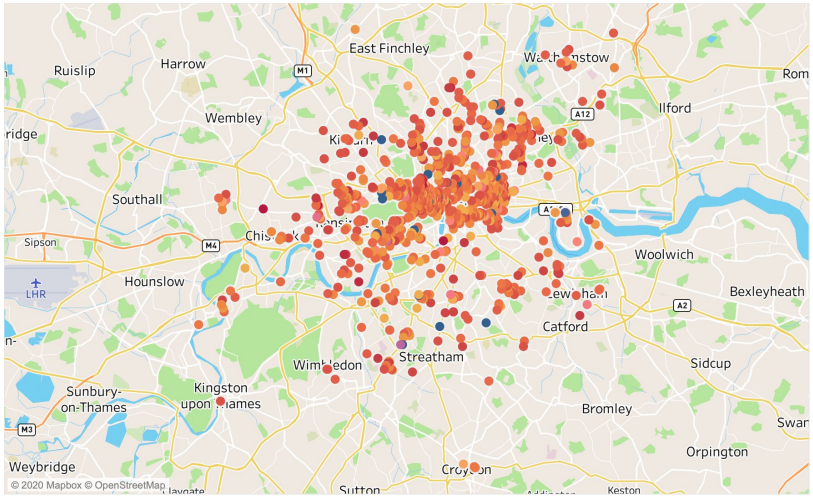
Sparsity
Nearest Tube
Proximity to Tourist Sites
Category reduction
Clustering

London Restaurants Map

All Restaurants



Restaurants with Critic Ratings



Google user average scores

1.000 5.000

Initial Exploratory Data Analysis (EDA)

Average Ratings:

- Timeout Critic (reviewer)	3.45
- Timeout Users	4.18
- Google Users	4.3

Average Users reviewing a restaurant

- Time Out	12.4
- Google	520.9

Top Rated Restaurants by Google Users (& Critic)

Endo at the Rotunda
Kebab Queen
(High end tasting menus - but only 50-60 reviewers)

Worst and Biggest Differential

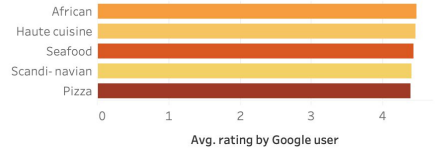
POTUS - Themed hotel bar & grill (1* critic score, 5* users)

Initial Observation

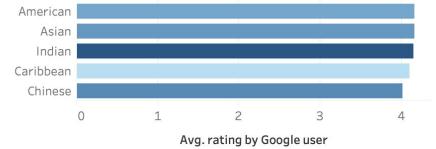
More reviewers tends to increase the average rating for a restaurant and lower the spread of rating scores across all restaurants

Initial Exploratory Data Analysis (EDA) - Cuisines

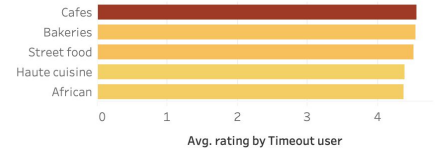
Top Rated Cuisines by Google Users



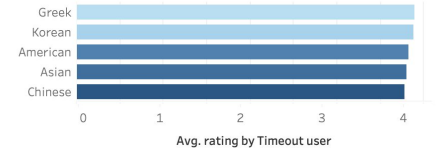
Bottom Rated Cuisines by Google Users



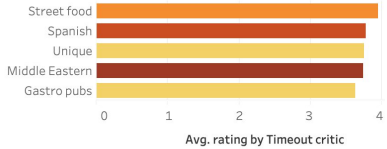
Top Rated Cuisines by Timeout Users



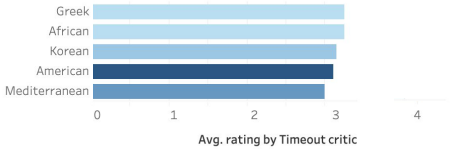
Bottom Rated Cuisines by Timeout Users



Top Rated Cuisines by Timeout Critic



Bottom Rated Cuisines by Timeout Critic



Average Ratings:

- Timeout Critic (reviewer)	3.45
- Timeout Users	4.18
- Google Users	4.3

Cuisines engineered from grouping of over 180 categories

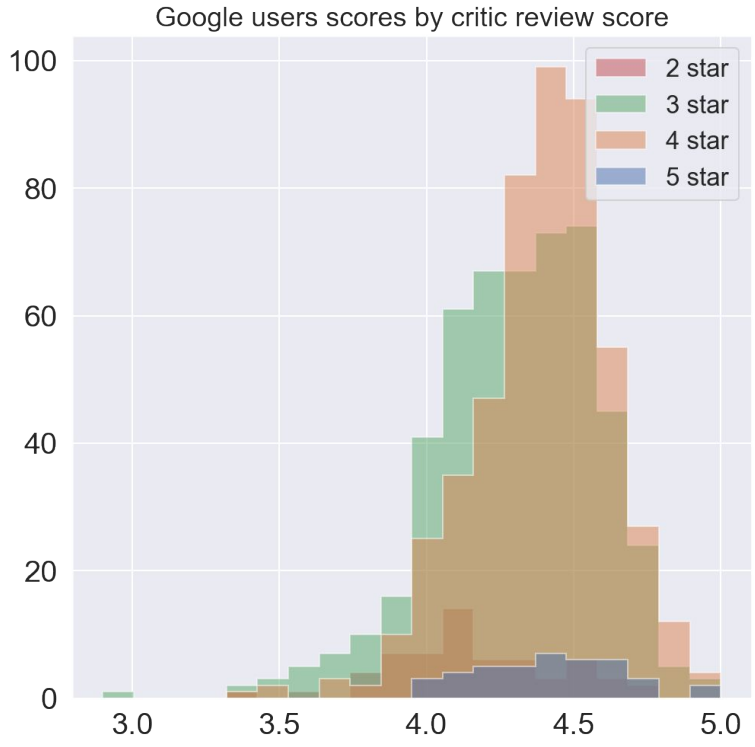
Greater agreement between Users than Critic on top and bottom rated cuisines, but not full agreement

Top Rated Cuisines - Users
Haute Cuisine, African

Top Rated Cuisines - Critic
Street food, Spanish

Bottom Rated Cuisines
American in Bottom Rated Cuisines for all three
Chinese in Users' Bottom Rates but not Critic's
African in Critic's Bottom Rated but in Users' Top Rated

Further EDA - Setting Model Target



Target for Modelling:
Google users scores

Features:
16 from TimeOut covering:
- Location (address, area, nearest tube, latitude, longitude)
- Category (180 distilled to 40 cuisines)
- Budget (scale of 1 to 4)
- Reviews overview info (metadata)

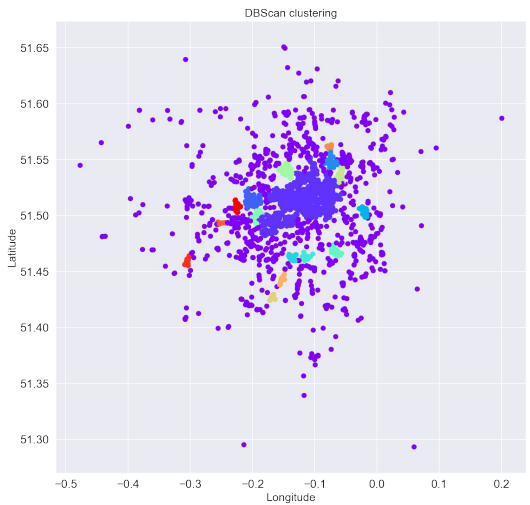
4 from Google covering
- Reviews overview info (metadata)
- Budget
- Category (260 of them)

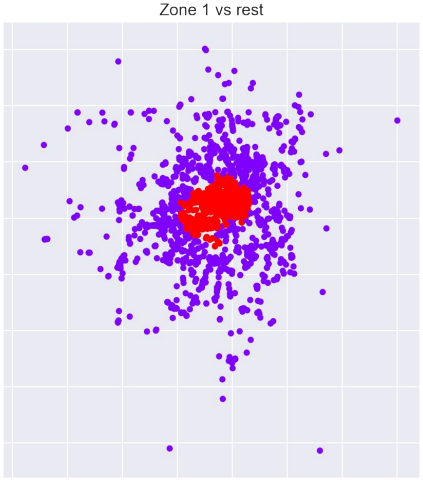
Location data
- Tube stations & locations - almost 200 unique values in each
- Zones, districts, cluster (engineered)
- Proximity to Tourist sites (engineered)
- Sparsity (engineered)

Lots of Null Values, Collinearity and Feature Engineering

Modelling - Unsupervised

Clustering Algorithms used to search for a new restaurant feature to better represent its location than London Transport Zones





CLUSTERING MODELS
DBScan
Agglomerative Clustering

SCORING CRITERIA
Adjusted precision and accuracy score to predict zone 1 vs the rest.
Less than 30 clusters

CHOSEN CLUSTER
DBScan
95% f1 score
93.9% precision (accuracy to zone 1)
17 clusters

Modelling - Supervised

Regression Models
Linear Regression
(ElasticNetCV - Ridge)
Decision Tree
Bagging
Random Forest
Support Vector Model
(Linear & SVR)

Classification Models
Logistic Regression
KNN
SVC Linear
Decision Tree

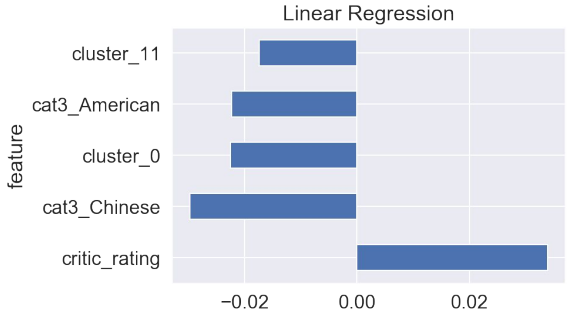
BEST SCORES	R2 (CV/Best)	'Accuracy' (Train)	'Accuracy' (Test)
Regression (Linear)	0.10	-	-
Regression (SVR)	0.09	-0.22	0.085
Classification (Logistic Regression)*	-	0.62	0.60

Classification Baseline - 0.50. Regression 'Accuracy' R2, Classification 'Accuracy' ROC-AUC

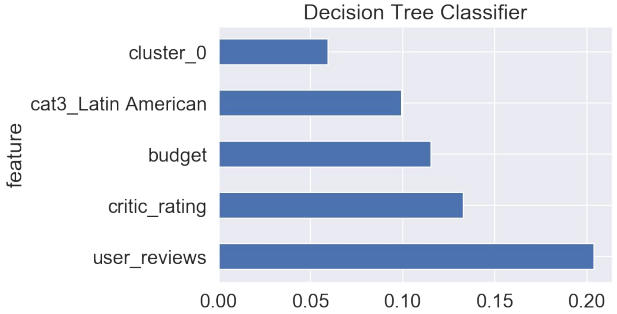
Most Important Features by Model

Inconsistencies between models seen when comparing Cuisines

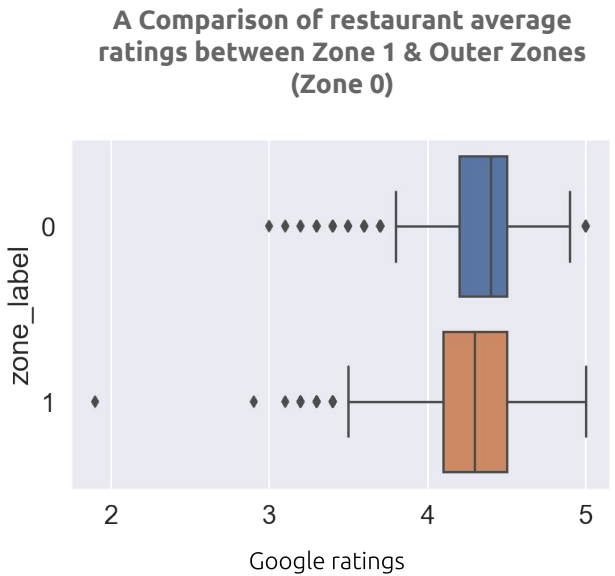
Cluster location feature one of the stronger predictors in all models



Critic rating of a restaurant the best predictor of Google User ratings for Linear and Logistic Regression but not for the Decision Tree model.



Location Bias Review



- Expectation from EDA and modelling that location does impact a restaurant's average ratings.
- Tests on each of the 3 ratings sources showed a statistically significant difference between the the average restaurant scores of those located in Zone 1 vs Outer Zones.
- Ratings were lower on average in Zone 1

Insights and Next Steps

Discoveries and Limitations

RATING SCALES

Average restaurant user ratings were narrowly spread and higher on average than expected - watch out for 'high' scores.
Critic ratings were spread across the 1-5 scale.

RATING SUBJECTIVITY

Little consistency between ratings and preferences from the 3 different sources. This reflects the expectation that the reviewers profiles across the 3 sources are different.

METADATA

Proximity to Zone 1 did show an influence on a restaurant's average rating.
Certain Cuisines are more popular with London restaurant reviewers than others.

BIAS & DATA SCIENCE

Defining bias in ratings is not easy - hard to find a 'baseline'.
Data collection, analysis and cleaning really is at least 80% of data science.

Future Possibilities

DATA

Gather different user (& critic) scores data to compare e.g. Tripadvisor. Update the metadata used.

NLP

Natural Language Processing Sentiment Analysis
e.g. hot-dinners.com, detailed user reviews.

RECOMMENDERS

Explore building a cold start recommendation engine with metadata.

Thanks!

Any questions?

Credits

Thank you to General Assembly and my Data Science Immersive peers for all their support and valuable feedback.

Presentation template by Slidesgo

Icons by Flaticon

Images & infographics by Freepik