Battle of the neighborhoods

London 2020

MARCH 9

Authored by: Pratish Patni

Contents

Categorization of London restaurants based on food safety, customer likes and price

Introduction

There are basically two types of people in this world based on how they look at the food they eat. People who eat to live, and people who live to eat 😋

While eating out has become an everyday reality in most urban cultures in the world today, restaurants are increasingly trying to suit their menu to cater to both the categories of people.

For the people who eat to live, organic food from local farmers with more stress on a vegetarian/vegan and other healthy menu items is being laid. For the people who live to eat, exotic food from faraway places including international menu is increasingly becoming popular.
In all the scenarios, it is very imperative that the food is not only tasty but also healthy.

How many times have you had food in a restaurant and wished you had never been there? There are many things which can put you away from a restaurant and food safety must be the most important of it. A bad tasting food can leave you in a bad taste and mood, but an unsafe food will cause danger to your health and can take you to a hospital. Hence, food safety should be of prime importance along with the taste and value for money.

Today, thanks to the various apps, customers have the power to rate various restaurants based on the taste of food, service provided, cleanliness, quality etc. This gives a very good outlook to other users of the app to evaluate a restaurant before going there.

Government agencies also put in effort to define food safety norms, exercise control and give safety ratings to places where food is procured, processed and served.

Problem statement

Can we analyze the various restaurants located in the neighborhoods of London based on the food safety rating received by them via an authorized government agency, online user ratings (likes) and price category, and publish this data for customers to choose the neighborhoods where eating out is the best option in London?

Stakeholders

This report will help the following stakeholders and is majorly meant for their consumption:

1. General population – of people living in London and visitors to London who can benefit from this report and choose a safe, healthy and popular place to have a meal in their neighborhood/vicinity
2. Restaurant owners – who can evaluate where they stand in terms of safety, taste and price of the food they serve. They can learn from restaurants in other neighborhoods which fall in a better category
3. Government officials – who can identify neighborhoods where poor category of restaurants is largely located and deal with them using mass communication programs and workshops so they can improve their standards of hygiene and food/service quality.

Background

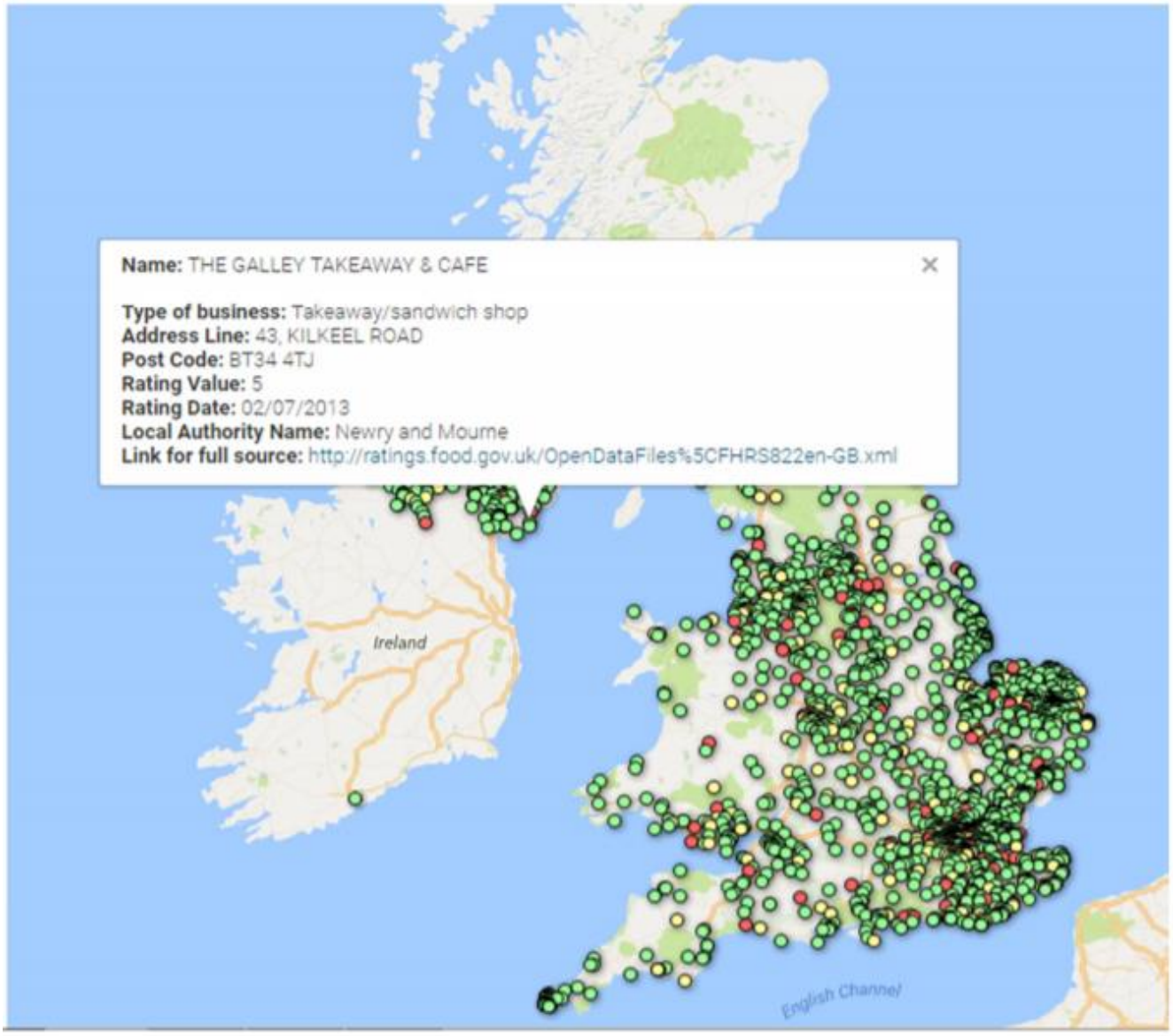
FSA – Food Standards Agency, United Kingdom

As per FSA, they are '*an independent Government department working across England, Wales and Northern Ireland to protect public health and consumers' wider interests in food. They make sure that food is safe and what it says it is.'*

The **Food Hygiene Rating Scheme** of the FSA is a *scheme that helps you choose where to eat out or shop for food by giving you clear information about the businesses' hygiene standards*.
The scheme gives businesses a rating from 5 to 0 which is displayed at their premises and online so you can make more informed choices about where to buy and eat food.

5 – hygiene standards are very good
4 – hygiene standards are good
3 – hygiene standards are generally satisfactory
2 – some improvement is necessary
1 – major improvement is necessary
0 – urgent improvement is required

In 2013, The Guardian displayed on an interactive map, presented below, the restaurants with the food hygiene grades.



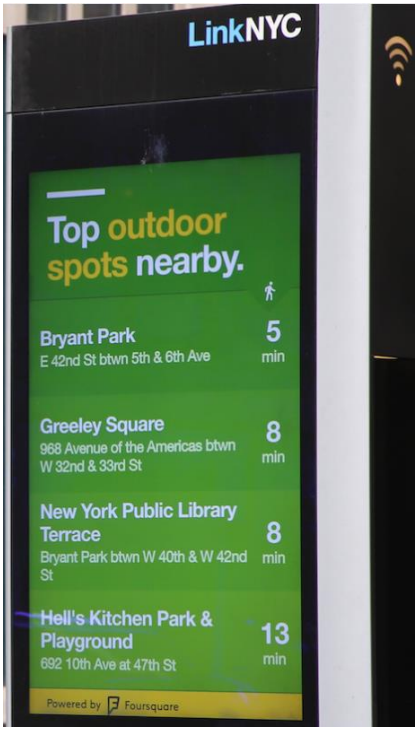
Foursquare API

Foursquare is a '*location technology platform dedicated to improving how people move through the real world*'.

As per Foursquare, '*Location is more than a data point. We believe that the places you go say a lot about who you are. Our technology's unparalleled sense of place and space has allowed us to help the world's leading brands and advertisers unlock valuable insights about their consumers and their businesses.*'

Developers and companies around the world have used Foursquare technology to create products that leverage the location data in a way that has helped and empowered people with detailed information about their surroundings.

Apps like Intersection tapped Foursquare's Places API to power the LinkNYC 'Popular Places' feature, which helps New Yorkers and visitors easily find trending spots nearby, delivering responsive, localized information to city dwellers in real time. 'Popular Places' surfaces restaurants, shops, outdoor spots, cultural institutions and nightlife venues that have the most check-ins within a 15-minute walking distance from a given Link's location, in real time.



One of the most popular use of Foursquare API is to explore restaurants in a particular neighborhood and check how good they are and what people are saying about them.

Generally, more the likes on a restaurant would mean the better and more popular the restaurant is.

In our analysis to produce this report, we will be using both FSA and Foursquare data to find a solution to our problem stated above.

Data description

Time to introduce the research methods and data sources used for the analysis and achieving the desired solution.

As mentioned above, the analysis is largely based on the food hygiene ratings given by Food Standards Agency (FSA) of the UK and Foursquare data of the restaurants in various neighborhoods of the city of London.

The data about the food safety ratings for establishments located in London can be obtained from https://ratings.food.gov.uk/open-data/en-GB
The data is updated regularly and currently has 1,804 establishments reviewed till last week, most of which are restaurants:

City of London Corporation (English language)	07/03/2020 at 12:00	1,804

The data is in XML file, and is typically in this schema format:

```
▼<FHRSEstablishment xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance">
  ▶<Header>...</Header>
  ▼<EstablishmentCollection>
    ▼<EstablishmentDetail>
      <FHRSID>294048</FHRSID>
      <LocalAuthorityBusinessID>PI/000008080</LocalAuthorityBusinessID>
      <BusinessName>1 Lombard Street, Restaurant</BusinessName>
      <BusinessType>Restaurant/Cafe/Canteen</BusinessType>
      <BusinessTypeID>1</BusinessTypeID>
      <AddressLine1>1 Lombard Street</AddressLine1>
      <AddressLine2>London</AddressLine2>
      <PostCode>EC3V 9AA</PostCode>
      <RatingValue>5</RatingValue>
      <RatingKey>fhrs_5_en-GB</RatingKey>
      <RatingDate>2019-12-19</RatingDate>
      <LocalAuthorityCode>508</LocalAuthorityCode>
      <LocalAuthorityName>City of London Corporation</LocalAuthorityName>
    ▼<LocalAuthorityWebSite>
        http://www.cityoflondon.gov.uk/Corporation/homepage.htm
      </LocalAuthorityWebSite>
      <LocalAuthorityEmailAddress>publicprotection@cityoflondon.gov.uk</LocalAuthorityEmailAddress>
      <Scores/>
      <SchemeType>FHRS</SchemeType>
      <NewRatingPending>False</NewRatingPending>
    ▼<Geocode>
        <Longitude>-0.08903100000000</Longitude>
        <Latitude>51.51291700000000</Latitude>
      </Geocode>
    </EstablishmentDetail>
```

A description of some key fields is given in the below table:

Field	Sample Value	Description
FHRSID	294048	Unique Code given to business by FSA
Business Name	1 Lombard Street, Restaurant	Name of the business
BusinessType	Restaurant/Cafe/Canteen	Type of business
PostCode	EC3V 9AA	Postal Code of the business location
RatingValue	5	Rating value, has been described earlier
Longitude	-0.08903100000000	Geo longitude of the location
Latitude	51.51291700000000	Geo latitude of the location

The data is for all business dealing with food, as seen below:

```
out_df['BusinessType'].value_counts()
```

```
Restaurant/Cafe/Canteen                846
Takeaway/sandwich shop                 345
Pub/bar/nightclub                      230
Other catering premises                168
Retailers - other                      134
Retailers - supermarkets/hypermarkets   35
Hotel/bed & breakfast/guest house       15
Mobile caterer                          12
Hospitals/Childcare/Caring Premises      8
School/college/university                7
Distributors/Transporters                2
Manufacturers/packers                    1
Importers/Exporters                      1
Name: BusinessType, dtype: int64
```

However, we will concentrate on various food outlets serving food like restaurants, pub, bar etc. which is also majority of these businesses as show below:

```
out_df=out_df[out_df.BusinessType.isin(['Restaurant/Cafe/Canteen','Takeaway/sandwich shop','Pub/bar/nightclub',
```

```
out_df.shape
```

```
(1436, 3)
```

The datasets used for this project to get online customer likes, popular categories, geolocation details, price ratings etc. were extracted using Foursquare API Venues Platform. To retrieve the necessary data types from the online platform a URL request was build using parameters from Foursquare repository.

The general elements that compose this URL for the project in case are presented as follows:

https://api.foursquare.com/v2/venues/search?client_id={}&client_secret={}&v={}&query={}&ll={},{}&radius={}&limit={}'.format(

CLIENT_ID,
CLIENT_SECRET,
VERSION,
query,
latitude,
longitude,
radius,
LIMIT)

This detail of the parameters can be found at Foursquare developer's page **https://developer.foursquare.com/docs/api.** A registration is required to obtain the client ID and secret. Query is the search term used to get required venues.

The search result is as follows (sample from Foursquare site):


```
{ "meta": { "code": 200, "requestId": "5ac51d7e6a607143d811cecb" }, "response": {
"venues": [ { "id": "5642aef9498e51025cf4a7a5", "name": "Mr. Purple", "location":
{ "address": "180 Orchard St", "crossStreet": "btwn Houston & Stanton St", "lat":
40.72173744277209, "lng": -73.98800687282996, "labeledLatLngs": [ { "label":
"display", "lat": 40.72173744277209, "lng": -73.98800687282996 } ], "distance": 8,
"postalCode": "10002", "cc": "US", "city": "New York", "state": "NY", "country":
"United States", "formattedAddress": [ "180 Orchard St (btwn Houston & Stanton
St)", "New York, NY 10002", "United States" ] }, "categories": [ { "id":
"4bf58dd8d48988d1d5941735", "name": "Hotel Bar", "pluralName": "Hotel Bars",
"shortName": "Hotel Bar", "icon": { "prefix":
"https://ss3.4sqi.net/img/categories_v2/travel/hotel_bar_", "suffix": ".png" },
"primary": true } ], "venuePage": { "id": "150747252" } } ] } }
```

We can get more details on the venue using the explore API as below:

https://api.foursquare.com/v2/venues/{}?client_id={}&client_secret={}&v={}'.format(venueid, CLIENT_ID, CLIENT_SECRET, VERSION, where venue ID is obtained above.


```
{
  "meta": {
    "code": 200,
    "requestId": "59a45921351e3d43b07028b5"
  },
  "response": {
    "venue": {
      "id": "412d2800f964a520df0c1fe3",
      "name": "Central Park",
      "contact": {
        "phone": "2123106600",
        "formattedPhone": "(212) 310-6600",
        "twitter": "centralparknyc",
        "instagram": "centralparknyc",
        "facebook": "37965424481",
        "facebookUsername": "centralparknyc",
        "facebookName": "Central Park"
      },
      "location": {
        "address": "59th St to 110th St",
        "crossStreet": "5th Ave to Central Park West",
        "lat": 40.78408342593807,
        "lng": -73.96485328674316,
        "postalCode": "10028",
        "cc": "US",
        "city": "New York",
        "state": "NY",
        "country": "United States",
        "formattedAddress": [
          "59th St to 110th St (5th Ave to Central Park West)",
          "New York, NY 10028",
          "United States"
        ]
      },
      "canonicalUrl": "https://foursquare.com/v/central-park/412d2800f964a520df0c1fe3",
      "categories": [
        {
          "id": "4bf58dd8d48988d163941735",
          "name": "Park",
          "pluralName": "Parks",
          "shortName": "Park",
          "icon": {
```


The categories returned by Foursquare are more elaborate than FSA BusinessType and will be used to analyze the restaurants better.

Once we are ready to further explore a particular venue, we can use the below url to get number of likes for that venue:

https://api.foursquare.com/v2/venues/{}/likes?client_id={}&client_secret={}&v={}'.format(venueid, CLIENT_ID, CLIENT_SECRET, VERSION


```
{ "meta": { "code": 200, "requestId": "5854c1f8351e3d5887252266" },
"notifications": [ { "item": { "unreadCount": 13 }, "type": "notificationTray" }
], "response": { "like": false, "likes": { "count": 28, "groups": [ { "count": 28,
"items": [], "type": "others" } ], "summary": "28 Likes" } } }
```


Analysis and outcome

The idea is to merge data from both the FSA and Foursquare sources to build up our own dataset that caters to our requirements.

Once we have the complete final dataset, we can define our categories based on food safety ratings, customer likes, price indicator.

The restaurants which have the highest food safety rating, likes > a specific count based on percentile of the likes, affordable price range indicator will be rated best. Likewise, we can define other categories.

Finally, we will run a KNN algorithm to cluster our final datasets on the London restaurants to cluster the restaurants into the defined catergories.

In order to evaluate our results, will find out the Jaccard index, F1-score and log loss of our predicted outcome and see which iteration of the KNN algorithm gives us the most accurate result.

Once the result is finalized, we will publish the clusters and inform our customers which neighborhood in London has the best rated hygienic, yummy, and affordable food.