

# Battle of the neighborhoods

## London 2020

---

MARCH 12

---

Authored by: Pratish Patni



---

## Contents

<b>Categorization of London restaurants based on food safety, customer rating and price .....</b>	<b>3</b>
<b>Introduction.....</b>	<b>4</b>
<b>Problem statement .....</b>	<b>5</b>
<b>Stakeholders .....</b>	<b>5</b>
<b>Background.....</b>	<b>6</b>
<b>FSA – Food Standards Agency, United Kingdom .....</b>	<b>6</b>
<b>Foursquare API.....</b>	<b>8</b>
<b>Data description .....</b>	<b>9</b>
<b>Methodology .....</b>	<b>14</b>
<b>Results .....</b>	<b>16</b>
<b>Discussion .....</b>	<b>18</b>
<b>Conclusion .....</b>	<b>19</b>

# Categorization of London restaurants based on food safety, customer rating 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 (rating) 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.*

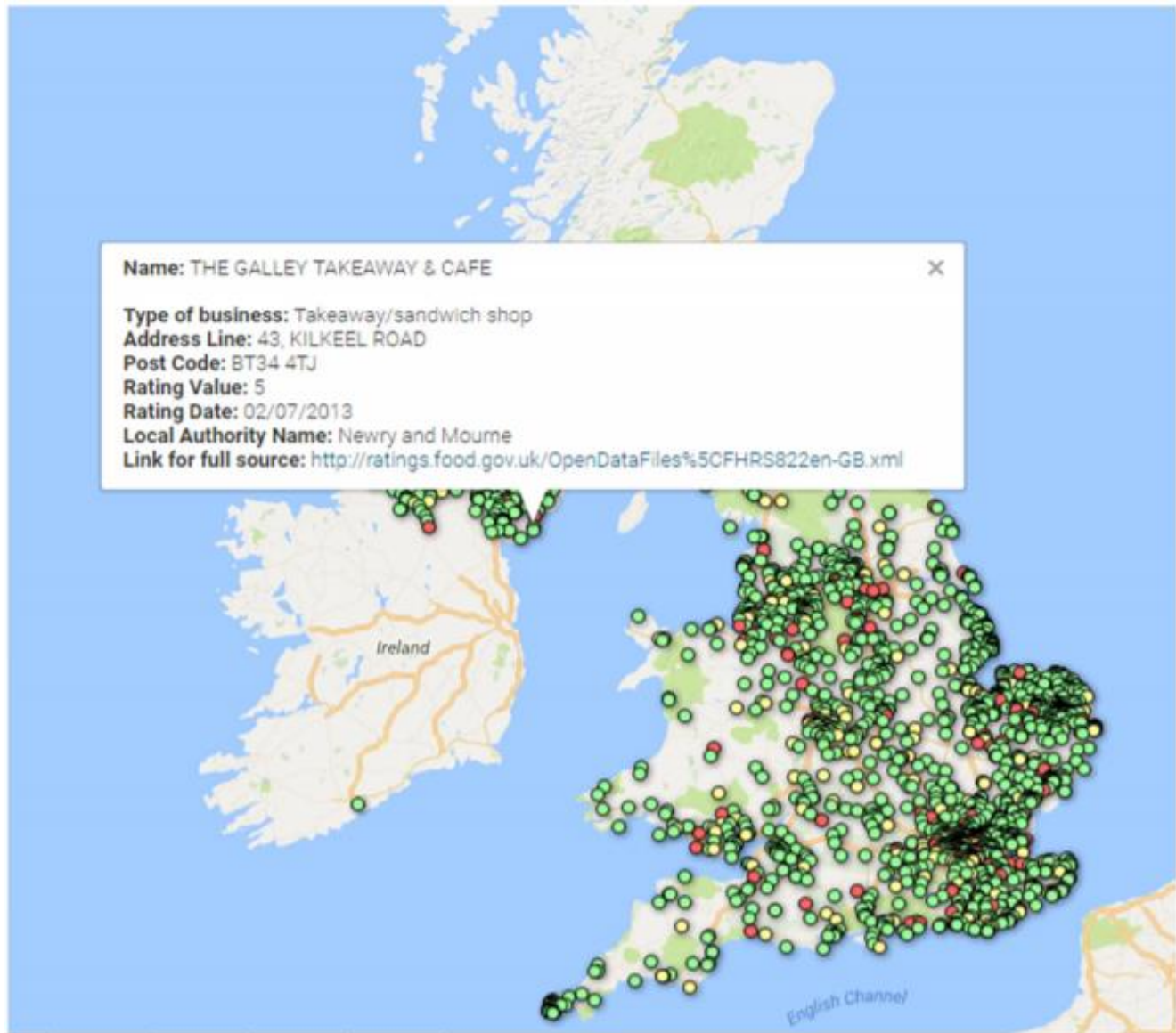
The scheme gives businesses a rating from 5 to 0 which is displayed at their premises and online so you can make more informed choices about where to buy and eat food.



- 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

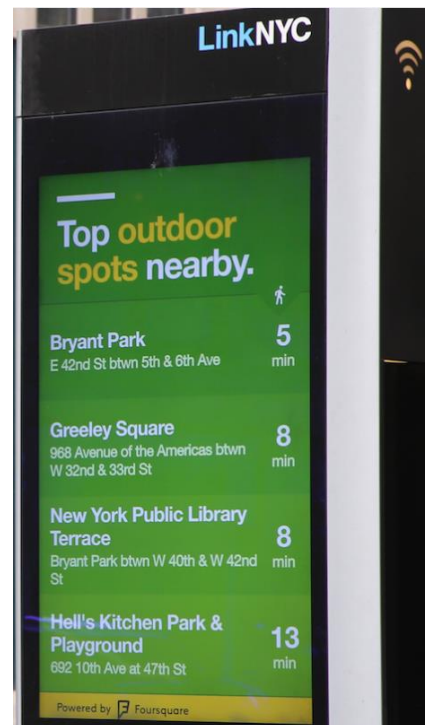
Foursquare is a *'location technology platform dedicated to improving how people move through the real world'*.

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 rating 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> — 1
      <LocalAuthorityBusinessID>PI/000008080</LocalAuthorityBusinessID>
      <BusinessName>1 Lombard Street, Restaurant</BusinessName> — 2
      <BusinessType>Restaurant/Cafe/Canteen</BusinessType> — 3
      <BusinessTypeID>1</BusinessTypeID>
      <AddressLine1>1 Lombard Street</AddressLine1>
      <AddressLine2>London</AddressLine2>
      <PostCode>EC3V 9AA</PostCode> — 4
      <RatingValue>5</RatingValue> — 5
      <RatingKey>fhrrs_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>FHR</SchemeType>
      <NewRatingPending>False</NewRatingPending>
      ▼<Geocode>
        <Longitude>-0.08903100000000</Longitude> — 6
        <Latitude>51.51291700000000</Latitude> — 7
      </Geocode>
    </EstablishmentDetail>
  </EstablishmentCollection>
</FHRSEstablishment>
```

---

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 datasets used for this project to get online customer rating, 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/{id}?client\\_id={}&client\\_secret={}&v={}.format\(venueid, CLIENT\\_ID, CLIENT\\_SECRET, VERSION\)](https://api.foursquare.com/v2/venues/{id}?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.

There are many categories of places returned by FSA data:

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

Restaurant/Cafe/Canteen	841
Takeaway/sandwich shop	344
Pub/bar/nightclub	229
Other catering premises	148
Retailers - other	134
Retailers - supermarkets/hypermarkets	34
Hotel/bed & breakfast/guest house	14
Mobile caterer	10
Hospitals/Childcare/Caring Premises	8
School/college/university	7
Distributors/Transporters	2
Manufacturers/packers	1
Importers/Exporters	1

Name: BusinessType, dtype: int64

We will just focus on places which serve food such as restaurants. Also, the fact that Foursquare has limit on number of API calls to be made, and we must make multiple calls including premium. So, we will limit our initial dataset.

```
out_df_R = out_df[out_df['BusinessType']=='Restaurant/Cafe/Canteen']
```

```
out_df_R.shape
```

```
(212, 7)
```

The above size is not the original because I have been slicing the data and then querying it against Foursquare API.

A subset of response from Foursquare merged with FSA data:

	Name	Business Type	Food Safety Rating	Venue ID	Venue Latitude	Venue Longitude	Venue Category	Total Likes	User Rating	Price
0	EAT.	Restaurant/Cafe/Canteen	5	4b19280af964a52094d823e3	51.519854	-0.097873	Sandwich Place	5.0	5.4	1.0
1	EAT.	Restaurant/Cafe/Canteen	5	4c36eec5ae2da593edf5fdc5	51.514508	-0.096357	Sandwich Place	5.0	5.9	1.0
2	EAT.	Restaurant/Cafe/Canteen	5	4d3ad9576de7721e8a92f249	51.513962	-0.095676	Sandwich Place	7.0	4.9	1.0
3	EAT.	Restaurant/Cafe/Canteen	5	572f63e8cd102de0088dbfd6	51.518014	-0.096905	Sandwich Place	0.0	5.8	1.0
4	Eatwell Restaurant	Restaurant/Cafe/Canteen	5	5d10c6e269ad4f0023b5921e	51.517162	-0.099004	Corporate Cafeteria	0.0	NaN	NaN

After cleaning the data by removing all fields with NaN values, removing duplicates, following is the data ready for analysis:

	Name	Business Type	Food Safety Rating	Venue ID	Venue Latitude	Venue Longitude	Venue Category	Total Likes	User Rating	Price
0	1 Lombard Street, Restaurant	Restaurant/Cafe/Canteen	5	4ac518bcf964a5207da220e3	51.513179	-0.088872	Restaurant	68.0	7.5	3.0
1	14 Hills	Restaurant/Cafe/Canteen	5	5d0f5a7044627d0023f4e721	51.512021	-0.081007	Restaurant	1.0	7.6	2.0
6	1901 Restaurant and Bar	Restaurant/Cafe/Canteen	5	5beac305d1a402002c7b31ea	51.513599	-0.087253	Restaurant	14.0	8.0	2.0
8	1901 Restaurant and Bar	Restaurant/Cafe/Canteen	5	4c1905d2f551ef3bd11d4768	51.518561	-0.093132	Pub	9.0	5.9	1.0
9	1901 Restaurant and Bar	Restaurant/Cafe/Canteen	5	594d318ba4b51b7d05bb26ed	51.525200	-0.082651	Japanese Restaurant	90.0	8.2	2.0
...	...	...	...	...	...	...	...	...	...	...
221	Notes Coffee Barrows	Restaurant/Cafe/Canteen	5	530dcb9a498ec337c000fcea	51.519407	-0.089254	Coffee Shop	7.0	6.7	1.0
222	Nusa Kitchen	Restaurant/Cafe/Canteen	5	593fe3f10d8a0f7b00a3c00f	51.515768	-0.090636	Soup Place	5.0	5.9	1.0
228	Original Bagel Bakery	Restaurant/Cafe/Canteen	4	4bcef39bcc8cd13a480ec5cf	51.522379	-0.097641	Bagel Shop	10.0	8.0	1.0
229	Osteria	Restaurant/Cafe/Canteen	4	56b25967498e51046e48118d	51.519273	-0.093718	Italian Restaurant	7.0	6.4	2.0
231	Panino Giusto	Restaurant/Cafe/Canteen	5	51715d33e4b0e5ceaa967f88	51.513976	-0.086589	Sandwich Place	26.0	6.9	1.0

202 rows × 10 columns

```
london_venues.shape
```

```
(237, 10)
```



## Methodology

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

The basic idea of this project is to rate the restaurants in the city of London based on:

- Food safety rating
- User Rating
- Price

A glimpse of all restaurants displayed on the map of London:



We do one hot encoding of our dataset based on the three fields: 'Food Safety Rating', 'User Rating', 'Price'

	Name	Food Safety Rating	User Rating	Price
0	1 Lombard Street, Restaurant	5	7.5	3.0
1	14 Hills	5	7.6	2.0
6	1901 Restaurant and Bar	5	8.0	2.0
8	1901 Restaurant and Bar	5	5.9	1.0
9	1901 Restaurant and Bar	5	8.2	2.0

Now we define the **cluster = 4**, and then run the **k-mean algorithm** on the data. The labels generated by the k-mean algorithm will be assigned to the dataset.

	Name	Business Type	Food Safety Rating	Venue ID	Venue Latitude	Venue Longitude	Venue Category	Total Likes	User Rating	Price	label
0	1 Lombard Street, Restaurant	Restaurant/Cafe/Canteen	5	4ac518bcf964a5207da220e3	51.513179	-0.088872	Restaurant	68.0	7.5	3.0	1
1	14 Hills	Restaurant/Cafe/Canteen	5	5d0f5a7044627d0023f4e721	51.512021	-0.081007	Restaurant	1.0	7.6	2.0	3
6	1901 Restaurant and Bar	Restaurant/Cafe/Canteen	5	5beac305d1a402002c7b31ea	51.513599	-0.087253	Restaurant	14.0	8.0	2.0	3
8	1901 Restaurant and Bar	Restaurant/Cafe/Canteen	5	4c1905d2f551ef3bd11d4768	51.518561	-0.093132	Pub	9.0	5.9	1.0	0
9	1901 Restaurant and Bar	Restaurant/Cafe/Canteen	5	594d318ba4b51b7d05bb26ed	51.525200	-0.082651	Japanese Restaurant	90.0	8.2	2.0	3

Finally, we rename the labels to the categories identified by us so we can base them as our recommendation

```
london_venues.loc[london_venues['label'] == 3, 'label'] = 'cheap and best'
london_venues.loc[london_venues['label'] == 2, 'label'] = 'very bad'
london_venues.loc[london_venues['label'] == 1, 'label'] = 'pricey and best'
london_venues.loc[london_venues['label'] == 0, 'label'] = 'cheap & ok but not liked'
```

## Results

The algorithm produced 4 clusters of data as shown below. We need to look at the data for each cluster to make sense out of it and find how it has been classified.

### Cluster 0:

	Name	Business Type	Food Safety Rating	Venue ID	Venue Latitude	Venue Longitude	Venue Category	Total Likes	User Rating	Price	label
8	1901 Restaurant and Bar	Restaurant/Cafe/Canteen	5	4c1905d2f551ef3bd11d4768	51.518561	-0.093132	Pub	9.0	5.9	1.0	0
37	Abokado	Takeaway/sandwich shop	5	4b13de37f964a520e79923e3	51.520128	-0.104246	Sushi Restaurant	18.0	6.7	2.0	0
43	Abokado	Takeaway/sandwich shop	5	4acdd443f964a52029cd20e3	51.523959	-0.087432	Sushi Restaurant	30.0	6.0	2.0	0
45	Abokado	Takeaway/sandwich shop	5	530b0630498e233d6936f4e0	51.513516	-0.073368	Sushi Restaurant	12.0	6.1	2.0	0
0	Bad Egg	Restaurant/Cafe/Canteen	4	5490afd9498ec53ea63bbf2c	51.519212	-0.089948	Diner	90.0	6.3	2.0	0

*If we look at the data above it indicates that the restaurants with label=0 are very cheap, with bad user rating, good food safety rating so we can call them 'cheap & ok but not liked'*

### Cluster 1:

	Name	Business Type	Food Safety Rating	Venue ID	Venue Latitude	Venue Longitude	Venue Category	Total Likes	User Rating	Price	label
0	1 Lombard Street, Restaurant	Restaurant/Cafe/Canteen	5	4ac518bcf964a5207da220e3	51.513179	-0.088872	Restaurant	68.0	7.5	3.0	1
13	Baraka Restaurant	Restaurant/Cafe/Canteen	5	50237f538055249578a0f772	51.519110	-0.086729	Seafood Restaurant	61.0	7.3	4.0	1
18	Be At One	Restaurant/Cafe/Canteen	5	4b6dfa98f964a520c5a02ce3	51.518909	-0.078269	Cocktail Bar	73.0	6.6	3.0	1
0	Cabotte	Restaurant/Cafe/Canteen	4	57dfb4a1498e923fc6548fc8	51.514984	-0.091293	French Restaurant	8.0	6.8	3.0	1
5	Camino	Restaurant/Cafe/Canteen	5	4fd110d4e4b0b63780abf524	51.510791	-0.081540	Tapas Restaurant	39.0	7.7	3.0	1

*If we look at the data above it indicates that the restaurants with label=1 are pricey, with good user rating, good food safety rating so we can call them 'pricey and best'*

### Cluster 2:

	Name	Business Type	Food Safety Rating	Venue ID	Venue Latitude	Venue Longitude	Venue Category	Total Likes	User Rating	Price	label
15	Barcelona Tapas Bar & Restaurant	Restaurant/Cafe/Canteen	3	4df9e831aeb785aedbeffc6d	51.511777	-0.083728	Spanish Restaurant	6.0	6.1	2.0	2
16	Barcelona Tapas Bar & Restaurant	Restaurant/Cafe/Canteen	3	4bb4f53329269c748be0ca92	51.512828	-0.083401	Tapas Restaurant	6.0	5.9	3.0	2
19	Bea's of Bloomsbury	Restaurant/Cafe/Canteen	3	4ccc16e897d0224bed705db8	51.513382	-0.095616	Tea Room	94.0	7.1	2.0	2
21	Bengal Tiger	Restaurant/Cafe/Canteen	3	4b9ff194f964a520764c37e3	51.513349	-0.102060	Indian Restaurant	18.0	5.6	2.0	2
27	Beppes Cafe	Restaurant/Cafe/Canteen	3	4c404717d691c9b67be38a0a	51.517750	-0.101410	Italian Restaurant	19.0	7.3	2.0	2



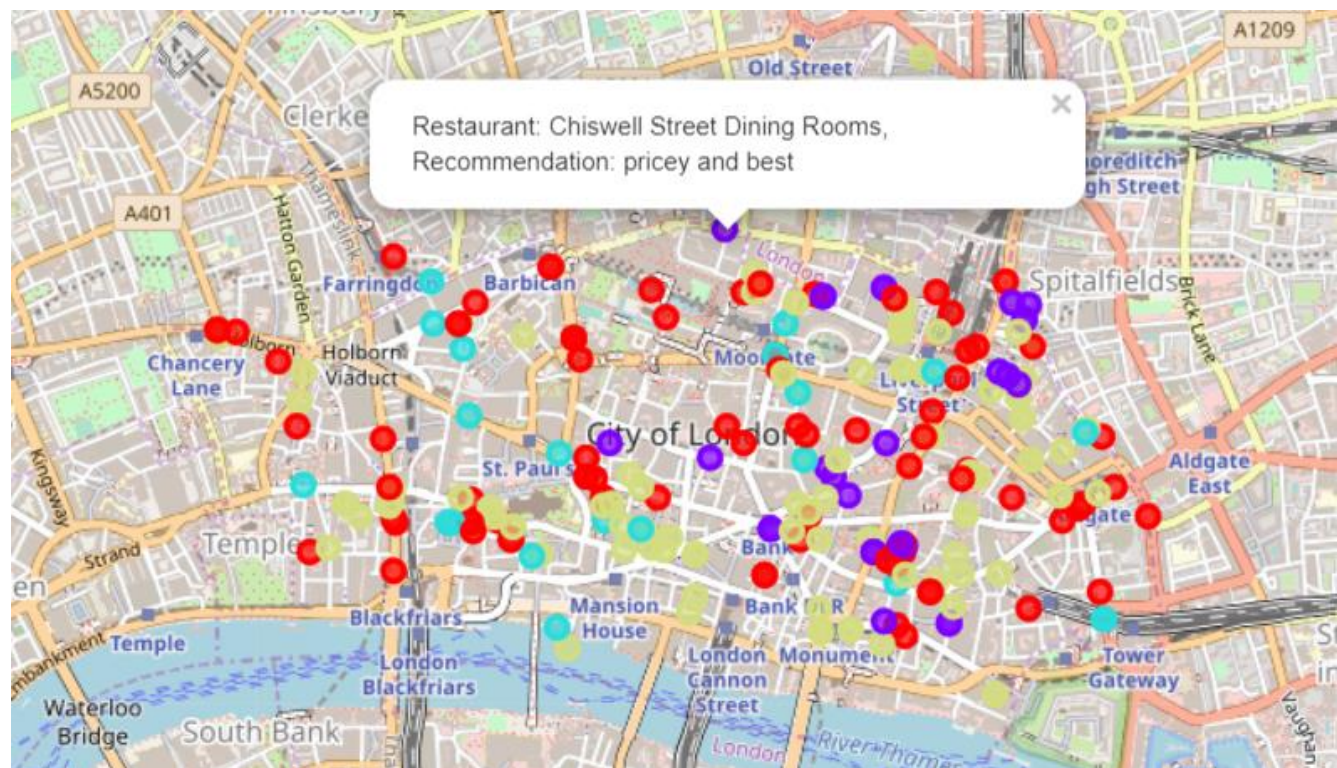
If we look at the data above it indicates that the restaurants with label=2 are pricey, with average user rating, bad food safety rating so we can call them '**very bad**'

### Cluster 3:

	Name	Business Type	Food Safety Rating	Venue ID	Venue Latitude	Venue Longitude	Venue Category	Total Likes	User Rating	Price	label
1	14 Hills	Restaurant/Cafe/Canteen	5	5d0f5a7044627d0023f4e721	51.512021	-0.081007	Restaurant	1.0	7.6	2.0	3
6	1901 Restaurant and Bar	Restaurant/Cafe/Canteen	5	5beac305d1a402002c7b31ea	51.513599	-0.087253	Restaurant	14.0	8.0	2.0	3
9	1901 Restaurant and Bar	Restaurant/Cafe/Canteen	5	594d318ba4b51b7d05bb26ed	51.525200	-0.082651	Japanese Restaurant	90.0	8.2	2.0	3
6	Bangalore Express	Restaurant/Cafe/Canteen	4	4b0da488f964a520f54c23e3	51.512928	-0.084752	Indian Restaurant	21.0	7.3	2.0	3
8	Banh Mi Bay	Restaurant/Cafe/Canteen	4	54cb7bae498e99dacf3c3a28	51.512621	-0.094847	Vietnamese Restaurant	40.0	8.1	2.0	3

If we look at the data above it indicates that the restaurants with label=3 are cheap, with very good user rating, good food safety rating so we can call them '**cheap and best**'

When projected on the map, we can see the cluster with our recommendation:



---

## Discussion

The various categories of restaurants can be further explained as below. It also has some actions and learning for the various stakeholders:

- **Cheap and best** - restaurants which are a value for money and are healthy too. They are also popular with people visiting London
- **Pricey and best** - restaurants which are popular, healthy to eat but may be heavier on your pocket
- **Cheap & ok but not liked** - restaurants which are with good health rating and cheap too but for some reasons are not very liked by people who have visited them. **This is a curious case which must be inspected by both the *restaurant owners* and the *food safety agency***
- **Very bad** - certainly a case where nothing is good. These have very bad food rating, average user rating and are costly as well. **The *restaurant owners* must certainly look at these and try to improve**



---

## Conclusion

We cannot be very judgmental about which neighborhood is better in terms of quality of restaurant because all categories of restaurants are pretty spread out but my **recommendations can certainly help** when you are in a **neighborhood** and want to visit a particular **restaurant**.

This report will prove pretty useful for **restaurant owners** who **can improve on their food quality, taste, service and affordability of the restaurants** so that they can get **better recommendations**

The **health officials** can look at the **few curious cases** where a good food safety rating has been assigned to restaurants which are not very much liked by their customers.