Eat Safe, Love

Notebook Set Up

In [10]:
```
# Import dependencies
from pymongo import MongoClient
import pandas as pd
from pprint import pprint
```

In [11]:
```
# Create an instance of MongoClient
mongo = MongoClient(port=27017)
```

In [12]:
```
# assign the uk_food database to a variable name
db = mongo['uk_food']
```

In [13]:
```
# review the collections in our database
print(db.list_collection_names())
```

```
['establishments']
```

In [14]:
```
# assign the collection to a variable
establishments = db['establishments']
```

Part 3: Exploratory Analysis

Unless otherwise stated, for each question:

- Use `count_documents` to display the number of documents contained in the result.
- Display the first document in the results using `pprint`.
- Convert the result to a Pandas DataFrame, print the number of rows in the DataFrame, and display the first 10 rows.

1. Which establishments have a hygiene score equal to 20?

In [15]:
```
# Find the establishments with a hygiene score of 20
query = {'scores.Hygiene': 20}
print("Number of documents in result: ", establishments.count_documents(
print("First result:")
results = establishments.find(query)
pprint(results[0])
```

```
Number of documents in result:  41
First result:
{'AddressLine1': '5-6 Southfields Road'
```

In [16]:

```
# Convert the result to a Pandas DataFrame
results = establishments.find(query)
result_df = pd.DataFrame(results)
# Display the number of rows in the DataFrame
print("Rows in DataFrame: ", len(result_df))
# Display the first 10 rows of the DataFrame
result_df.head(10)
```

```
Rows in DataFrame:  41
```

Out[16]:

	_id	FHRSID	ChangesByServerID	LocalAuthorityBusinessID	Busines
0	65c7de6df61833133ba96879	110681	0	4029	The Res
1	65c7de6df61833133ba96bfa	612039	0	1970/FOOD	Bren
2	65c7de6ef61833133ba96f04	730933	0	1698/FOOD	Melros
3	65c7de6ef61833133ba970f2	172735	0	PI/000023858	Seafor
4	65c7de6ef61833133ba97101	172953	0	PI/000024532	Golden
5	65c7de6ff61833133ba97aa0	512854	0	12/00816/BUTH	A B

2. Which establishments in London have a `RatingValue` greater than or equal to 4?

In [20]:

```
# Find the establishments with London as the Local Authority and has a RatingValue greater thar
query = {'LocalAuthorityName': {'$regex':'London'}, 'RatingValue': {'$gte': 4}}
print("Number of documents in result: ", establishments.count_documents(query))
print("First result:")
results = establishments.find(query)
pprint(results[0])
```

```
Number of documents in result:  33
First result:
{'AddressLine1': 'Oak Apple Farm Building 103 Sheernes Docks',
 'AddressLine2': 'Sheppy Kent',
 'AddressLine3': '',
 'AddressLine4': '',
 'BusinessName': "Charlie's",
 'BusinessType': 'Other catering premises',
 'BusinessTypeID': 7841,
 'ChangesByServerID': 0,
 'Distance': 4627.439467780196,
 'FHRSID': 621707,
 'LocalAuthorityBusinessID': 'PI/000025307',
 'LocalAuthorityCode': '508',
 'LocalAuthorityEmailAddress': 'publicprotection@cityoflondon.gov.uk',
 'LocalAuthorityName': 'City of London Corporation',
 'LocalAuthorityWebSite': 'http://www.cityoflondon.gov.uk/Corporation/homepage.htm',
 'NewRatingPending': False,
 'Phone': '',
 'PostCode': 'ME12',
 'RatingDate': '2021-10-18T00:00:00',
 'RatingKey': 'fhrs_4_en-gb',
```

In [21]:

```
# Convert the result to a Pandas DataFrame
results = establishments.find(query)
London_df = pd.DataFrame(results)
# Display the number of rows in the DataFrame
print('Number of Documents in dataframe: ', len(London_df))
# Display the first 10 rows of the DataFrame
London_df.head(10)
```

```
Number of Documents in dataframe:  33
```

Out[21]:

	_id	FHRSID	ChangesByServerID	LocalAuthorityBusinessID	Busines
0	65c7de6ff61833133ba98294	621707	0	PI/000025307	C
1	65c7de6ff61833133ba985ba	1130836	0	PI/000034075	Mv City (E
2	65c7de70f61833133ba99107	293783	0	PI/000002614	Benflee Yac
3	65c7de71f61833133ba99f07	1315095	0	PI/000036464	C Catering Lock a

3. What are the top 5 establishments with a `RatingValue` rating value of 5, sorted by lowest hygiene score, nearest to the new restaurant added, "Penang Flavours"?

In [22]:

```
# Search within 0.01 degree on either side of the latitude and longitude
# Rating value must equal 5
# Sort by hygiene score

degree_search = 0.01
latitude = 51.49014200
longitude = 0.08384000

query = {'geocode.latitude': {'$gte':latitude-degree_search, '$lte':lati
         'geocode.longitude': {'$gte': longitude-degree_search, '$lte':
         'RatingValue': 5}
#sort by the hygiene scores from lowest to greatest
sort =  [('score.Hygiene', 1)]
#set a limit of 5 establishments
limit = 5


# Print the results
pprint(list(establishments.find(query).sort(sort).limit(limit)))
```

In [24]:

```
# Convert result to Pandas DataFrame
results = establishments.find(query)
near_penang_df = pd.DataFrame(results)
# Display the number of rows in the DataFrame
print('Number of rows in dataframe: ', len(near_penang_df))
# Display the first 10 rows of the DataFrame
near_penang_df
```

```
Number of rows in dataframe:  87
```

Out[24]:

scores	SchemeType	geocode	RightToReply	Distance	NewRatingPending	
giene': 5, ctural': 5, nceInM...	FHRS	{'longitude': 0.0927429, 'latitude': 51.4870351}		4646.930146	False	{'dat 'ext '00(
giene': 0, ctural': 5, nceInM...	FHRS	{'longitude': 0.0924199968576431, 'latitude': ...		4646.946071	False	{'dat 'ext '00(
giene': 0, ctural': 0, nceInM...	FHRS	{'longitude': 0.0925370007753372, 'latitude': ...		4646.955931	False	{'dat 'ext '00(
giene': 0, ctural': 0, nceInM...	FHRS	{'longitude': 0.09208, 'latitude': 51.4873437}		4646.965635	False	{'dat 'ext '00(
giene': 0, ctural': 0, nceInM...	FHRS	{'longitude': 0.0859939977526665, 'latitude': ...		4646.974010	False	{'dat 'ext

4. How many establishments in each Local Authority area have a hygiene score of 0?

In [25]:

```
# Create a pipeline that:
# 1. Matches establishments with a hygiene score of 0
# 2. Groups the matches by LocalAuthorityName
# 3. Sorts the matches from highest to lowest
# Mongosh version:
# db.establishments.aggregate( [ { $match: {'scores.Hygiene': 0}}, { $gr
pipeline = [
    {'$match': {'scores.Hygiene': 0}},
    {'$group': {'_id': "$LocalAuthorityName", 'count': { '$sum': 1 }}},
    {'$sort': { 'count': -1 }}
]

results = list(establishments.aggregate(pipeline))

# Print the number of documents in the result
print("Number of documents in result: ", len(results))

# Print the first 10 results
pprint(results[0:10])
```

```
Number of documents in result:  55
[{'_id': 'Thanet', 'count': 1130},
 {'_id': 'Greenwich', 'count': 882},
 {'_id': 'Maidstone', 'count': 713},
 {'_id': 'Newham', 'count': 711},
 {'_id': 'Swale', 'count': 686},
 {'_id': 'Chelmsford', 'count': 680},
```

In [27]:

```
# Convert the result to a Pandas DataFrame

# Display the number of rows in the DataFrame

# Display the first 10 rows of the DataFrame

result_df = pd.DataFrame(results)

print("Rows in DataFrame: ", len(result_df))
result_df.head(20)

```

```
Rows in DataFrame:  55
```

Out[27]:

	_id	count
0	Thanet	1130
1	Greenwich	882
2	Maidstone	713
3	Newham	711
4	Swale	686
5	Chelmsford	680
6	Medway	672
7	Bexley	607
8	Southend-On-Sea	586