# Sentiment and Data Analysis on Restaurant Reviews Final Report

## Group 2

### Ally Ryan
Computer Science and Engineering
UMKC
Kansas City Missouri
aerc4d@umsystem.edu

### Mustavi Islam
Computer Science and Engineering
UMKC
Kansas City Missouri
mizmc@umsystem.edu

### Mufrad Islam
Computer Science and Engineering
UMKC
Kansas City Missouri
migh8@umsystem.edu

### Nabila Hashim
Computer Science and Engineering
UMKC
Kansas City Missouri
nntdc@umsystem.edu

### Jack Wojcicki
Computer Science and Engineering
UMKC
Kansas City Missouri
jswfff@umsystem.edu

### Sai Poorna Chowdary Dama
Computer Science and Engineering
UMKC
Kansas City Missouri
sd8h6@umsystem.e

## ABSTRACT

The purpose of this report is to showcase techniques that can be used to identify important insights within restaurant and/or food reviews. The restaurant and/or food industry is filled with critics and is supported across the world. Any person who enjoys going out to eat is interested in spending their money on a quality dining experience, either with good food, good service, or a good atmosphere. Likewise, any person visiting a restaurant may be interested in the specific dishes to order. To help the consumer have a quality dining experience, sentiment analysis and data analysis techniques were used. The sentiment analysis techniques help to classify the consumer reviews as positive, negative, or neutral. The model we were able to develop classified positive, negative, and neutral reviews with an 88% level of accuracy. The data analysis techniques help to identify the best food types and dishes within a specific city. The data analysis techniques are also helpful in generating maps to locate a city's best restaurants and dishes. Not only are the maps helpful to the consumer looking for a great dining experience, but also to the restaurant manager looking for the best spot to open their new restaurant. Being able to locate where the successful restaurants are, and also where there is a need for a certain restaurant type in a city has potential to be very helpful for a restaurant's success.

## KEYWORDS

Sentiment, Data, Analysis, Restaurant, Reviews

## 1 Introduction: Dataset

The data set being used to perform sentiment and data analysis was found on Kaggle and contains TripAdvisor restaurant reviews for the city of London. As the class requirement for the file size had to be greater than 500 MB, the London restaurant review dataset in use fulfills that requirement with a 942.33 MB size. Inside the London restaurant review dataset, there are a million restaurant reviews. The notable columns include review_full, rating_review, and restaurant_name. The review_full column contains the TripAdvisor restaurant review column made by the customer on a respective restaurant. The review_full column will contain insight regarding the customer's experience, the quality of the food, the quality of the service, the quality of the ambience, and (but not limited to) perhaps the exact food they ordered. The restaurant name provides the name of the restaurant with all spaces removed and replaced by an underscore. The rating_review column provides a numerical ranking from 1-5 pertaining to what the customer ranked their experience, with 1 being a terrible experience and 5 being a fantastic experience.

## 2 Introduction: Sentiment Analysis

The first goal of the project is to perform sentiment analysis on the London restaurant reviews. Through sentiment analysis, the goal is to develop a model that determines whether London restaurant reviews are positive or negative. The interesting component of the London dataset is that it already contains a column that contains 'Positive' or 'Negative' dispositions. One may be able to conclude from such data that sentiment analysis has already been performed, and thus that we should simply develop a model that tests the datasets accuracy. However, upon further analysis, we discovered that any ranking assigned a rating_review from 1-3 were automatically given a 'negative' disposition, and any ranking assigned a rating_review from 4-5 were automatically given a 'positive' disposition. As such, we seek to continue our sentiment analysis and perhaps perform some data analysis to determine if the customer's review matches up with their inputted ranking.

## 3 Initial Data Cleaning Steps

To perform sentiment analysis and in turn data analysis and visualization, the first step was to gain a deep understanding of our data and clean the data. The first thing we did to clean the data was to remove inaccurately named restaurants. To gain a clear understanding of our dataset, we first uploaded our dataset into SQLite via DBeaver, as this system allowed us to scroll through the output of our completed queries. However, we performed all our cleaning and preprocessing in PySpark via Google Colab. Upon completion of a 'select distinct restaurant names' query, we recognized that there were several restaurant names with reviews in the columns (Figure 1). In turn, we deleted such rows.



**Figure 1: Inaccurate Restaurant Names**

The second step of our cleaning process was to remove null restaurant names (Figure 2).



**Figure 2: Remove Null Restaurant Names**

The third step of our cleaning process was to remove inaccurate city names (Figure 3). As our dataset is supposed to contain restaurant reviews only from the city of London, we wanted to ensure that all reviews did indeed come from London. As a result of a 'select distinct city' names query, we were able to identify one restaurant that had a review in the city column. As such, the review in the city column was deleted.



**Figure 3: Remove Inaccurate City Names**

The fourth step of our cleaning process was to check for null values in the review_full column, as this column is deeply important to our sentiment analysis model (Figure 4). We did not find any, and thus could proceed without taking any further action in this step.



**Figure 4: Check for Null Reviews**

The fifth step of our cleaning process was to check that the rating_review column contained only values 1-5 (Figure 5). Upon completion of a 'select distinct rating_review' query, we found that the contents of this column only contained values 1-5.



**Figure 5: Check Rating Column**

As such we concluded that there was no content of concern in this column, and that we could proceed without taking any further action.

## 4 Preprocessing Steps

After the dataset as a whole was cleaned, it was necessary for us to further preprocess the data to get the text ready for use in a sentiment analysis model. The first step of this sentiment analysis preprocessing activity was to drop irrelevant columns in the dataset (Figure 6). As such, we only kept the restaurant_name, rating_review, sample (which contains the positive and negative dispositions), and review_full columns.



**Figure 6: Drop Irrelevant Columns**

The columns mentioned above are the only columns needed to perform sentiment analysis, and thus the rest were removed for the duration of the sentiment analysis project portion. It should be noted that some columns such as date and website_link may be returned for the data analysis portion of the project. The second and third steps of our sentiment analysis preprocessing included making the review_full column all lowercase letters (Figure 7) and removing all punctuation (Figure 8).

```
+------------------+-------------+------+-----------------------------------------------------+
|restaurant_name   |rating_review|sample|review_full                                          |
+------------------+-------------+------+-----------------------------------------------------+
|Cocotte_Notting_Hill|5.0        |Positive|i was away for couple of months and i am so happy to be back
|Cocotte_Notting_Hill|5.0        |Positive|i recently ordered over the phone from cocotte to collect for
|Cocotte_Notting_Hill|5.0        |Positive|came here because we missed our table at farmacy across the r
|Cocotte_Notting_Hill|5.0        |Positive|my first time in cocotte and was amazed by how friendly this
|Cocotte_Notting_Hill|5.0        |Positive|a healthy-homemade dishes using farm-grown ingredients really
+------------------+-------------+------+-----------------------------------------------------+
only showing top 5 rows
```

**Figure 7: Change Review Column to Lowercase**

```
+------------------+-------------+------+-----------------------------------------------------+
|restaurant_name   |rating_review|sample|review_full                                          |
+------------------+-------------+------+-----------------------------------------------------+
|Cocotte_Notting_Hill|5.0        |Positive|i was away for couple of months and i am so happy to be back for my
|Cocotte_Notting_Hill|5.0        |Positive|i recently ordered over the phone from cocotte to collect for a gat
|Cocotte_Notting_Hill|5.0        |Positive|came here because we missed our table at farmacy across the road th
|Cocotte_Notting_Hill|5.0        |Positive|my first time in cocotte and was amazed by how friendly this member
|Cocotte_Notting_Hill|5.0        |Positive|a healthyhomemade dishes using farmgrown ingredients really lovely
+------------------+-------------+------+-----------------------------------------------------+
only showing top 5 rows
```

**Figure 8: Remove Punctuation from Review Column**

The fourth step included removing all stop words, which included words like 'in,' 'a,' and 'as' (Figure 9). This step also placed all words in the review_full column into an array with a comma separating each word.

```
+------------------+-------------+------+-----------------------------------------------------+
|restaurant_name   |rating_review|sample|review_full                                          |
+------------------+-------------+------+-----------------------------------------------------+
|Cocotte_Notting_Hill|5.0        |Positive|[away, couple, months, happy, back, regular, cocotte, meal, collecting
|Cocotte_Notting_Hill|5.0        |Positive|[recently, ordered, phone, cocotte, collect, gathering, home, calling,
|Cocotte_Notting_Hill|5.0        |Positive|[came, missed, table, farmacy, across, road, waitress, accommodated, u
|Cocotte_Notting_Hill|5.0        |Positive|[first, time, cocotte, amazed, friendly, member, front, house, staff,
|Cocotte_Notting_Hill|5.0        |Positive|[healthyhomemade, dishes, using, farmgrown, ingredients, really, lovel
+------------------+-------------+------+-----------------------------------------------------+
only showing top 5 rows
```

**Figure 9: Remove Stop Words from Review Column**

The fifth step of our sentiment analysis preprocessing including lemmatization of the review_full column. This step returns the lemma of each word, meaning that the word 'dishes' would be turned into the base form of the word 'dish' in this step.

```
+------------------+-------------+------+-----------------------------------------------------+
|restaurant_name   |rating_review|sample|review_full                                          |
+------------------+-------------+------+-----------------------------------------------------+
|Cocotte_Notting_Hill|5.0        |Positive|away, couple, month, happy, back, regular, cocotte, meal, collec
|Cocotte_Notting_Hill|5.0        |Positive|recently, ordered, phone, cocotte, collect, gathering, home, cal
|Cocotte_Notting_Hill|5.0        |Positive|came, missed, table, farmacy, across, road, waitress, accommodat
|Cocotte_Notting_Hill|5.0        |Positive|first, time, cocotte, amazed, friendly, member, front, house, st
|Cocotte_Notting_Hill|5.0        |Positive|healthyhomemade, dish, using, farmgrown, ingredient, really, lov
+------------------+-------------+------+-----------------------------------------------------+
only showing top 5 rows
```

**Figure 10: Review Column Lemmatization**

Once we started the data analysis side of the project, we noticed that a low number of restaurant reviews caused the best ranked restaurants to be skewed. For example, an average review of 5.0 likely means that there were just 1-3 people that left a review for that restaurant. (This concept is further talked about in section 5 of this research paper.) As such, for the sentiment analysis side of this project, we decided to delete all restaurants that had less than 5 reviews. Not only did this help make more interesting and accurate results for our project, but it also got rid of the need to do many of the cleaning steps ahead. The restaurant names were no longer placed in the wrong category, as those reviews were already a unique value. Five (5) reviews were chosen for the sentiment analysis section because 5 represented the first number that resulted in multiple unique reviews. Figure 11 below represents the identification of unique restaurant reviews, and figure 11 below represents the restaurants used to conduct sentiment analysis.

```
    breaded chicken strips over cooked fried in old oil..." ,  1
    Merlot_Garden_Bar_Restaurant ,  5
```

**Figure 11: Unique Restaurant Reviews**

```
+------------------+------------------+
|   restaurant_name|count(review_full)|
+------------------+------------------+
|Merlot_Garden_Bar...|               5|
|       Chotto_Matte|               5|
|      Buns_and_Buns|               7|
|Sweetsmile_Bakery...|               8|
|Roxie_Steak_Earls...|               8|
+------------------+------------------+
only showing top 5 rows
```

**Figure 12: Restaurants Used for Sentiment Analysis**

Once the uniqueness of the restaurant reviews were satisfactory, the next step was to split the dataset into a training and test set. We used 80% of the data for the training set, which was used to create our model. We used the remaining 20% of the data for the testing set, which was used to test the model's accuracy. Figure 13 below represents the results from the training data. Looking at the results, we can see that almost all of the predictions are accurate. There are a few that are not, but if they are incorrect they are just 1 point off.

```
+------------------+------------------+-------------+----------+
|   restaurant_name|       review_full|rating_review|prediction|
+------------------+------------------+-------------+----------+
|280_Degrees_Afric...|bar, man, one, co...|         1.0|       1.0|
|280_Degrees_Afric...|called, , degree,...|         1.0|       1.0|
|280_Degrees_Afric...|get, empty, resta...|         1.0|       1.0|
|280_Degrees_Afric...|god, bad, bad, ba...|         1.0|       1.0|
|280_Degrees_Afric...|ordered, food, on...|         1.0|       1.0|
|280_Degrees_Afric...|receiving, groupo...|         1.0|       2.0|
|280_Degrees_Afric...|visited, restaura...|         1.0|       2.0|
|280_Degrees_Afric...|watson, puff, puf...|         1.0|       1.0|
|280_Degrees_Afric...|arrived, hr, took...|         2.0|       2.0|
|280_Degrees_Afric...|booked, table, , ...|         2.0|       2.0|
|280_Degrees_Afric...|food, alright, ni...|         2.0|       3.0|
|280_Degrees_Afric...|restaurant, stand...|         2.0|       2.0|
|280_Degrees_Afric...|went, try, someth...|         2.0|       2.0|
|280_Degrees_Afric...|went, voucher, st...|         2.0|       2.0|
|280_Degrees_Afric...|bought, voucher, ...|         3.0|       3.0|
|280_Degrees_Afric...|came, deal, knowi...|         3.0|       3.0|
|280_Degrees_Afric...|day, ago, pm, aft...|         3.0|       3.0|
|280_Degrees_Afric...|dinner, recently,...|         3.0|       3.0|
|280_Degrees_Afric...|eatery, standard,...|         3.0|       3.0|
|280_Degrees_Afric...|groupon, worth, t...|         3.0|       3.0|
+------------------+------------------+-------------+----------+
```

**Figure 13: Training Data Results**

Figure 14 below represents the results from the testing data. We can see that the results are again fairly accurate, but not perfect. There is only one review where the prediction was off by more than one point.

```
+-------------------+-------------------+-------------+----------+
|    restaurant_name|        review_full|rating_review|prediction|
+-------------------+-------------------+-------------+----------+
| Cahoots_Underground|first, ever, , st...|          5.0|       5.0|
|       The_Monsoon|visit, monsoon, r...|          5.0|       5.0|
| Gaucho_Tower_Bridge|one, best, locati...|          4.0|       5.0|
|Park_Terrace_Rest...|park, terrace, ro...|          4.0|       4.0|
|Thali_Pickles_Ind...|tasty, food, serv...|          5.0|       5.0|
|L_Escargot_Restau...|last, time, ate, ...|          3.0|       3.0|
|Pollen_Street_Social|disappointing, oc...|          3.0|       3.0|
|Sketch_Lecture_Ro...|visited, restaura...|          5.0|       5.0|
|   The_Hereford_Arms|researched, place...|          5.0|       5.0|
|            Cosmoba|reservation, cosm...|          2.0|       4.0|
| Franco_Manca_Balham|luck, heart, balh...|          5.0|       5.0|
|           GONG_Bar|, visited, drink,...|          5.0|       5.0|
|Bentley_s_Oyster_...|service, good, so...|          3.0|       3.0|
|              Hunan|need, book, well,...|          5.0|       5.0|
|        Mother_Mash|meal, potion, big...|          4.0|       4.0|
|          Balthazar|great, restaurant...|          5.0|       5.0|
|         Heist_Bank|food, nice, every...|          4.0|       4.0|
|O_Neill_s_Kings_C...|classic, oneills,...|          3.0|       4.0|
|SUSHISAMBA_Heron_...|husband, went, su...|          5.0|       5.0|
+-------------------+-------------------+-------------+----------+
```

**Figure 14: Testing Data Results**

Figure 15 below represents the table used to compute the model's accuracy. To determine the model's accuracy, we passed a filter across the test data to identify the points where the actual and predicted reviews match. Once we identified the matching points, we divided the result by the total number of rows in our test data. This calculation yielded an accuracy of 68%, which represented the accuracy for when the numerical values matched. However, since the goal of this project was to identify positive, negative, or neutral reviews, we also looked at how frequently the qualitative reviews matched. This accuracy calculation yielded an improved 88.5%.

```
+-------------------+-------------------+--------+---------+
|    restaurant_name|        review_full|off_sent|pred_sent|
+-------------------+-------------------+--------+---------+
|     The_Park_Room|whats, like, deli...|Positive| Positive|
|            Dishoom|found, restaurant...| Neutral| Positive|
|Baltic_Restaurant...|boyfriend, took, ...|Positive| Positive|
|       Hankies_Cafe|first, time, hank...|Positive| Positive|
|        Sale_e_Pepe|every, time, come...|Positive| Positive|
|        ASK_Italian|problem, getting,...|Positive| Positive|
| Cote_Brasserie_Soho|visited, meal, th...|Positive| Positive|
|      Electric_Diner|love, nothing, gr...|Negative| Negative|
|  Hache_Shoreditch|food, staff, amaz...|Positive| Positive|
|Spaghetti_House_W...|pleasant, atmosph...|Positive| Positive|
|         Veeraswamy|wanted, go, somew...|Positive| Positive|
|           Vietfood|lunch, venue, rec...| Neutral|  Neutral|
|       Franco_Manca|want, enjoy, best...|Positive| Positive|
|    Honey_and_Smoke|friend, introduce...|Positive| Positive|
|     The_Wellington|find, well, run, ...|Positive| Positive|
|Cote_Brasserie_Sl...|eaten, hear, plac...|Positive| Positive|
+-------------------+-------------------+--------+---------+
```

**Figure 15: Model Used to Calculate Accuracy**

The above table completed the sentiment analysis portion of our project. Given our dataset had over 1,000,000 restaurant reviews, we are quite pleased we were able to achieve an 88.5% accurate model in determine positive, negative, and neutral reviews.

## 5 Data Analysis

During the preliminary steps of our data analysis and visualization, we completed a few queries that identified the restaurants with the best steak and lobster (Figure 16), tacos and enchiladas (Figure 17), sushi and tempura (Figure 18), and Korean and bibimbap (Figure 19). Upon completion of these queries, we recognized that some average rating columns contained the value 5.0. This means that there were not very many restaurant reviews that contained our

desired foods for this restaurant. To solve this problem, we required that a restaurant have greater than 50 restaurant reviews with a certain food type to be eligible for our ranking system.

```
+------------------+-------------------+
|avg(rating_review)|    restaurant_name|
+------------------+-------------------+
|               5.0|   The_Pearson_Room|
|               5.0|Laurent_at_Cafe_R...|
|               5.0|Granary_Square_Br...|
|               5.0|Bloomsbury_Street...|
|               5.0|      Ormer_Mayfair|
|               5.0|  Augustine_Kitchen|
|               5.0|  Galvin_La_Chapelle|
|               5.0|    Brasserie_Zedel|
|               5.0|       Ciullosteria|
|               5.0|         Bella_Cosa|
|               5.0|          Pentolina|
| 4.666666666666667|            Smith_s|
| 4.666666666666667|Bel_Canto_Restaurant|
|               4.5|  Locanda_Locatelli|
|               4.5|Mac_And_Wild_Fitz...|
|               4.5|InterContinental_...|
|               4.5|Holborn_Dining_Room|
| 4.444444444444445|Butlers_Wharf_Cho...|
|              4.25|Boisdale_of_Bisho...|
|              4.25| Plum_and_Spilt_Milk|
+------------------+-------------------+
only showing top 20 rows
```

**Figure 16: Restaurants with the Best Steak and Lobster**

```
+------------------+-------------------+
|avg(rating_review)|    restaurant_name|
+------------------+-------------------+
|               5.0|         Aqua_Nueva|
|               5.0|              Anglo|
|               5.0|Busaba_Westfield_...|
|               5.0|     Berners_Tavern|
|               5.0|      AW_Restaurant|
|               5.0|The_Blacksmith_s_...|
|               5.0|     The_Good_Earth|
|               4.9|           Paladar|
|               4.8|  The_Fire_Station|
|               4.8|Donnelly_s_Restau...|
| 4.777777777777778|            Farmacy|
|4.6923076923076925|  May_Fair_Kitchen|
|4.676829268292683|Leicester_Square_...|
| 4.666666666666667|      Yuu_Kitchen|
| 4.615384615384615|   COYA_Angel_Court|
|               4.6|Dirty_Bones_Shore...|
|               4.5|              Momo|
|4.473684210526316|SUSHISAMBA_Heron_...|
| 4.353535353535354|       Jinjuu_Soho|
| 4.317073170731708|              COYA|
+------------------+-------------------+
only showing top 20 rows
```

**Figure 17: Restaurants with the Best Tacos and Enchiladas**

```
+------------------+-------------------+
|avg(rating_review)|    restaurant_name|
+------------------+-------------------+
|               5.0|          Tanakatsu|
|               5.0|Ichi_Sushi_and_Sa...|
|               4.5|SUSHISAMBA_Heron_...|
|               3.5|Flesh_Buns_Covent...|
|               3.0|     Ping_Pong_Soho|
+------------------+-------------------+
```

**Figure 18: Restaurants with the Best Sushi and Tempura**

```
+------------------+-------------------+
|avg(rating_review)|    restaurant_name|
+------------------+-------------------+
|               5.0|           BAO_Soho|
|               5.0|Steak_Co_Garrick_...|
|               5.0|Flat_Three_Restau...|
|               5.0|       Yuu_Kitchen|
|               5.0|              Chojo|
| 4.888888888888889|  The_Petite_Coree|
|4.055555555555555|Flesh_Buns_Covent...|
| 4.049019607843137|       Jinjuu_Soho|
|               4.0|         Yama_Momo|
| 3.559322033898305|        On_the_Bab|
|3.333333333333335|          Wagamama|
|               3.0|Temper_Covent_Garden|
|               2.0|       Ceviche_Soho|
+------------------+-------------------+
```

**Figure 19: Restaurants with the Best Korean Food and Bibimbap**

The above images were important to our project because it helped us recognize that the results were not useful for finding important or accurate insights. The remaining images in this section will showcase the best restaurants in London offering the best specific dishes. The following image (Figure 20) showcases the top 10

London restaurants with the best steak with the greater than 50 restaurant reviews constraint.

| avg(rating_review) | restaurant_name |
|---|---|
| 4.6794871795 | Smith_s |
| 4.6376811594 | Laurent_at_Cafe_Royal |
| 4.6046948357 | Steak_Co_St_Martins_Ln |
| 4.5907335907 | Blacklock_Soho |
| 4.55 | Melange |
| 4.5471698113 | Chez_Elles_Bistroquet |
| 4.4793713163 | Buenos_Aires_Argentine_Steakhouse |
| 4.4493464052 | Steak_Co_Garrick_Street_Covent_Gard |
| 4.4234693878 | The_Banc |
| 4.2878787879 | El_Vaquero_Whetstone |

**Figure 20: Restaurants with the Best Steak, >50 Constraint**

The following image (Figure 21) showcases the top 10 London restaurants with the best burger with the greater than 50 restaurant reviews constraint.

| avg(rating_review) | restaurant_name |
|---|---|
| 4.9305555556 | Bruce_Burgers |
| 4.4736842105 | Hard_Rock_Cafe_Piccadilly_Circus |
| 4.3375 | Dip_Flip |
| 4.3269230769 | Dirty_Bones_Shoreditch |
| 4.3080357143 | The_Banc |
| 4.2567567568 | Comptoir_Gascon |
| 4.2333333333 | Eat_17 |
| 4.1965811966 | Band_of_Burgers |
| 4.0948275862 | The_Fire_Station |
| 4.0530973451 | Mac_And_Wild_Fitzrovia |

**Figure 21: Restaurants with the Best Burger, >50 Constraint**

The following image (Figure 22) showcases the top 10 London restaurants with the best Chinese food with the greater than 50 restaurant reviews constraint.

| avg(rating_review) | restaurant_name |
|---|---|
| 4.2767857143 | Silk_Road |
| 4.2180451128 | Opium_Cocktail_Dim_Sum_Parlour |
| 4.1730769231 | The_Good_Earth |
| 4.1398601399 | Chinese_Cricket_Club |
| 4.0497925311 | Mandarin_Kitchen |
| 4.0388349515 | The_Duck_Rice |
| 4.0153256705 | Tao_Tao_Ju |
| 3.9945799458 | Wan_Chai_Corner |
| 3.993485342 | Gold_Mine |
| 3.5466666667 | Ping_Pong_Soho |

**Figure 22: Restaurants with the Best Chinese, >50 Constraint**

The following image (Figure 23) showcases the top 10 London restaurants with the best Indian food with the greater than 50 restaurant reviews constraint.

| avg(rating_review) | restaurant_name |
|---|---|
| 4.8367346939 | Kamasutra_Indian_Restaurant |
| 4.8131868132 | Shahi_Pakwaan |
| 4.810483871 | The_Sitara |
| 4.7413793103 | Nepal_Authentic_Dining |
| 4.734939759 | The_Jasmine_by_Spice_Nouv |
| 4.7198177677 | Indian_Express_West_Kensing |
| 4.7045454545 | Dhaba_49 |
| 4.6948356808 | Pure_Indian_Cooking |
| 4.6427503737 | Indian_Moment |
| 4.6078431373 | Radhuni_Southgate |

**Figure 23: Restaurants with the Best Indian, >50 Constraint**

The following image (Figure 24) showcases the top 10 London Pubs with the greater than 50 restaurant reviews constraint.

| avg(rating_review) | restaurant_name |
|---|---|
| 4.6617647059 | The_Peasant |
| 4.5905797101 | The_Blacksmith_s_Arms |
| 4.577092511 | The_Wilmington |
| 4.5405405405 | The_One_Tun |
| 4.5 | The_Duchy_Arms |
| 4.4867256637 | The_Black_Prince_Pub |
| 4.4297520661 | The_Nell_Gwynne_Tavern |
| 4.4137931034 | The_Old_Fountain |
| 4.3931623932 | The_Queen_s_Larder |
| 4.3423423423 | The_Dean_Swift |

**Figure 24: Best Pubs, >50 Constraint**

The following image (Figure 25) showcases the top London restaurants that play live music with the greater than 50 restaurant reviews constraint.

| avg(rating_review) | restaurant_name |
|---|---|
| 4.7986111111 | The_Sitara |
| 4.7358490566 | Cahoots_Underground |
| 4.6170212766 | Galvin_La_Chapelle |
| 4.3846153846 | Brasserie_Zedel |
| 4.3076923077 | Boisdale_of_Bishopsgate |
| 4.2586206897 | The_Real_Greek_Bankside |
| 4.145631068 | The_Piano_Works_Farringdon |
| 4.0396039604 | The_Archduke |

**Figure 25: Best Live Music, >50 Constraint**

The following image (Figure 26) showcases the top London Greek restaurants with the greater than 50 restaurant reviews constraint.

| avg(rating_review) | restaurant_name |
|---|---|
| 4.7467532468 | Liman_Restaurant |
| 4.7142857143 | CERU |
| 4.568627451 | Kibele_Restaurant_Bar |
| 4.56 | Ishtar |
| 4.5277777778 | Sappho_Meze_Bar |
| 4.3220338983 | Genzo_Greek |
| 4.2678571429 | Wine_and_Mousaka |
| 4.256097561 | The_Life_Goddess_Store_Stre |
| 4.2222222222 | Al_Santo_Restaurant |
| 4.1730769231 | Likya_Ocakbasi_Restaurant |

**Figure 26: Best Greek Restaurants, >50 Constraint**

The following image (Figure 27) showcases the top London Fried Chicken restaurants with the greater than 50 restaurant reviews constraint. The minimal number of reviews suggests that London restaurant goers do not tend to order or appreciate fried chicken.

| avg(rating_review) | restaurant_name |
|---|---|
| 4.3267973856 | BAO_Soho |
| 4.1 | Chick_n_Sours |
| 4.0642857143 | On_the_Bab |
| 4.0511627907 | Jinjuu_Soho |

**Figure 27: Best Fried Chicken Restaurants, >50 Constraint**

The following image (Figure 28) showcases the top London Sushi restaurants with the greater than 50 restaurant reviews constraint. It is interesting to note that the only restaurant that remains on the

greater than 50 constraint list as compared to Figure 13 without the constraint is Sushisamba Heron Tower.

| avg(rating_review) | restaurant_name |
|---|---|
| 4.7969639469 | Rock_Star_Sushi_Bar |
| 4.7804878049 | Laurent_at_Cafe_Royal |
| 4.6721311475 | Tomoe |
| 4.2678571429 | Yama_Momo |
| 4.1374045802 | Yashin_Sushi |
| 4.0735930736 | Ichi_Sushi_and_Sashimi_Bar |
| 4.0335811945 | SUSHISAMBA_Heron_Tower |
| 4 | Flesh_Buns_Covent_Garden |
| 3.9578947368 | Itsu |
| 3.9400921659 | Sake_no_Hana |

**Figure 28: Best Sushi Restaurants, >50 Constraint**

The following image (Figure 29) showcases the top London Korean restaurants with the greater than 50 restaurant reviews constraint. The minimal number of restaurant reviews suggests that London restaurant goers do not tend to order or appreciate Korean food.

| avg(rating_review) | restaurant_name |
|---|---|
| 4.7719298246 | The_Petite_Coree |
| 4.2097902098 | Flesh_Buns_Covent_Garden |
| 4.0759162304 | Jinjuu_Soho |
| 3.9162995595 | On_the_Bab |

**Figure 29: Best Korean Restaurants, >50 Constraint**

The following image (Figure 30) showcases the top London Thai restaurants with the greater than 50 restaurant reviews constraint.

| avg(rating_review) | restaurant_name |
|---|---|
| 4.7383419689 | Awesome_Thai_Barnes |
| 4.6428571429 | Yum_Sa |
| 4.5555555556 | Kaosarn |
| 4.51 | The_Blacksmith_s_Arms |
| 4.4736842105 | Jasmine_Thai_Restaurant |
| 4.41 | Poppy_s |
| 4.392 | The_Kings_Arms |
| 4.3307086614 | The_Lemon_Tree |
| 4.1972789116 | Monkey_and_Me |
| 4.1847507331 | The_Churchill_Arms |

**Figure 30: Best Thai Restaurants, >50 Constraint**

The following image (Figure 31) showcases the top London Lobster restaurants with the greater than 50 restaurant reviews constraint.

| avg(rating_review) | restaurant_name |
|---|---|
| 4.6949152542 | Smith_s |
| 4.6910569106 | Ormer_Mayfair |
| 4.6825396825 | The_Ledbury |
| 4.5344827586 | Seven_Park_Place_By_William_Drabble |
| 4.5 | Mere |
| 4.4736842105 | SUSHISAMBA_Heron_Tower |
| 4.2250489237 | Steak_And_Lobster_Warren_Street |
| 4.2205882353 | Scalini |
| 4.1050955414 | Mandarin_Kitchen |
| 4.037037037 | Berners_Tavern |

**Figure 31: Best Lobster Restaurants, >50 Constraint**

The following image (Figure 32) showcases the top London Italian restaurants with the greater than 50 restaurant reviews constraint.

| avg(rating_review) | restaurant_name |
|---|---|
| 4.9411764706 | Da_Moreno |
| 4.8831168831 | Pizza_Union_Hoxton |
| 4.8442622951 | Fine_Foods |
| 4.8415841584 | Mio_Restaurant |
| 4.8363636364 | Sfizio_Italian_Tapas |
| 4.8121546961 | Trattoria_Raffaele |
| 4.8050314465 | Oi_Vita_Pizzeria |
| 4.805 | Morso_Abbey_Road |
| 4.7727272727 | Tomo_Pizzeria_Restaurant |
| 4.7703927492 | Purezza_Camden |

**Figure 32: Best Italian Restaurants, >50 Constraint**

The following image (Figure 33) showcases the top London Fish and Chips restaurants with the greater than 50 restaurant reviews constraint.

| avg(rating_review) | restaurant_name |
|---|---|
| 4.8343949045 | Fishers_Restaurant_Takeaway |
| 4.7204724409 | Baileys_Fish_and_Chips |
| 4.6179775281 | Fishcotheque |
| 4.1090909091 | Toff_s |
| 4.0925925926 | The_Goat |
| 3.9885057471 | The_Cutty_Sark |
| 3.9655172414 | The_Old_Shades_London_Pub_Dining |
| 3.8 | Butlers_Wharf_Chop_House |
| 3.7976190476 | Garfunkel_s |

**Figure 33: Best Fish and Chips Restaurants, >50 Constraint**

London residents and tourists can look at the above tables identifying London's best restaurants to find a respectable place to eat. Considering the restaurant has to have over 50 restaurant reviews, each restaurant in the above tables is likely to be a trustworthy choice and should offer great food, service, and ambience.

# 6   Data Visualization

Each of the above figures depicting London's best restaurants with a specific food type were visualized individually on a map made in Tableau. The maps were made by collecting the latitude and longitude coordinates from the TripAdvisor Link provided for each restaurant review. Though each of the above figures were visualized, for the sake of this report we will choose to showcase the most interesting ones.

The following figure (Figure 34) showcases the map of London's best Italian restaurants. By viewing the map, we can see that Italian restaurants are well spread out across London, and not just centered in the heart of the city. As such, a restaurant manager may look to find a location in London where there is not a nearby Italian restaurant, as Italian restaurants seem to do well all across the city. Further, London restaurant goers wanting to eat good Italian food can look at this map to locate their nearest option.
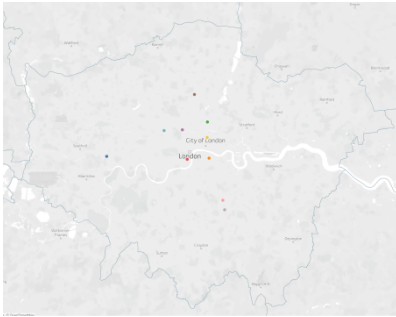
**Figure 34: Map of Best Italian Restaurants in London**

The following figure (Figure 35) showcases the map of London's best Chinese restaurants. By viewing the map, we can see that the best Chinese restaurants are located only in the heart of London. As such, a restaurant manager may look to open up a Chinese restaurant in the heart of the city for optimal success. Further, London restaurant goers can look at this map and expect to find the best Chinese food in downtown London.
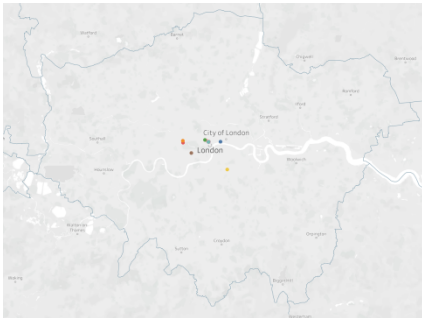


**Figure 35: Map of Best Chinese Restaurants in London**

The following figure (Figure 36) showcases the map of London's best Lobster restaurants. By viewing the map, we can see that the best Lobster restaurants are located in the heart of the city. Upon a closer view, we also see that the restaurants are located near a body of water. Perhaps London restaurants prefer to order Lobster at these restaurants because it is fresh.
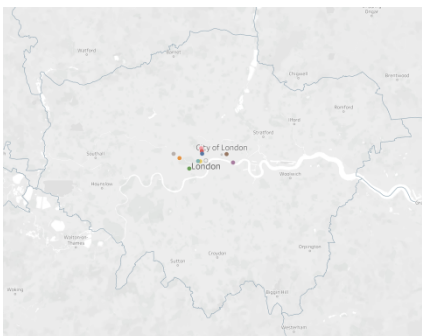


**Figure 36: Map of Best Lobster Restaurants in London**

The following figure (Figure 37) showcases all of the restaurants identified to have the best dishes or ambience. Figure 38 showcases

the key for Figure 37. The purpose of the compiled map was to see if common restaurants formed groups. From this map, we can see that there is not an obvious common pattern between any of the restaurant types. As such, a London restaurant goer can expect to find a well-liked restaurant all across London.
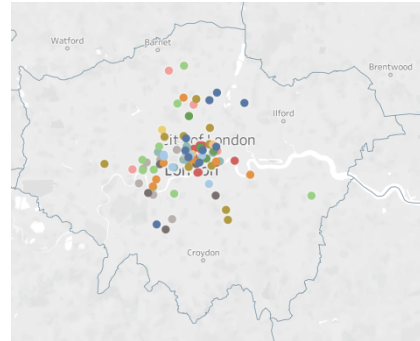


**Figure 37: Compiled Best Restaurants in London**



**Figure 38: Compiled Best Restaurants in London Key**

One additional map that we think provides insightful information regarding the best restaurants in London is the map containing the restaurants with the most 5 star reviews. Figure 39 showcases the names of the restaurants and the counts of the five star reviews. Figure 40 showcases the map of London's most well-liked restaurants. From the map, we can see that the most liked restaurants are located in the heart of the city. This would make sense as downtown London is the most populated part of the city, and also attracts the most tourists.

| restaurant_name | count(rating_review) | rating_review |
|---|---|---|
| Galvin_La_Chapelle | 2,318 | 5 |
| The_Ledbury | 2,182 | 5 |
| SUSHISAMBA_Heron_Tower | 1,994 | 5 |
| Cahoots_Underground | 1,573 | 5 |
| Little_India | 1,471 | 5 |
| The_Monsoon | 1,393 | 5 |
| Hard_Rock_Cafe_Piccadilly_Circus | 1,370 | 5 |
| Smith_s | 1,309 | 5 |
| Steak_Co_Garrick_Street_Covent_Garden | 1,205 | 5 |
| Madison | 1,181 | 5 |

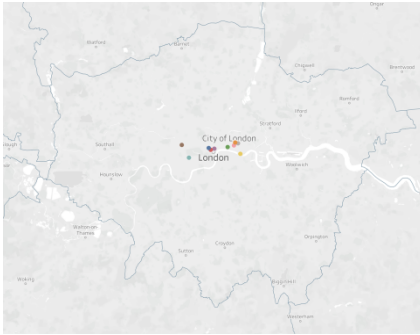**Figure 39: London Restaurants with Most 5-Star Reviews**

**Figure 40: Map of London Restaurants with Most 5-Star Reviews**

Most of the time when people leave restaurant reviews, it is because they had a really good experience or a really bad experience. We wanted to check this theory within our dataset. Using tableau, we created a graph that identifies the number of reviews given each ranking. From the graph below (Figure 41), we can see that many people gave reviews of a 5-star experience, but very few gave reviews of a terrible 1-star experience.
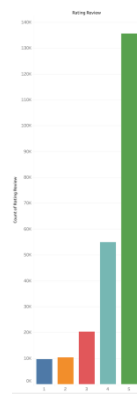


**Figure 41: Dispersion of Restaurant Review Rankings**

Another visualization we made was to showcase the above theory, but for one restaurant. Galvin La Chapelle is London's restaurant that received the most 5-star reviews. From the graph below (Figure 42), we see an identical theme to that of Figure 41. The most reviews were left for a 5-star review, and the least amount of reviews were left for a 1-star review.
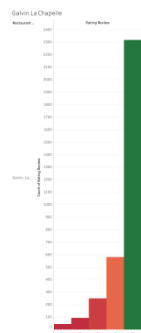


**Figure 42: Galvin La Chapelle Review Dispersion**

Lastly, a word cloud was created for the positive reviews in the file (Figure 43). We can see positive words such as "excellent," "lovely," "delicious," and "fantastic." These words are as expected for a positive review.



**Figure 43: Word Cloud for Positive Reviews**

## 7 Conclusion

In conclusion, the results of this study show that a model can be created to determine the sentiment of restaurant reviews as positive, negative, or neutral with reasonably high accuracy. The tables in this project show that the best dishes and food types in a city can easily be identified. The maps show that the best dishes and food types in a city can found. Both the tables identifying the restaurants with the best dishes and the maps locating such restaurants are helpful to the restaurant manager looking to open up a new restaurant. The restaurant manager can see where certain food types do and do not do well. The restaurant manager can also see which food types the restaurant goers of London value and like. For example, many of the highest ranked restaurants were Michelin rated, and were very fancy and expensive. This project has potential to have a vast impact on the restaurant industry, as current restaurant review websites such as TripAdvisor or Yelp do not showcase interesting data insights or maps.

## REFERENCES

[1] Botana, Inigo, Six TripAdvisor Datasets for NLP Tasks; https://www.kaggle.com/datasets/inigolopezrioboo/a-tripadvisor-dataset-for-nlp-tasks