Analyzing Google Map Reviews of Cheesecake Factory Restaurants in New Jersey

Group 4

Group Members: Heng Zhao, Zimeng Zhao, Wencheng Qiu

Introduction

In today's digital age, online reviews hold significant importance when it comes to judging a business's reputation and customer satisfaction. They provide valuable feedback to businesses and influence public opinion, thus affecting their performance.

Our project aims to analyze customer feedback from online sources to extract valuable insights.By applying advanced machine learning techniques for sentiment analysis and trend identification, this study seeks to unravel the nuances of customer feedback, ranging from service quality and food taste to the overall ambiance of the restaurants. The objective is not only to understand what customers are saying but also to identify why they are saying it. This includes dissecting both positive and negative feedback to gain a holistic view of the customer experience. In addition, our analysis goes beyond just interpreting data. Our final goal is to turn these insights into practical recommendations for The Cheesecake Factory. We hope to help improve customer satisfaction, enhance service quality, and maintain a competitive advantage in New Jersey's ever-changing restaurant industry.

Research Questions

1. Customer Satisfaction and Dissatisfaction:
    - What specific factors contribute to customer satisfaction or dissatisfaction at The Cheesecake Factory restaurants? This includes an exploration of food quality, service, ambiance, and overall dining experience.
    - How do these aspects compare across different Cheesecake Factory locations? We aim to identify the strengths and weaknesses of each Cheesecake Factory location to understand why some locations are better than others in the eyes of customers.

2. Temporal Changes in Reputation:
    - How has the reputation of each Cheesecake Factory restaurant evolved over time concerning service quality, food taste, and customer experience?
    - We seek to assess both areas where each restaurant has improved over time and areas of decline in customer satisfaction.

To effectively investigate and address these research questions, we utilized the following methodologies:

I. **Sentiment Analysis Method:** Implement a method to analyze the sentiment of each sentence in every customer comment, categorizing it as positive, negative, or neutral.

II. **Topic Categorization Method:** Develop a method to categorize sentences in reviews accurately, identifying common themes and subjects addressed by customers like food, service, atmosphere, etc.

III. **Data Presentation Strategy:** Design a strategy that clearly presents the data and information we have mined. This will involve creating visualizations and reports that succinctly convey our findings, making it easier for both customers and restaurant managers to understand the key insights from the analysis.

Dataset Collection

Data collection process for our analysis is a critical first step. It involves systematically gathering the necessary review data from the web. We use web scraping tools to gather all the store ratings and customer reviews.

Firstly, we created a “getRestaurantsLinks” function to extract the URLs of different Cheesecake Factory restaurant pages. These URLs are then added to a list, which serves as the

foundation for data collection in subsequent stages. The list is important because it provides a direct route to each restaurant's specific page, where customer reviews are posted. Then, we extract the comments through the “getReview” function. It uses the URL to navigate to the detailed page for each restaurant. Once you reach the page, the feature is designed to automatically access the comments section. To ensure a comprehensive collection of comments, the feature is programmed to perform 3,000 to 4,000 automatic rolls. In the end, we collected 15000 customer reviews from a total of 10 restaurants.

Dataset Processing

During the data processing phase, we apply specially designed functions to the customer review dataset.

Firstly, we developed an “Emoji Removal Function” to simplify textual analysis. We noticed that emojis, although expressive, could complicate the linguistic analysis process. This function efficiently strips away emojis from the review text data, ensuring that the remaining text is primed for more precise linguistic analysis. Then, we created a “Sentence Splitting Function” to enable a more detailed analysis of customer feedback. This function breaks down review content into individual sentences, allowing for more detailed sentiment analysis at the sentence level instead of just the broader review level. Thirdly, we created a “Tokenization Function” to further dissect the text data. This function is customizable and it tokenizes the text data by breaking it down into individual words or tokens. This process helps to identify key themes and words within the reviews and also facilitates deeper linguistic analysis. Finally, we developed a function to extract all sentences from the reviews and compile them into a single, consolidated list. This approach was crucial in creating an extensive dataset for analysis. After

post-processing, we were able to successfully compile a total of 21,254 individual customer review sentences, each offering unique insights into customer experiences and perceptions.

Dataset Labeling

Data labeling is very important. For unsupervised learning, we need datasets to evaluate the performance of our models; for supervised learning, a large dataset is an indispensable part of the training process.

In order to label data, we wrote an automatic data labeling script using OPENAI's GPT-3.5 turbo API, and set a global Custom Instruction to control GPT to label data for us. The content of the custom instruction is as follows:

```
"role": "system",
"content": "You are a professional comment dataset tagger and can find the most suitable label for each \
comment. Now I will give you some comments about restaurant, you will give them a label represent their \
topic, there are eight topics, food, hygiene, atmosphere, service, location, parking, transportation \
and none. So no matter what I tell you, just answer one of these eight words"
```

With this setup, we can instruct GPT to categorize individual sentences obtained from our data processing into eight main categories: food, hygiene, atmosphere, service, location, parking, transportation, and None, which refers to no clear theme.

Dataset Clean

GPT is a language generation model that mimics human language logic using confidence levels, and therefore makes the same kinds of mistakes as humans. In the continuous labeling of 20,000 data entries, GPT may make errors in the capitalization of some data, include commas and periods in the results beyond what is required for labels, and even produce multi-label outputs that are not permitted. In addition to this, we observed that comments on the topics of

parking and transportation are very few, numbering only in the double digits. Such a small amount of data can adversely affect the training effectiveness. Therefore, we need to perform the following data cleansing tasks:

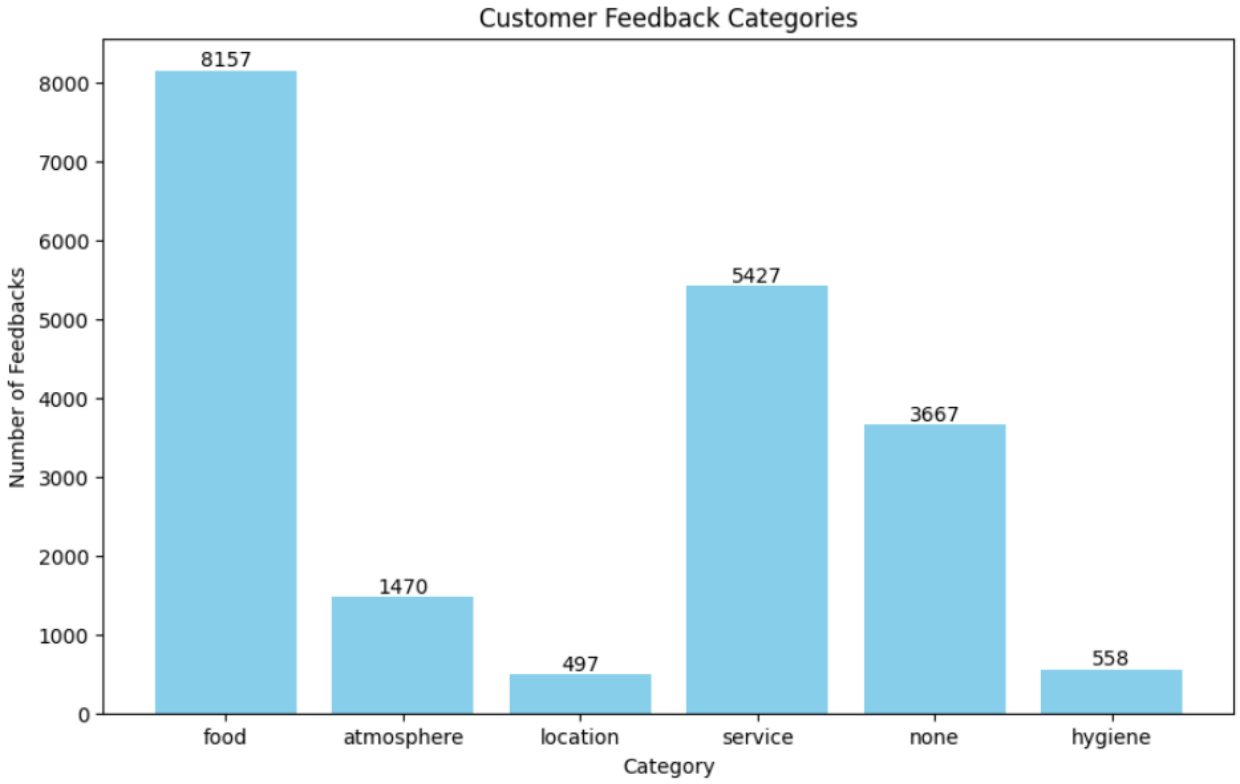
1. Lowercase all the labels.
2. Use Regular Expression to handle the irregular labels.

```
label = re.sub(r'^[^a-zA-Z]+|[^a-zA-Z]+$', '', data[1])
```

3. delete multiple labels.
4. Combine categories with very few numbers and are harder to classify.

```
elif label == "location" or label == "parking" or label == "transportation":
    X.append([sentense,"location"])
```

After data cleansing, we obtained the following data entries.



Dataset Balance

From the results after data cleansing, it is evident that our data is still imbalanced. This imbalance can significantly affect the performance of learning algorithms:

1. Bias Towards Majority Classes: Models trained on imbalanced datasets often favor majority classes, leading to poor recognition of minority classes (He & Garcia, 2009).
2. Weak Generalization: Such models struggle to generalize well to new data, particularly for the underrepresented classes (Haixiang et al., 2017).
3. Inaccurate Feature Significance: Imbalanced data can result in overestimating the importance of features associated with the majority class, reducing the model's effectiveness (Krawczyk, 2016).
4. Misleading Evaluation Metrics: Traditional metrics like accuracy can be misleading in imbalanced settings, as they might not accurately reflect the model's performance on minority classes (Jeni et al., 2013).

SMOTE (Synthetic Minority Over-sampling Technique) is a widely-used method for addressing the issue of class imbalance in supervised learning. Developed by Chawla et al. (2002), SMOTE works by creating synthetic samples from the minority class, rather than simply over-sampling with replacement. This is achieved by randomly selecting a point from the minority class and then creating synthetic instances by interpolating between this point and its neighbors. We will use this method to augment our data, and observe the effectiveness of this data balancing approach in our subsequent model research.

```
class SMOTE(
    *,
    sampling_strategy: str = "auto",
    random_state: Any | None = None,
    k_neighbors: int = 5,
    n_jobs: Any | None = None
)
```

Class to perform over-sampling using SMOTE.

This object is an implementation of SMOTE - Synthetic Minority Over-sampling Technique as presented in [1]_.

Read more in the `User Guide <smote_adasyn>` .

Unsupervised Model Research

Our model research begins with unsupervised learning. In our assignment 5, the topic classification method based on three clustering models achieved considerable accuracy. Therefore, in this part, we will study the classification effects of the K-means, Gaussian Mixture Model, and LDA models that we used previously in this topic.

1. K-means

   Whether to use SMOTE to balance the dataset: Yes

   Our tested best TF-IDF parameter setting is:

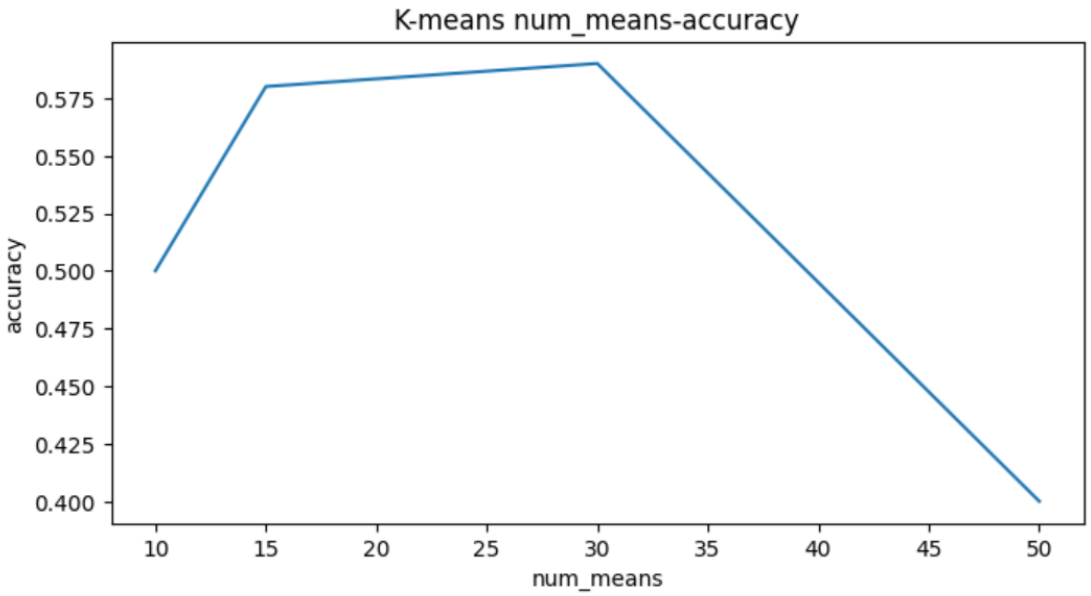
stop_words	min_df	lowercase	Others
“English”	0.0001	True	default

   Our tested best KMeansCluster parameter setting is:

num_means	distance	repeats	Others
30	cosine_distance	1	default

   Explanation:

We applied the “majority vote” rule to dynamically map the predicted cluster IDs to real labels. So the more number of clusters the less likely there are several topics in one cluster. Guided by this principle, we tested a variety of num_means settings, resulting in the following line graph.



Our model's final accuracy is as shown in the following graph:

atmosphere	0.35	0.21	0.26	277
food	0.68	0.74	0.71	1641
hygiene	0.51	0.32	0.39	107
location	0.58	0.37	0.45	94
none	0.44	0.37	0.40	763
service	0.58	0.67	0.62	1074
accuracy			0.59	3956
macro avg	0.52	0.45	0.47	3956
weighted avg	0.58	0.59	0.58	3956

2. Gaussian Mixture Model

Whether to use SMOTE to balance the dataset: No

Our tested best TF-IDF parameter setting is:

stop_words	min_df	max_features	lowercase	Others
“English”	0.001	1000	True	default

Our tested best GaussianMixture parameter setting is:

n_components	covariance_type	n_init	Others
range(5,20)	cv_type	10	default

Explanation:

Since the GMM (Gaussian Mixture Model) method requires creating and storing a large probability distribution matrix in memory, we decided to take a subset of the dataset, randomly selecting 500 data points from each category, and replicating them if insufficient. In addition, we need to limit the max_features, as too many features can lead to memory overflow. To balance performance and accuracy, we set the n_components to 5-20 and n_init to 10. GMM performs poorly in this scenario because most individual sentences have token counts in the single digits, but TF-IDF is a matrix with thousands of features, resulting in a very sparse matrix. There is no guarantee that this high-dimensional sparse matrix follows Gaussian Distribution.

Our model's final accuracy is as shown in the following graph:

	precision	recall	f1-score	support
atmosphere	0.31	0.39	0.34	101
food	0.28	0.50	0.36	103
hygiene	0.38	0.24	0.29	95
location	0.49	0.40	0.44	107
none	0.32	0.24	0.28	95
service	0.27	0.19	0.22	99
accuracy			0.33	600
macro avg	0.34	0.33	0.32	600
weighted avg	0.34	0.33	0.33	600

This is not a good model.

3. Latent Dirichlet Allocation

   Whether to use SMOTE to balance the dataset: Yes

   Our tested best TF-IDF parameter setting is:

stop_words	min_df	max_features	lowercase	ngram_range	Others
“English”	0.005	3000	True	(1,1)	default

Our tested best LatentDirichletAllocation parameter setting is:

n_components	max_iter	n_jobs	Others
30	40	1	default

For the LDA cluster, we use two ways to build the mapping between the real topic and the generated topic. One is only keeping the topic with the highest score and then using vote to

build the mapping. Another is keeping all topics and their scores and using weighted voting to build this mapping.

Below are the result displays and comparison charts for the two mapping models:

1. Only keeping the topic with the highest score and then using vote to build the mapping:

	precision	recall	f1-score	support
atmosphere	0.00	0.00	0.00	277
food	0.53	0.75	0.62	1641
hygiene	0.00	0.00	0.00	107
location	0.00	0.00	0.00	94
none	0.39	0.39	0.39	763
service	0.48	0.39	0.43	1074
accuracy			0.49	3956
macro avg	0.23	0.25	0.24	3956
weighted avg	0.42	0.49	0.45	3956

2. keeping all topics and their scores and using weighted voting to build this mapping:

	precision	recall	f1-score	support
atmosphere	0.16	0.06	0.09	277
food	0.56	0.59	0.58	1641
hygiene	0.13	0.33	0.19	107
location	0.24	0.22	0.23	94
none	0.28	0.17	0.21	763
service	0.37	0.45	0.41	1074
accuracy			0.42	3956
macro avg	0.29	0.30	0.28	3956
weighted avg	0.41	0.42	0.41	3956

Although Method 1 can achieve higher accuracy, this accuracy is primarily provided by the largest category, 'food,' with the recall rate for many other categories being 0, meaning this model can only identify certain categories. Method 2, while less accurate than Method 1, is capable of recognizing all six categories. The performance of both models is not ideal.

Supervised Model Research

Many studies on topic classification have mentioned the combination of unsupervised and supervised learning methods (Daniel et al., 2009), using the results of unsupervised learning as features and classifying these features with supervised learning methods. Based on this concept, in order to further enhance the performance of our classification model, we have decided to use SVM (Support Vector Machine) and neural network methods to classify the topic probability distribution results of LDA (Latent Dirichlet Allocation) to obtain the true topic labels of sentences.

1. LDA+SVM

   Whether to use SMOTE to balance the dataset: Yes

   Our tested best TF-IDF parameter setting is:

stop_words	min_df	max_features	lowercase	ngram_range	Others
“English”	0.005	3000	True	(1,1)	default

   Our tested best LatentDirichletAllocation parameter setting is:

n_components	max_iter	n_jobs	Others
30	40	1	default

We used OneVsRestClassifier and LinearSVC to train and test on the LDA results, and the classification results are as follows:

	precision	recall	f1-score	support
atmosphere	0.00	0.00	0.00	277
food	0.61	0.34	0.43	1641
hygiene	0.00	0.00	0.00	107
location	0.00	0.00	0.00	94
none	0.00	0.00	0.00	763
service	0.57	0.12	0.20	1074
micro avg	0.60	0.17	0.27	3956
macro avg	0.20	0.08	0.11	3956
weighted avg	0.41	0.17	0.23	3956
samples avg	0.17	0.17	0.17	3956

Accuracy: 0.172143579373104 15

This is a very poor classification result, which is just slightly better than a random classification result. Therefore we will not do further research on it.

2. LDA+ANN

Whether to use SMOTE to balance the dataset: Yes

Our tested best TF-IDF parameter setting is:

stop_words	min_df	lowercase	ngram_range	Others
“English”	0.005	True	(1,1)	default

Our tested best LatentDirichletAllocation parameter setting is:

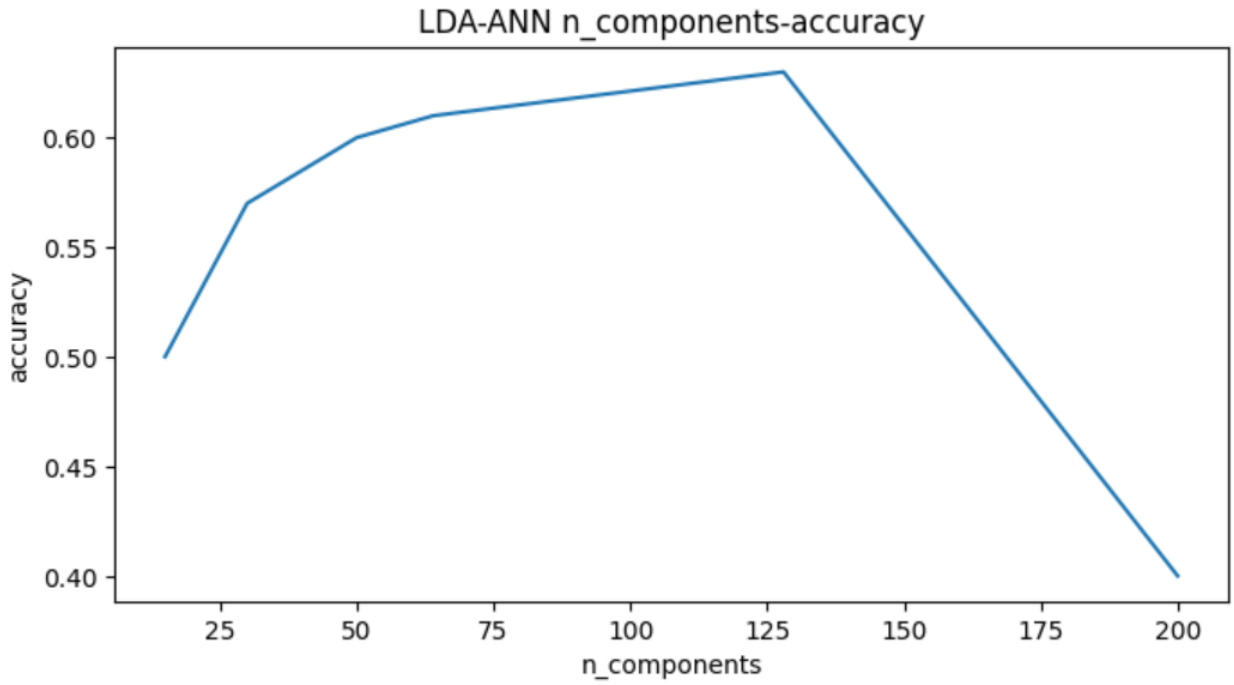
n_components	max_iter	n_jobs	Others
128	40	1	default

Our tested best ANN configuration is:

```
    model = Sequential([
    Dense(16, activation='relu', input_shape=(train_topic_distributions.shape[1],)),
    Dense(32, activation='sigmoid'),
    Dense(32, activation='relu'),
    Dense(6, activation='softmax')
])
```

Explanation:

The most important parameter in this model is n_components, for which we chose 128. This is because in our multiple experiments, we found that the finer the granularity of clustering, i.e., the more clusters, the better our ANN classification performance. However, beyond 128, the classification accuracy rapidly declines, so we chose 128 as our optimal parameter.



The classification results are as follows:

	precision	recall	f1-score	support
atmosphere	0.42	0.30	0.35	277
food	0.73	0.77	0.75	1641
hygiene	0.40	0.19	0.25	107
location	0.49	0.36	0.41	94
none	0.51	0.57	0.54	763
service	0.66	0.64	0.65	1074
micro avg	0.64	0.64	0.64	3956
macro avg	0.53	0.47	0.49	3956
weighted avg	0.63	0.64	0.63	3956
samples avg	0.64	0.64	0.64	3956

Accuracy: 0.6382709807886754

The accuracy of 63.8% is much higher than 49% and 42% of two vote methods. The idea of using topic distributions extracted with LDA as the feature to predict the real classification of sentences has finally started to show its value.

In fact, an accuracy rate of 63% still does not meet our expectations. We continue to explore other topic classification models in an attempt to achieve higher accuracy. Next, we decided to directly use ANN and CNN to classify the TF-IDF feature vector of a document to explore this possibility.

1. ANN

   Whether to use SMOTE to balance the dataset: No

   Our tested best TF-IDF parameter setting is:

stop_words	min_df	lowercase	ngram_range	Others
“English”	0.005	True	(1,1)	default

Our tested best ANN configuration is:

```
    model = Sequential([
    Dense(16, activation='relu', input_shape=(train_dtm.shape[1],)),
    Dense(32, activation='sigmoid'),
    Dense(32, activation='relu'),
    Dense(6, activation='softmax')
])
```

Explanation:

Although it may seem unusual, directly using an ANN (Artificial Neural Network) to classify TF-IDF feature vectors without data balancing results in better performance. The reason for this phenomenon is that a model balanced for minority categories is more likely to overfit on those categories, while simplifying the model parameters and reducing the number of training iterations can lead to underfitting in the majority categories. Therefore, in this section of model training, we will not use SMOTE to balance the dataset. The optimal ANN model configuration remains consistent with the previous section and does not require adjustment.

The classification results are as follows:

	precision	recall	f1-score	support
0	0.49	0.28	0.36	277
1	0.84	0.81	0.82	1641
2	0.62	0.26	0.37	107
3	0.59	0.44	0.50	94
4	0.49	0.73	0.59	763
5	0.78	0.70	0.74	1074
micro avg	0.71	0.71	0.71	3956
macro avg	0.64	0.54	0.56	3956
weighted avg	0.72	0.71	0.70	3956
samples avg	0.71	0.71	0.71	3956

Accuracy: 0.705005055611729

2. CNN

Whether to use SMOTE to balance the dataset: No

Our tested best TF-IDF parameter setting is:

stop_words	min_df	lowercase	ngram_range	Others
“English”	0.005	True	(1,1)	default

Our tested best CNN configuration is:

```
model = Sequential()

model.add(Conv1D(filters=64, kernel_size=3, activation='relu', input_shape=(train_dtm.shape[1],1)))
model.add(MaxPooling1D(pool_size=2))
model.add(Flatten())

model.add(Dense(32, activation='relu'))
model.add(Dropout(0.6))
model.add(Dense(6, activation='softmax'))
```

Explanation:

CNNs use convolutional and pooling layers to gather local features of data, which is similar to the LDA layer in the LDA+ANN model. In this model, we set the size of the convolutional kernel to 3, as we tested larger kernels and found that the impact was minimal and could even pose a risk of overfitting. The appropriate size of our hidden layers can extract sufficient information, and setting the dropout rate of the dropout layer to 0.6 effectively suppresses overfitting. This model, based on extensive testing, achieved the best results among all the models we tested.

The classification results are as follows:

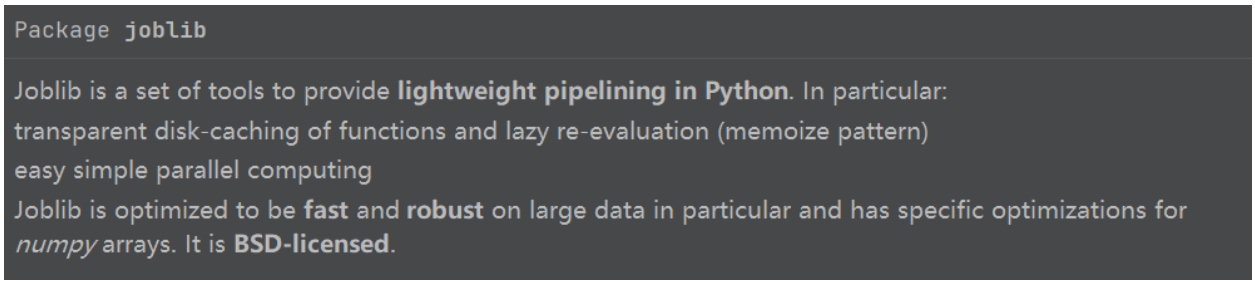
	precision	recall	f1-score	support
0	0.55	0.29	0.38	277
1	0.82	0.85	0.83	1641
2	0.78	0.27	0.40	107
3	0.60	0.40	0.48	94
4	0.52	0.67	0.58	763
5	0.76	0.73	0.74	1074
micro avg	0.71	0.71	0.71	3956
macro avg	0.67	0.53	0.57	3956
weighted avg	0.72	0.71	0.71	3956
samples avg	0.71	0.71	0.71	3956

Accuracy: 0.7143579373104145

We will use this model to complete the topic classification for the entire dataset as well as future data.

Packing of the model and its components

To ensure our model can be used for subsequent data, we need to save the dependencies of the model, which include LabelBinarizer, TfidfVectorizer, and MinMaxScaler. We will use the joblib library to package these trained instances, ensuring the model's usability.




```
dump(lb, 'my_lb.joblib')
```

```
dump(tfidf, 'tfidf_vectorizer.joblib')
```

```
dump(scaler,"my_scaler.joblib")
```

The trained model will be saved as 'my_CNN_TOPIC.h5'.

```
model.save("my_CNN_TOPIC.h5")
```

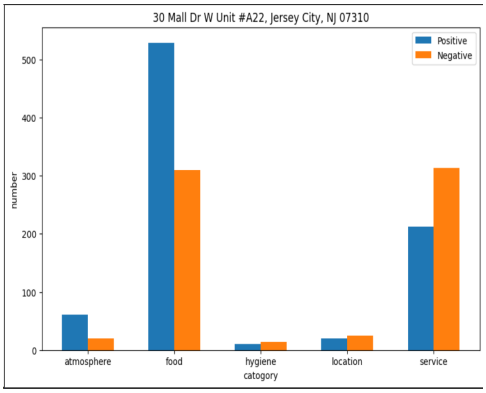
Result Analyzation

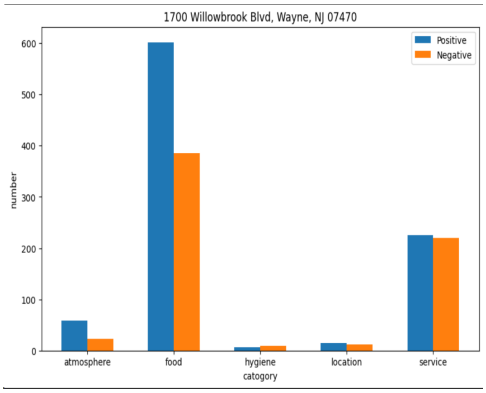
1. Positive & Negative Aspects Of Various Types Of Restaurants

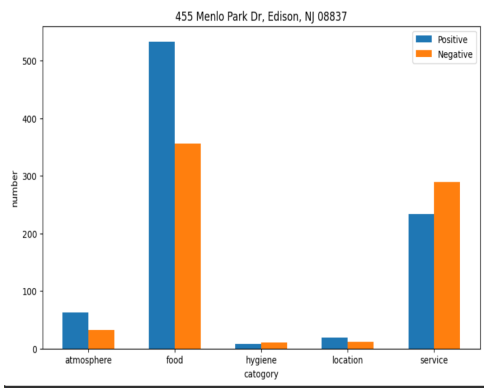
After thorough research, we categorized all the reviews into six distinct groups. Our focus then shifted to a detailed analysis of reviews from ten selected restaurants. To start, we compiled aggregate data for each establishment. Within this dataset, we established a structure

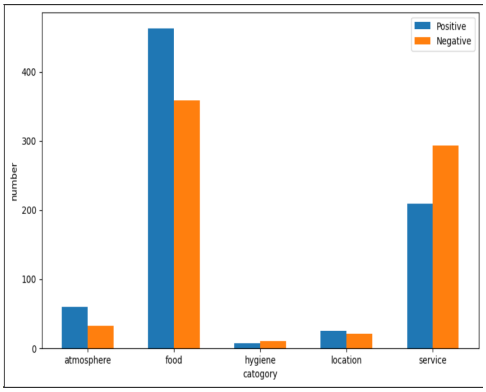
(labeled "overall" for each restaurant) to track the frequency and sentiment (positive or negative) of specific labels. This involved meticulously iterating through each review and sentence, updating counts and sets based on the identified sentiment and topic. The culmination of our analysis was the creation of bar charts, which effectively illustrate the distribution of sentiments—both positive and negative—across various categories for each restaurant.

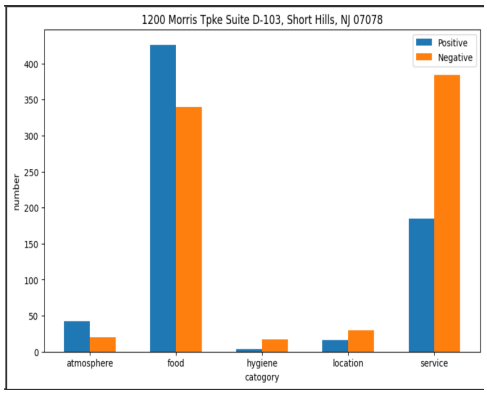
The reason why we didn't include a 'none' category in our classification is that our objective was to garner clearer and more accurate keywords specific to each defined category, thereby avoiding any confusion that might arise from ambiguous keywords. Moreover, upon scrutinizing all comments initially falling under the 'none' category, we found that they indeed lacked meaningful attributes that would warrant a separate classification. This approach ensured that our analysis remained focused and relevant to the distinct characteristics of each identified category.

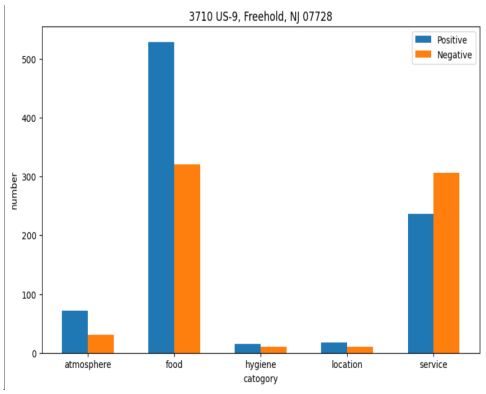


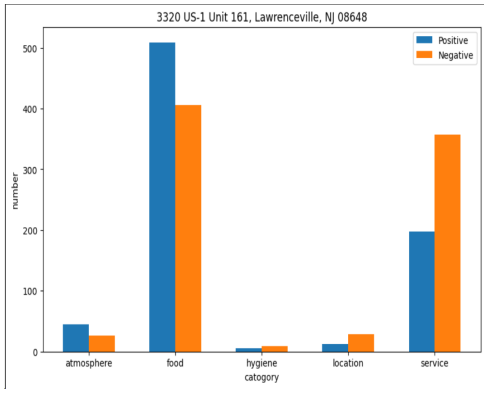


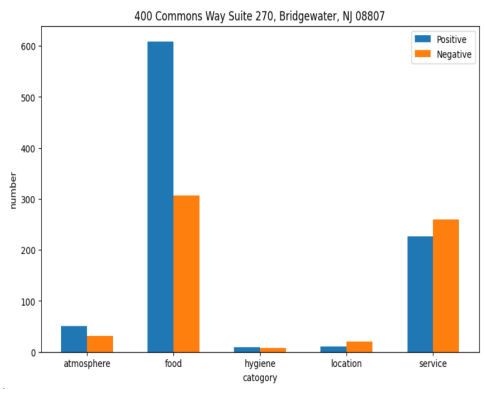

455 Menlo Park Dr, Edison, NJ 08837
Positive
Negative
number
atmosphere
food
hygiene
location
service
catogory

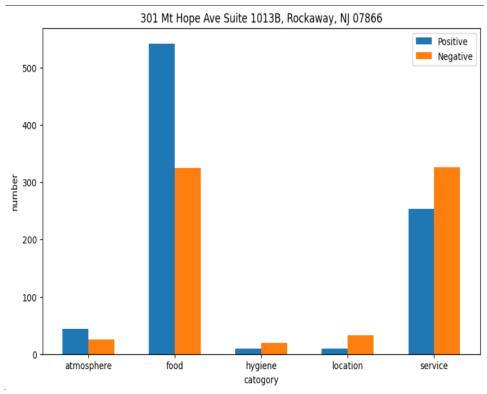

Positive
Negative
number
atmosphere
food
hygiene
location
service
catogory

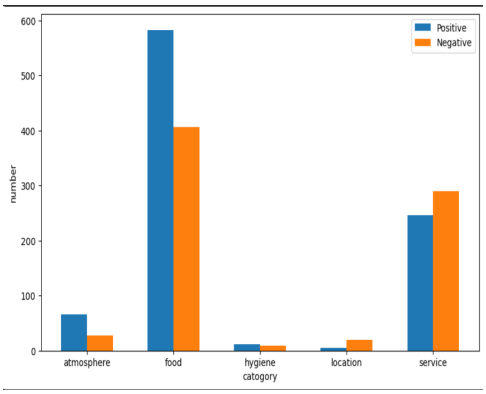

1200 Morris Tpke Suite D-103, Short Hills, NJ 07078
Positive
Negative
number
atmosphere
food
hygiene
location
service
catogory


3710 US-9, Freehold, NJ 07728
Positive
Negative
number
atmosphere
food
hygiene
location
service
catogory


3320 US-1 Unit 161, Lawrenceville, NJ 08648
Positive
Negative
number
atmosphere
food
hygiene
location
service
catogory


400 Commons Way Suite 270, Bridgewater, NJ 08807
Positive
Negative
number
atmosphere
food
hygiene
location
service
catogory

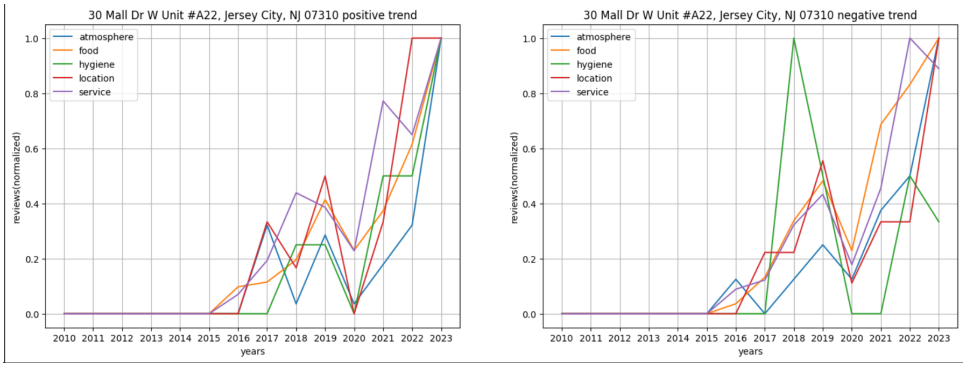

301 Mt Hope Ave Suite 1013B, Rockaway, NJ 07866
Positive
Negative
number
atmosphere
food
hygiene
location
service
catogory

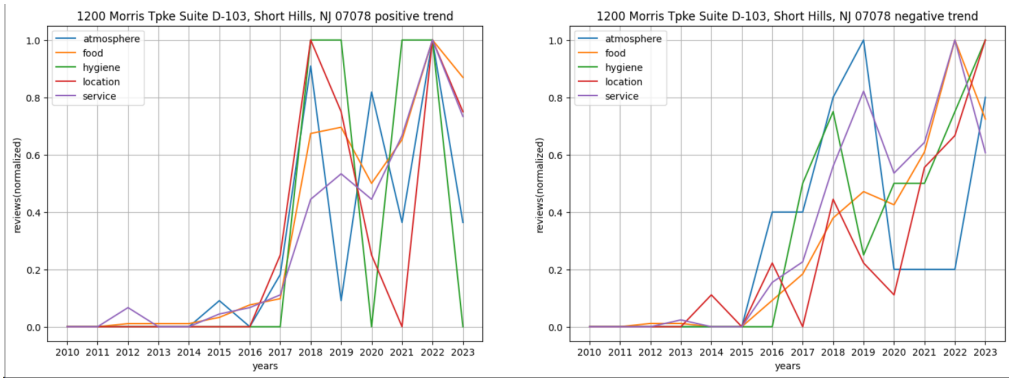

Positive
Negative
number
atmosphere
food
hygiene
location
service
catogory

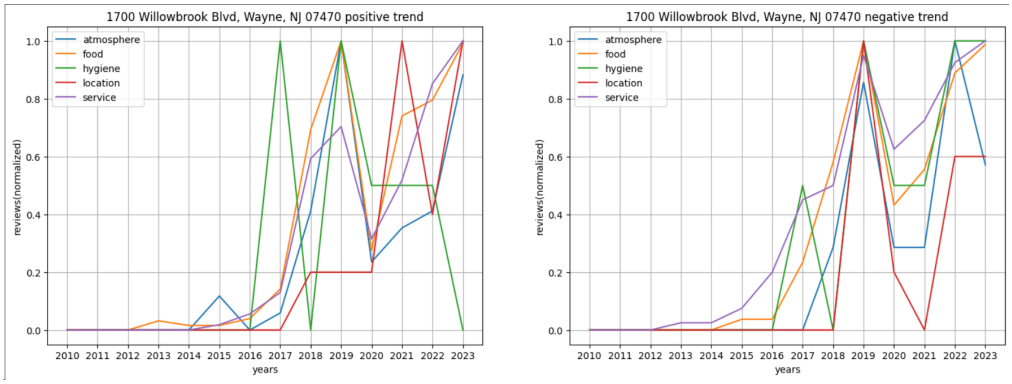
2. General Overview Of Restaurant Reviews Categorized By Year

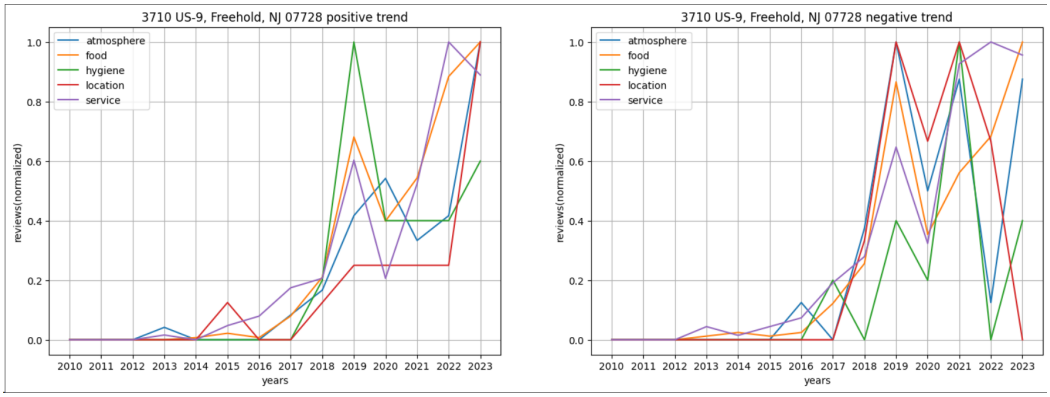
Our analysis revealed that the bar charts effectively depict the distribution of positive and negative comments across various categories for each restaurant. However, this leads to pertinent questions regarding the accuracy and temporal dynamics of these observations. For instance, could there be variations in comments attributable to different years? It's essential to consider whether a restaurant's overall rating has shown improvement or decline in a specific year. Such temporal analysis could provide deeper insights into the evolving nature of customer feedback and the restaurant's performance over time.

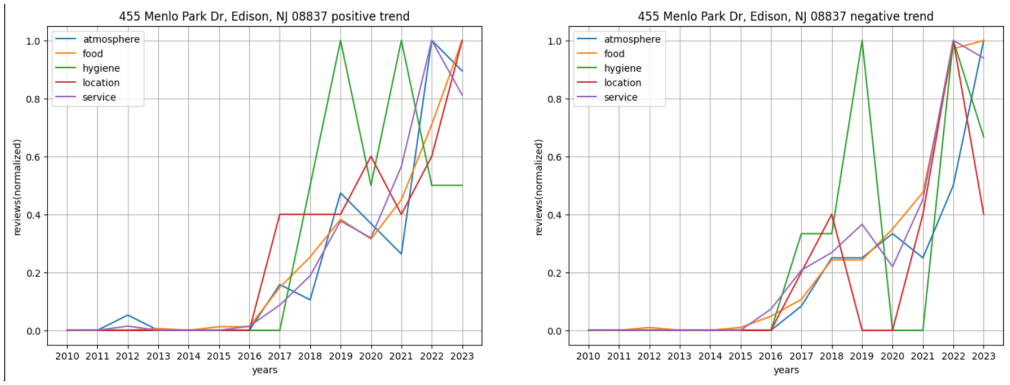
To address these questions, we developed a timeline structure within each restaurant's dataset. This structure is designed to track the frequency of positive and negative sentiments across various categories, spanning from 2010 to 2023. During our analysis, a noticeable decline in negative reviews was observed for all restaurants between 2020 and 2021. Initially, this might suggest an improvement in service quality. However, a closer examination revealed that the overall volume of restaurant reviews had significantly decreased during this period, primarily due to the impacts of the COVID-19 pandemic. Consequently, this unusual trend in the data necessitates caution in interpreting the charts as they might not accurately represent the true nature of positive and negative feedback over the years.

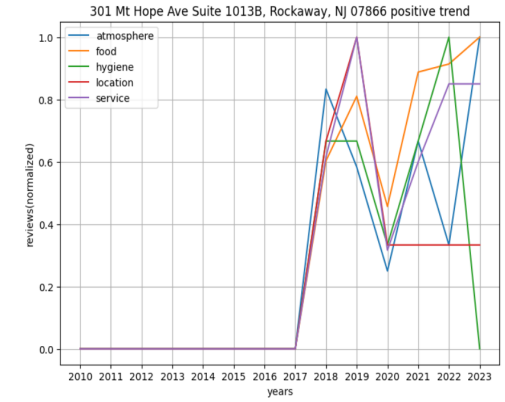


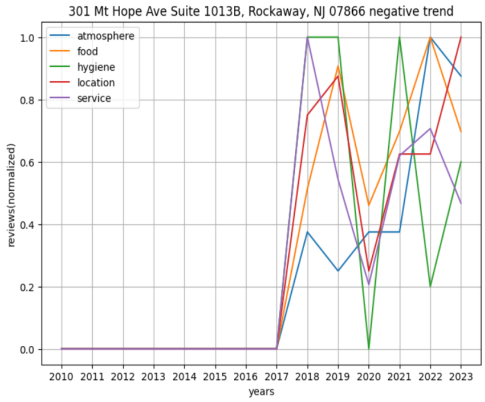

1200 Morris Tpke Suite D-103, Short Hills, NJ 07078 positive trend
1200 Morris Tpke Suite D-103, Short Hills, NJ 07078 negative trend
atmosphere
food
hygiene
location
service
reviews(normalized)
years

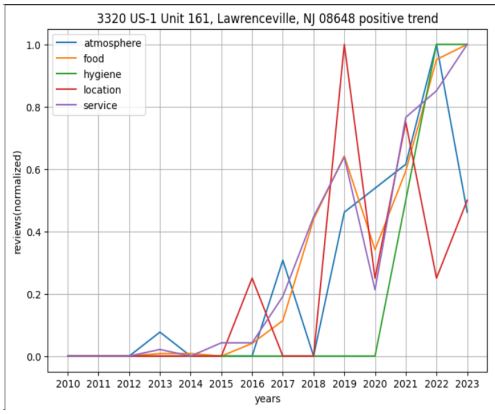

1700 Willowbrook Blvd, Wayne, NJ 07470 positive trend
1700 Willowbrook Blvd, Wayne, NJ 07470 negative trend
atmosphere
food
hygiene
location
service
reviews(normalized)
years

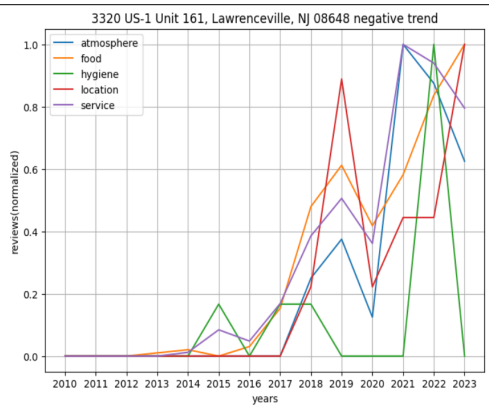

3710 US-9, Freehold, NJ 07728 positive trend
3710 US-9, Freehold, NJ 07728 negative trend
atmosphere
food
hygiene
location
service
reviews(normalized)
years

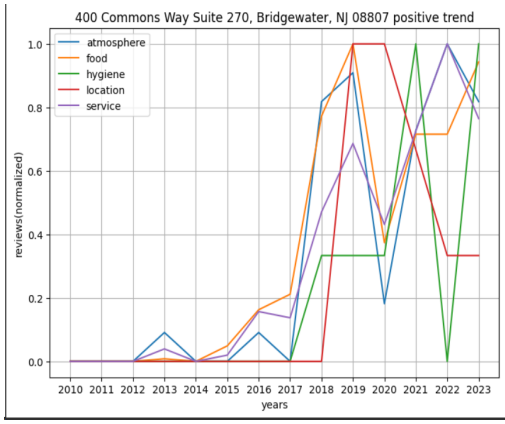

455 Menlo Park Dr, Edison, NJ 08837 positive trend
455 Menlo Park Dr, Edison, NJ 08837 negative trend
atmosphere
food
hygiene
location
service
reviews(normalized)
years

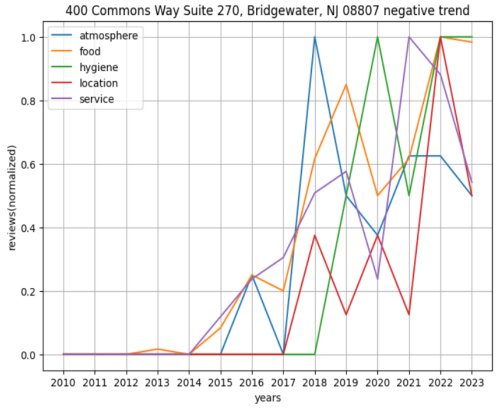










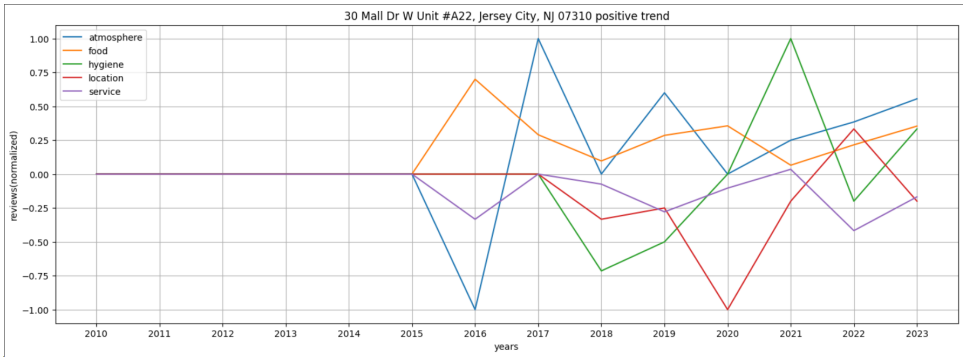


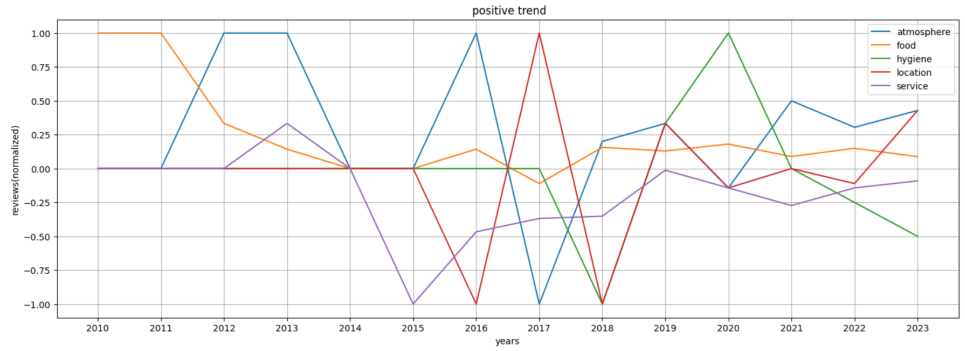
3. Improvement of Positive And Negative Aspects

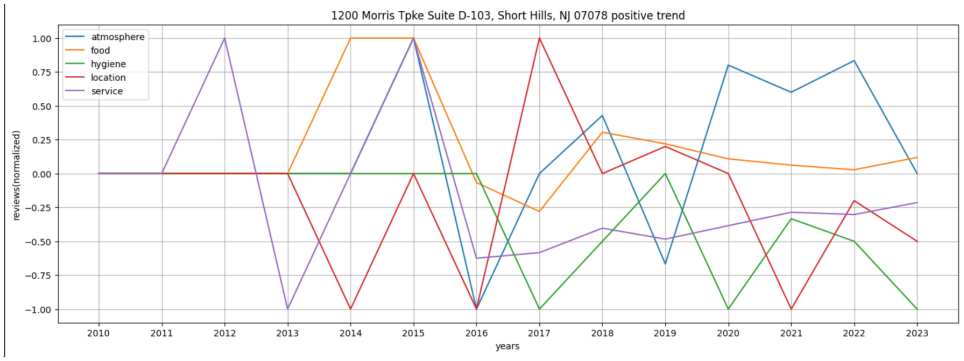
In our approach to more effectively represent customer sentiment trends, we devised a metric that calculates the net positivity rate as (positive reviews - negative reviews)/total reviews. This calculation allowed us to create line charts that more accurately depict the sentiment dynamics over time.

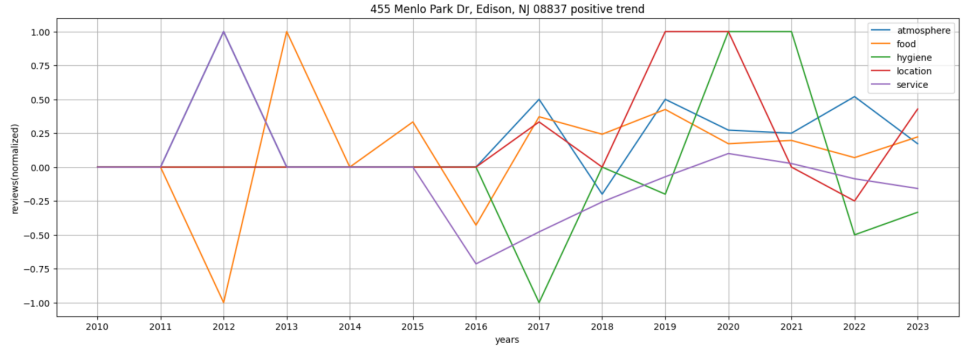
Taking a closer look at the first chart, specifically for the year 2023, we can visually discern that this restaurant received predominantly positive feedback in the areas of atmosphere, food, and hygiene. In contrast, the aspects of service and location were met with largely negative reviews. This contrasts with the 2022 data for the same establishment, where the atmosphere, location, and food categories garnered positive remarks, but the service and hygiene areas were viewed negatively.

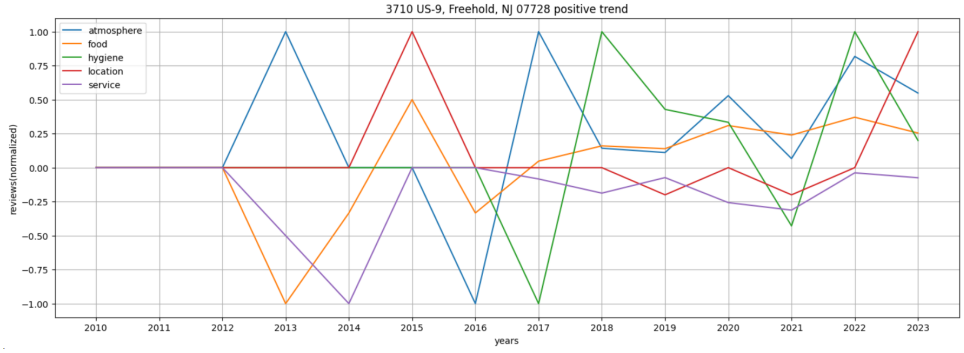
This method of representation enables us to present a more nuanced and accurate depiction of the customer review sentiment distribution for each restaurant, capturing the shifting preferences and experiences of customers over time.

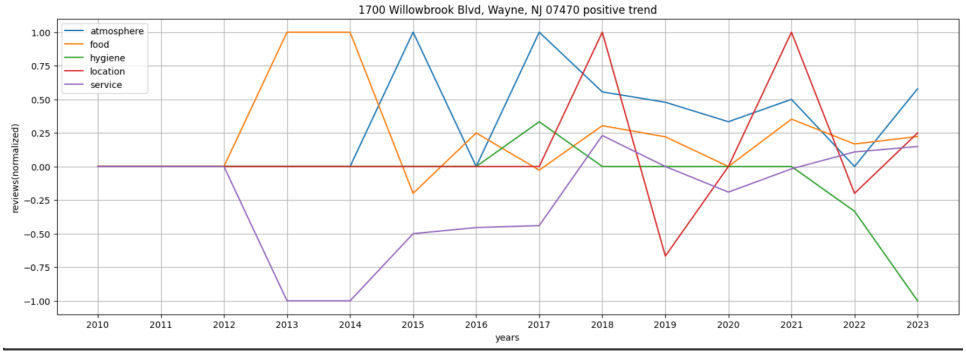


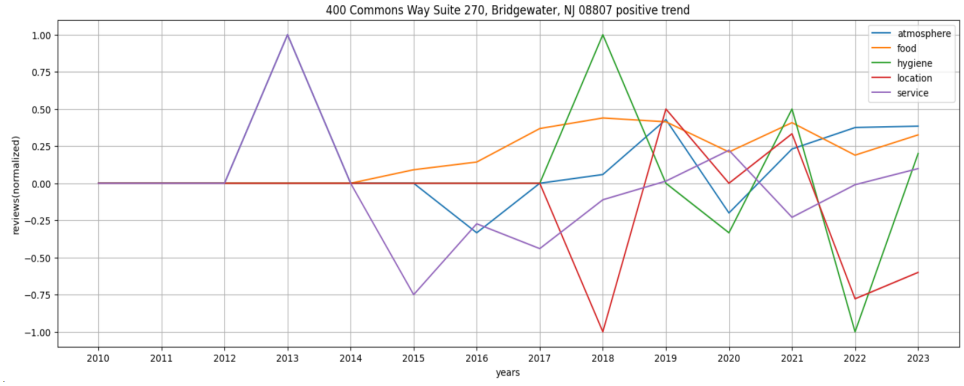


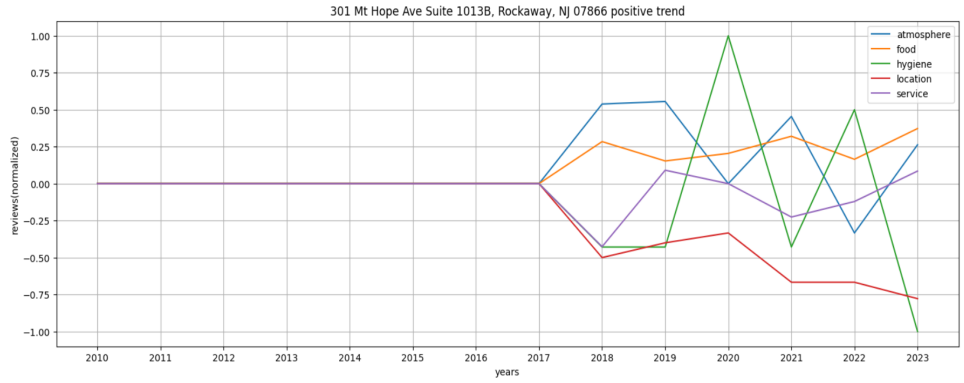

1200 Morris Tpke Suite D-103, Short Hills, NJ 07078 positive trend
atmosphere
food
hygiene
location
service
reviews(normalized)
years

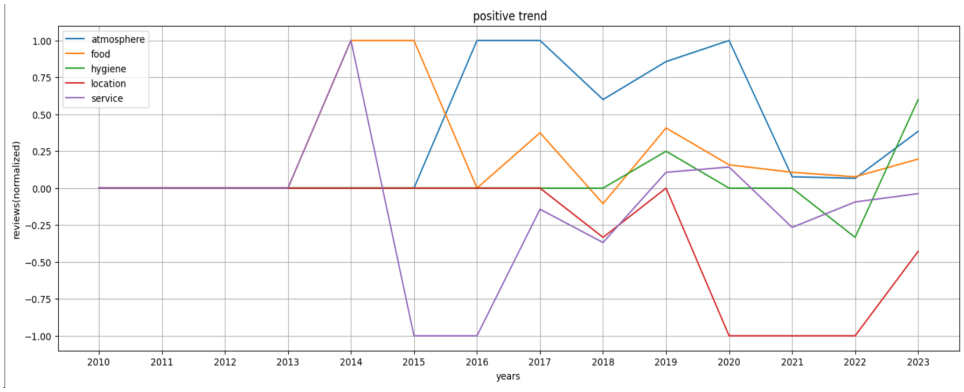

455 Menlo Park Dr, Edison, NJ 08837 positive trend
atmosphere
food
hygiene
location
service
reviews(normalized)
years

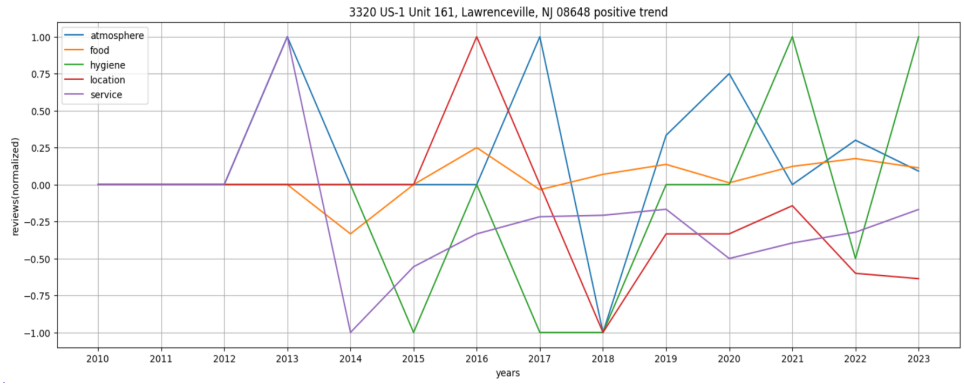

3710 US-9, Freehold, NJ 07728 positive trend
atmosphere
food
hygiene
location
service
reviews(normalized)
years


1700 Willowbrook Blvd, Wayne, NJ 07470 positive trend
atmosphere
food
hygiene
location
service
reviews(normalized)
years


400 Commons Way Suite 270, Bridgewater, NJ 08807 positive trend
atmosphere
food
hygiene
location
service
reviews(normalized)
years

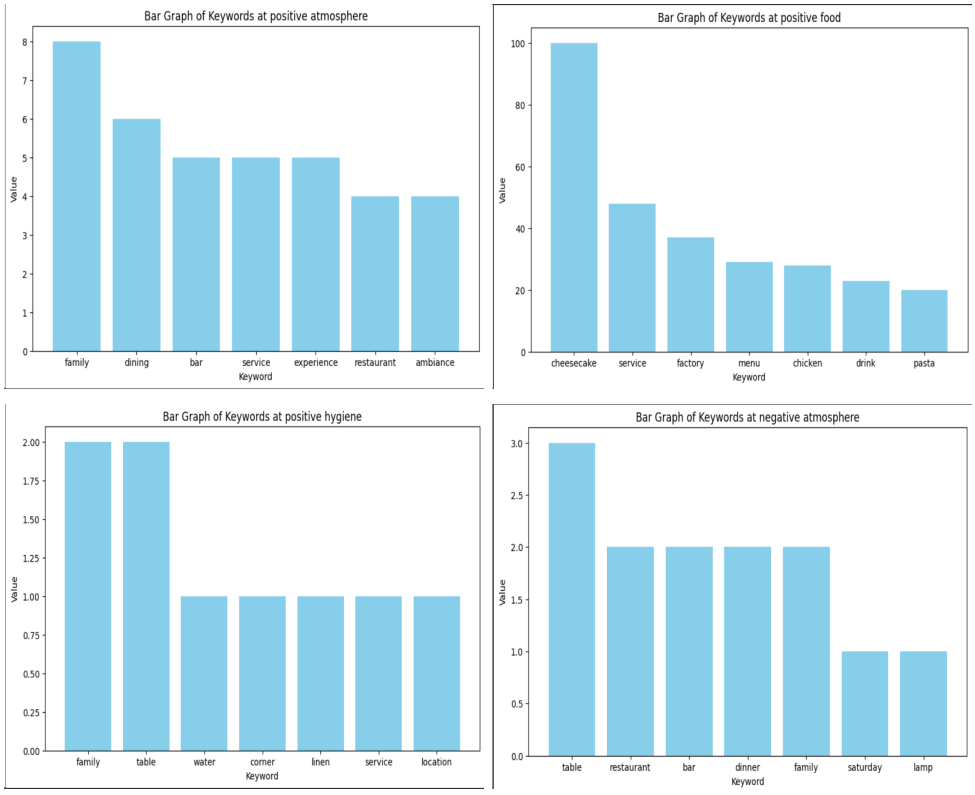

301 Mt Hope Ave Suite 1013B, Rockaway, NJ 07866 positive trend
atmosphere
food
hygiene
location
service
reviews(normalized)
years

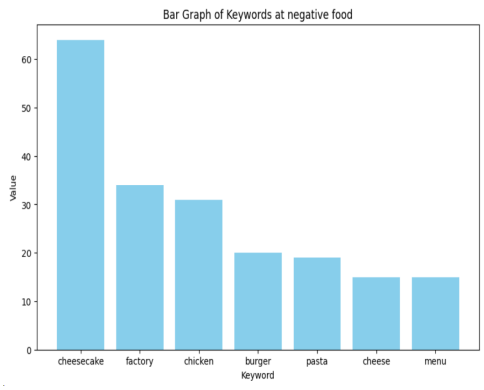

positive trend
atmosphere
food
hygiene
location
service
reviews(normalized)
years

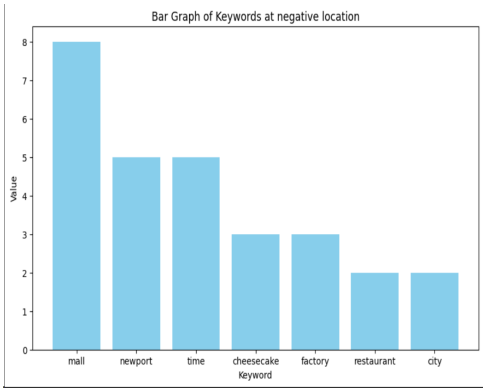

3320 US-1 Unit 161, Lawrenceville, NJ 08648 positive trend
atmosphere
food
hygiene
location
service
reviews(normalized)
years

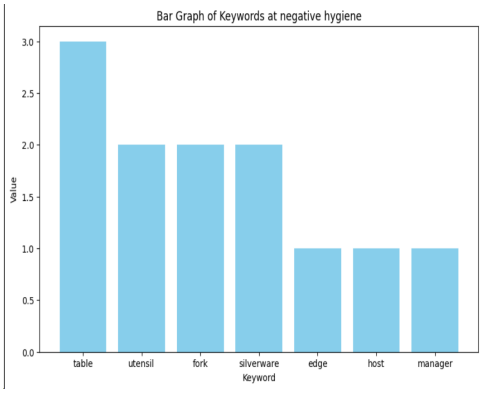
4. Keywords Results Of Positive And Negative Reviews In Each Type

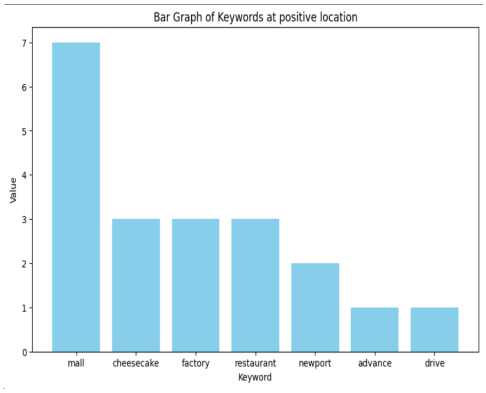
In the final stage of our analysis, we meticulously iterated through the dataset of each restaurant, focusing on identifying the most common keywords associated with both positive and negative sentiments within different categories. Our approach involved visualizing the top seven keywords for each sentiment type in every category, utilizing bar graphs for clear representation. By restricting our keyword selection to nouns, these graphs enable us to precisely pinpoint specific areas of concern or praise in various aspects of each restaurant's operations. This method ensures that the feedback provided to each restaurant is both clear and directly relevant to their service and offerings.

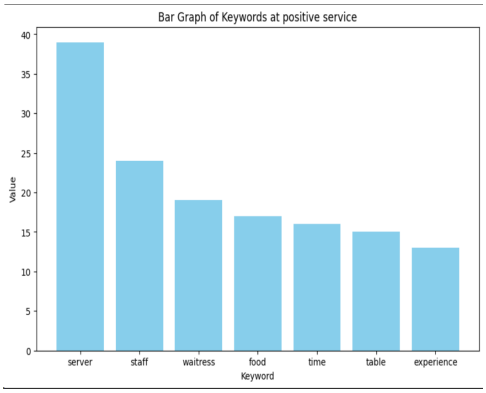


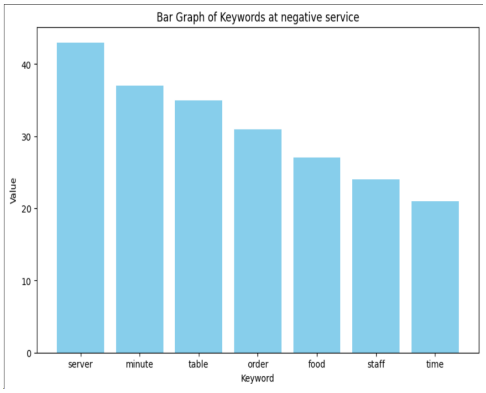

Bar Graph of Keywords at negative food
Value
Keyword
cheesecake
factory
chicken
burger
pasta
cheese
menu

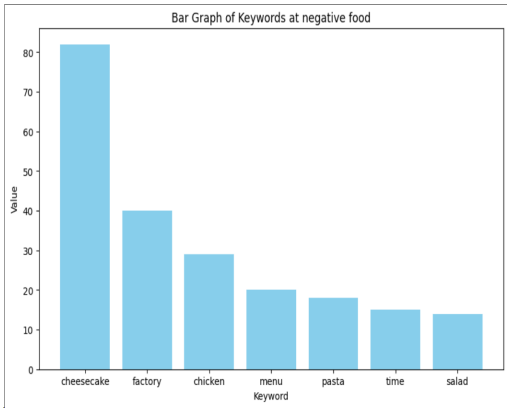

Bar Graph of Keywords at negative location
Value
Keyword
mall
newport
time
cheesecake
factory
restaurant
city

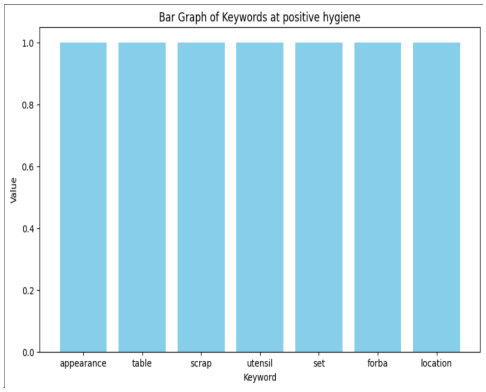

Bar Graph of Keywords at negative hygiene
Value
Keyword
table
utensil
fork
silverware
edge
host
manager

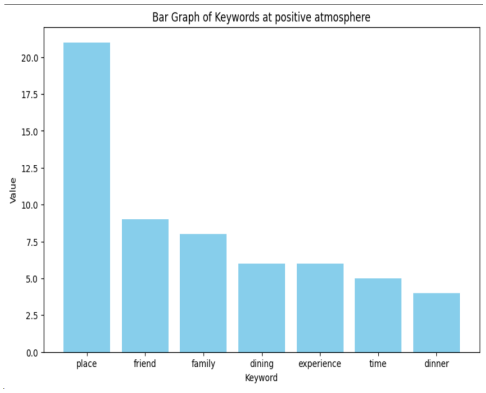

Bar Graph of Keywords at positive location
Value
Keyword
mall
cheesecake
factory
restaurant
newport
advance
drive

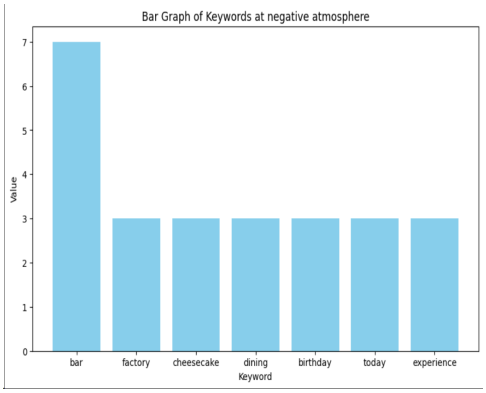

Bar Graph of Keywords at positive service
Value
Keyword
server
staff
waitress
food
time
table
experience


Bar Graph of Keywords at negative service
Value
Keyword
server
minute
table
order
food
staff
time

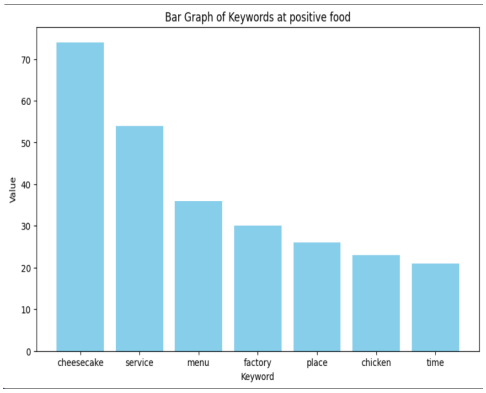

Bar Graph of Keywords at negative food
Value
Keyword
cheesecake
factory
chicken
menu
pasta
time
salad

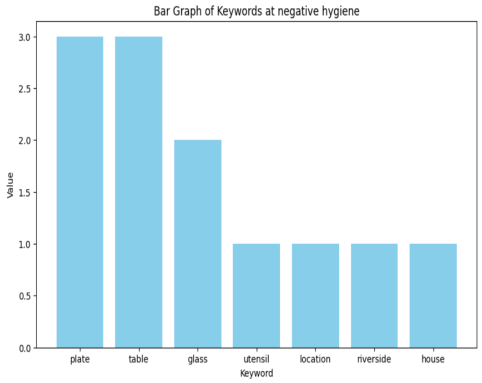

Bar Graph of Keywords at positive hygiene
Value
Keyword
appearance
table
scrap
utensil
set
forba
location

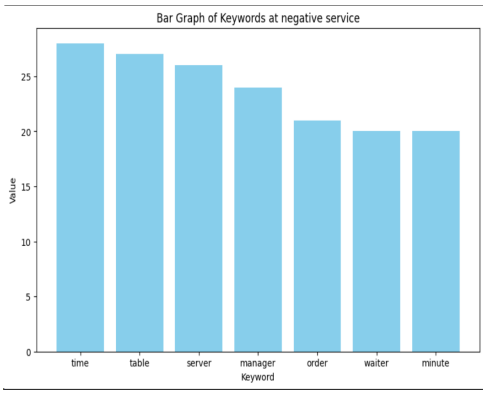

Bar Graph of Keywords at positive atmosphere
Value
Keyword
place
friend
family
dining
experience
time
dinner

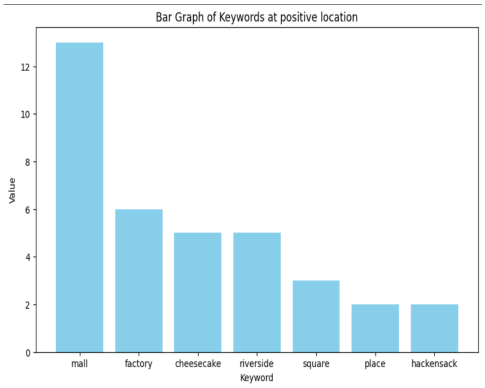

Bar Graph of Keywords at negative atmosphere
Value
Keyword
bar
factory
cheesecake
dining
birthday
today
experience

Z


Bar Graph of Keywords at positive food
Value
Keyword
cheesecake
service
menu
factory
place
chicken
time


Bar Graph of Keywords at negative hygiene
Value
Keyword
plate
table
glass
utensil
location
riverside
house


Bar Graph of Keywords at negative service
Value
Keyword
time
table
server
manager
order
waiter
minute


Bar Graph of Keywords at positive location
Value
Keyword
mall
factory
cheesecake
riverside
square
place
hackensack

Conclusion

Through our detailed analysis of customer reviews across ten restaurants, we have uncovered significant insights into customer perceptions and identified key areas for improvement. The analysis distinctly shows that, overall, customers rate 'food' and 'atmosphere' highly across these establishments, suggesting that these areas are strengths for the majority of the restaurants.

However, a consistent pattern of dissatisfaction emerges in the 'service' category across all restaurants. This indicates a critical area where improvement is necessary to enhance overall customer satisfaction. The feedback on 'location' and 'hygiene' shows more variability between different restaurants, suggesting that these aspects are influenced by individual restaurant circumstances and may require tailored approaches to improvement.

For example, at CheeseFactory, the positive atmosphere is frequently associated with keywords like 'family,' 'dining,' and 'experience,' reflecting a favorable environment. However, negative comments often pointed to service-related issues, like problems with 'tables' and 'service' itself, highlighting a need for operational improvements.

In the food category, the prominence of 'cheesecake' in both positive and negative contexts underscores its importance to the restaurant's identity. The presence of 'service' in positive reviews shows customer appreciation for attentive staff, but negative mentions in the food category, like 'chicken' and 'pasta,' may signal specific menu items needing attention. Hygiene-related comments, though fewer, underline the importance of cleanliness, with negative mentions of 'table' and 'utensils' suggesting areas that could significantly impact customer satisfaction. Location-related comments varied, indicating that each restaurant's setting influences customer perception differently.

To conclude, this analysis not only clarifies customer sentiment towards these restaurants but also provides actionable feedback for restaurant operators. Addressing service-related issues should be a priority, while maintaining the high standards in food and atmosphere. By focusing on areas with mixed reviews like location and hygiene, restaurants can further tailor their strategies to enhance overall customer experience and satisfaction.

Reference

[1]. Baumer, E. P., Katz, S., Freeman, J., Adams, P., & Gonzales, A. L. (2017). *How We Ask Matters: Understanding Survey Questions to Inform Design in HCI*. In Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (pp. 4754-4765).

[2]. Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). *Latent Dirichlet Allocation*. Journal of Machine Learning Research, 3, 993-1022.

[3]. Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). *SMOTE: Synthetic Minority Over-sampling Technique*. Journal of Artificial Intelligence Research, 16, 321-357.

[4]. Dellarocas, C., Zhang, X., & Awad, N. F. (2007). *Exploring the value of online product reviews in forecasting sales: The case of motion pictures*. Journal of Interactive Marketing, 21(4), 23-45.

[5]. Ghose, A., & Ipeirotis, P. G. (2011). *Estimating the helpfulness and economic impact of product reviews: Mining text and reviewer characteristics*. IEEE Transactions on Knowledge and Data Engineering, 23(10), 1498-1512.

[6]. Hu, M., & Liu, B. (2004). *Mining and summarizing customer reviews*. In Proceedings of the 10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 168-177).

[7]. Hutto, C. J., & Gilbert, E. (2014). *VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text*. In Proceedings of the Eighth International Conference on Weblogs and Social Media.

[8]. Liu, B. (2012). *Sentiment analysis and opinion mining*. Synthesis Lectures on Human Language Technologies, 5(1), 1-167.

[10]. O'Leary, D. E. (2016). *Customer satisfaction and dissatisfaction in China: A comprehensive review*. International Journal of Business and Social Science, 7(10), 97-114.

[11]. Pyun, J. Y., Kwak, H., & Smith, M. (2017). *Human computation and collaborative data processing: A research agenda*. Journal of Computer-Mediated Communication, 22(2), 61-76.

[12]. He, H., & Garcia, E. A. (2009). Learning from Imbalanced Data. IEEE Transactions on Knowledge and Data Engineering, 21(9), 1263-1284.

[13]. Haixiang, G., Yijing, L., Shang, J., Mingyun, G., Yuanyue, H., & Bing, G. (2017). *Learning from class-imbalanced data: Review of methods and applications.* Expert Systems with Applications, 73, 220-239.

[14]. Krawczyk, B. (2016). *Learning from imbalanced data: open challenges and future directions*. Progress in Artificial Intelligence, 5(4), 221-232.

[15]. Jeni, L. A., Cohn, J. F., & De La Torre, F. (2013). *Facing Imbalanced Data--Recommendations for the Use of Performance Metrics.* Humaine Association Conference on Affective Computing and Intelligent Interaction (ACII), 2013.

[16]. Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). *SMOTE: Synthetic Minority Over-sampling Technique*. Journal of Artificial Intelligence Research, 16, 321-357.

[17]. Daniel Ramage, David Hall, Ramesh Nallapati and Christopher D. Manning.(2009). *Labeled LDA: A supervised topic model for credit attribution in multi-labeled corpora.* Proceedings of the 2009 conference on empirical methods in natural language processing. 2009: 248-256.