Yelp-Sentiment-Rating-of-Restaurants

Amin Fesharaki†
Applied Data Science Master's Program
Shiley Marcos School of Engineering / University of San Diego
afesharaki@sandiego.edu

Luis Perez
Applied Data Science Master's Program
Shiley Marcos School of Engineering / University of San Diego
ldperez@sandiego.edu

ABSTRACT

*Article Title Footnote needs to be captured as **Title Note**
†Name Footnote to be captured as **Author Note**

KEYWORDS

Sentiment analysis; Online reviews; Opinion mining; Topic modeling; Recommender systems; Customer preferences**;**

1 Introduction

As the restaurant industry expands, customers face overwhelming dining options. Sentiment analysis and topic modeling have been successful in the restaurant industry (Boya et al, 2017). A potential application is a recommender system for restaurants in popular downtown areas. By leveraging sentiment analysis, which involves analyzing customer feedback to determine the overall tone and sentiment towards a particular restaurant or dish, a recommender system could provide personalized recommendations based on customer preferences and help them make more informed dining decisions (Salehi-Esfahani & Kang, 2019).

Another potential application for sentiment analysis on Yelp reviews is identifying areas for improvement in restaurant operations. By analyzing customer feedback, restaurant managers can gain insights into what aspects of their business are performing well and what areas need improvement. For example, if many customers mention slow service, restaurant managers can focus on improving wait times and staffing levels to enhance the overall customer experience. Identifying trends and patterns in customer feedback can also help restaurants better understand their target market and develop more effective marketing strategies (Meyer-Waarden & Sabadie, 2023). By utilizing sentiment analysis, restaurant managers can make data-driven decisions that improve the customer experience and ultimately lead to increased profitability. The following literature review examines research on sentiment analysis and topic modeling.

2 Background

Sentiment analysis and topic modeling have become increasingly popular research topics in the restaurant industry. The proliferation of social media and online review platforms has enabled customers to share their experiences and opinions about restaurants more efficiently than ever before.

Using sentiment analysis can help restaurant operators better understand customers' feedback and adapt accordingly (Huang et al, 2014). Additionally, topic modeling can provide insight into specific features of restaurants that customers appreciate or would like to see improved (Boya et al, 2017).

The potential application of sentiment analysis and topic modeling in developing a recommender system for restaurants highlights the relevance and importance of these research areas in the industry (Salehi-Esfahani & Kang, 2019). Research studies have shown that online reviews and ratings can have a significant impact on a restaurant's reputation and revenue (Luca, 2011). Therefore, leveraging sentiment analysis and topic modeling can provide valuable insights for restaurant operators to improve customer satisfaction and ultimately drive business success.

2.1 **Problem Identification and Motivation (Heading Level 2)**

The restaurant industry is highly competitive, and customers have access to a vast array of dining options. While online review platforms such as Yelp provide ratings and reviews of restaurants, they are often subjective and may not accurately reflect customers' overall sentiment. Moreover, traditional rating systems do not consider the nuances of customers' experiences or preferences. As a result, customers may need help making informed dining decisions, and restaurant operators may need help to improve their businesses based on customers' feedback

The restaurant industry can enhance customer satisfaction and drive business success by addressing these issues. Therefore, there is a need for a more detailed approach to evaluating restaurants' performance and providing personalized recommendations to customers. Comprehensively examining customers' feedback and preferences is essential for attaining our objectives and addressing the issue. Gaining insights into their needs, wants, and opinions allow us to make informed decisions and enhance customer satisfaction (Meyer-Waarden & Sabadie, 2023). A recommender system based on sentiment analysis and topic modeling can help customers make more informed dining decisions and assist restaurant operators in improving their businesses based on customers' preferences.

2.1.1 Personalized Restaurant Ratings: The restaurant industry is a highly dynamic and evolving field where customer satisfaction is of utmost importance. In this information age, customers can access a vast amount of information about restaurants and their offerings. However, with the rise of online review platforms, customers may need help differentiating between subjective and objective feedback. Yelp and other review platforms provide ratings and reviews, but these ratings may not accurately represent the customers' overall sentiment. Additionally, traditional rating systems do not consider the nuances of customer experiences or preferences. This can make customers feel uncertain about their choices, and restaurant operators may need help to improve their businesses based on customer feedback. Therefore, there is a need for a more personalized and nuanced approach to rating restaurants that considers customers' preferences and experiences. By leveraging sentiment analysis and topic modeling, this project aims to provide a more accurate and comprehensive rating system for restaurants in popular downtown areas. This will help customers make informed dining decisions and support restaurant operators in enhancing their businesses.

2.1.1.1 Optimizing Restaurant Operations: Additionally, sentiment analysis and topic modeling can provide valuable insights for restaurant operators to optimize their marketing strategies and enhance their brand reputation. By identifying a restaurant's most frequently mentioned positive and negative aspects, operators can leverage their strengths and address their weaknesses. For example, if a restaurant receives consistently positive feedback on their pizza but negative feedback on their service, they can focus on promoting their pizza offerings while working to improve their service quality. Additionally, by identifying trending topics and sentiments in customers' reviews, operators can stay ahead of market trends and adapt their menu and service offerings accordingly. Overall, using sentiment analysis and topic modeling in the restaurant industry can lead to better decision-making and improved business outcomes for customers and operators.

2.2 Definition of Objectives

This project aims to perform sentiment analysis on restaurants in popular downtown areas and provide users with a rating based on the sentiment score. The ultimate goal is to offer valuable information to users that can assist them in making informed dining decisions. By understanding the sentiments expressed in online reviews, users can improve their dining experience. The sentiment analysis and rating system are just the means to achieve the larger goal of enhancing the dining experience for users. The future scope of the project includes expanding to additional areas, but this study will focus on analyzing a limited number of downtown areas to showcase the methodology's effectiveness.

Furthermore, the project will use topic modeling to investigate specific restaurant features in more detail than is currently available on Yelp. By providing more personalized recommendations, we aim to create a better dining experience for customers and improve the overall reputation of restaurants in popular downtown areas. If our sentiment analysis and topic modeling methods are successful, the restaurant industry could benefit from increased revenue and a better reputation.

2.2.1 Benefits and Applications: Moreover, the success of this project can also lead to the development of new tools for restaurant owners to understand their customers better and improve their services. For instance, a dashboard that displays sentiment scores and topic trends could allow restaurant owners to quickly identify areas of improvement and track the impact of changes made to their menu or services. This can lead to improved customer satisfaction, higher ratings, and ultimately, increased revenue for the restaurant. Overall, this project has the potential to create a win-win situation for both customers and restaurant owners, ultimately contributing to the growth and success of the restaurant industry.

3 Literature Review (related works)

Yelp reviews hold plenty of weight in the restaurant industry. Reviews can lift expected revenue, highlight important improvement areas, and assist restaurant operators in understanding consumer input. Reviews also play an essential role in restaurant selection and can exhibit restaurant-client relationships in a public forum.

3.1 Reviews, reputation, and revenue: The case of yelp.com

The themes expressed in this article highlight the relationship between Yelp ratings, revenue, and reputation for independent restaurants. An overall take is that a one-star increase in Yelp rating leads to a 5-9% increase in revenue for independent restaurants. On the other hand, Chain restaurants' revenue is unaffected due to Yelp ratings. Online reviews are practical tools for small businesses to increase their revenue and reputation. A potential gap in the study is that it only focuses on independent restaurants. Medium or multiple franchisee restaurants were not part of the study (Luca, 2011).

3.2 Improving Restaurants by Extracting Subtopics from Yelp Reviews

Restaurant operators can make focused improvements that enhance client happiness and eventually boost income by identifying critical areas for improvement. This journal investigates subtopic extraction and identifies crucial improvement areas for restaurants. Based on client reviews, the subtopic extraction method may efficiently pinpoint important areas where restaurants can improve. The study does not evaluate potential biases in Yelp reviews or how they might affect the subtopic extraction (Huang et al, 2014).

3.3 Identifying Restaurant Features via Sentiment Analysis on Yelp Reviews.

The article analyzed sentiment analysis to identify restaurant features. Utilizing a dataset of Yelp reviews, the paper explained how they used sentiment analysis and identifying restaurant features to create a model. The model was successful in detecting restaurant features. The authors suggested a strategy that could assist restaurant operators in better-comprehending consumer input and adapting accordingly (Boya et al, 2017).

3.4 Why do you use yelp? analysis of factors influencing customers' website adoption and dining behavior.

The common themes found in this journal are factors that influence customers' adoption of Yelp, customers' dining behavior, the importance of online reviews in restaurant selection, the role of Yelp in customer decision-making, and the relationship between customer satisfaction and Yelp usage. The ability to identify high-quality restaurants and the desire for restaurant information were the two primary motivators for customers to use Yelp. Positive reviews had a more substantial influence on customer decision-making than negative reviews. Customers who were highly satisfied with their dining experiences were more likely to use Yelp to write reviews, whereas those who were less satisfied were less likely to use the platform. Additionally, the influence of online reviews on customer decision-making varied depending on the type of restaurant (Salehi-Esfahani & Kang, 2019).

3.5 Relationship quality matters: How restaurant businesses can optimize complaint management.

Meyer-Waarden and Sabadie examine how restaurant organizations might enhance complaint management to raise client satisfaction and retention rates. The report underlines that a crucial aspect of complaint management is the caliber of the company-client relationship. According to the authors, companies can handle complaints well and use them as an opportunity to improve their connections with customers by taking a customer-centric strategy. The article gives various suggestions for handling complaints, such as active listening, timely responses, empathy, and customized solutions. The journal emphasizes the value of complaint management in boosting customer satisfaction and achieving business success overall (Meyer-Waarden & Sabadie, 2023).

4 Methodology

The methodology section describes a Python script developed to perform sentiment analysis on customer reviews of a restaurant. The script categorizes the reviews based on different aspects such as food, service, atmosphere, price, alternative diet, and location.. To accomplish this task, several libraries were employed, including NLTK and TextBlob.

In the first step of the methodology, the necessary libraries were loaded and aspect keywords were defined for each aspect. Next, the customer reviews data was loaded into a pandas dataframe. The reviews were then split into phrases based on conjunctions, punctuation, and other rules. Following this, the phrases were filtered based on the aspect keywords using a filtering function. The filtered phrases were used to create new columns in the dataframe for each aspect. Finally, the sentiment score was calculated for each aspect using a sentiment analysis function.

In the last step of the methodology, the data frame was grouped by location and the average sentiment score for each aspect was calculated. Only the location and aspect sentiment columns were selected, and the result was printed. This methodology provides a unique approach to sentiment analysis that allows for a more nuanced understanding of customer sentiment towards different aspects of a restaurant.

4.1 **Data Acquisition and Aggregation**

Our team acquired data from the Yelp open dataset and Yelp Fusion API. The open Yelp database is a subset of businesses, reviews, and user data for personal, educational, and academic purposes. Yelp's open dataset is a set of six compressed JSON files. Within these files are 11 metropolitan areas, 150,346 businesses, 6,990,280 reviews, 200,100 pictures, 908,915 tips by 1,987,897 users, and Over 1.2 million business attributes like hours, parking, availability, and ambiance.

In contrast, Yelp's Fushion API provides a way to pull the most accurate and current local data source available with some limitations. The pull data is secondary in importance because of daily rate limits. Additionally, Yelp limits reviews to 3 per business and truncates reviews to 160 character limit. Yet, there is still value in having updated data.

Yelp's Open data JSON files consisted of around 9GB in total. We focused on two of the six files--- business and reviews--- to make the data more manageable. We continued to filter results to a specific metropolitan area and combined the business and review files based on business id. The team will add the pulled data from the API in segments to work with the daily limits set by Yelp simultaneously.

4.1.1 Exploratory Data Analysis

The Exploratory Data Analysis (EDA) section focused on the data preprocessing steps necessary to prepare the text data for further analysis. We began by filtering the dataset by location (Table 1). We focused on the Santa Barbara metropolitan area: Santa Barbara, Isla Vista, Goleta, Carpinteria, Montecito, Port Hueneme, Summerland, Santa Clara, Truckee, and Santa Maria.

Table 1

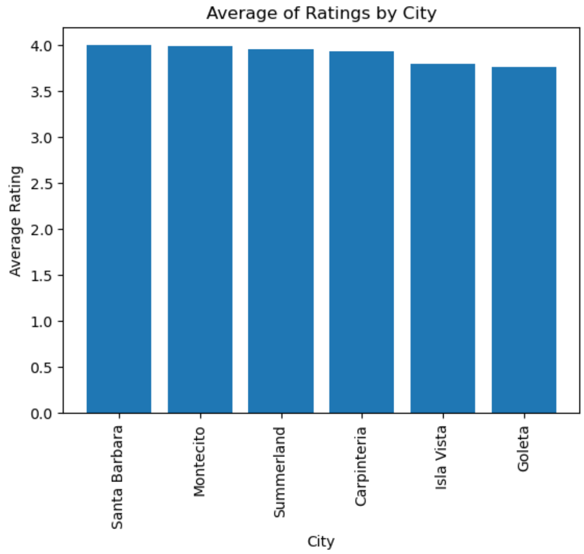
Restaurants In Santa Barbara metropolitan area

City	Number of Restaurants
Santa Barbara	270002
Goleta	45689
Carpinteria	17391
Isla Vista	7583
Montecito	5042
Summerland	2418
Mission Canyon	19
Santa Ynez	17
Oxnard	13
Total: 239065	

Figure 1, describes the average ratings of all the cities in the Santa Barbara metropolitan area. The barplot displayed the average ratings for each city on the y-axis, while the x-axis represented the different cities in the region. The height of each bar represented the average rating for each city. By visualizing a barplot, we could quickly compare the average ratings of each city in the region and identify any cities with high or low ratings. We found that the average ratings were relatively consistent across all the cities in the region, with only slight variations between them. This could suggest that the quality of the restaurants, as perceived by the customers who left ratings, is consistent across the different cities in the region.

Figure 1

Average of Ratings by City



4.1.1.2 NLP Pipeline

We began the NLP pipeline by removing any stopwords and punctuation from the text. Then tokenized the text and defined a lemmatize function, which combined all words with the same stem into a single term. Finally, we created a pipeline to execute all of these steps systematically. This data preprocessing step was necessary to ensure that further analysis would be conducted on clean and uniform data. By eliminating stopwords and punctuation and lemmatizing any words with the same stem, we confirmed that the data was free of any unnecessary noise that could interfere with further analysis.

Furthermore, by utilizing a pipeline, we were able to ensure that each of these steps was executed in a consistent and reliable manner. Overall, this EDA section was essential in ensuring the text data was ready for further analysis. By preprocessing the data this way, we confirmed that any further research would be conducted on clean and uniform data.

4.2 **Data Quality**

Once the data was filter and tokenized we could visualize the average comment length by restaurant. **Figure 2**, shows the top 25 restaurants by comment length. This gave an inside into how in-depth reviews for restaurants with high ratings. The opposite was true for the bottom 25 restaurants. In **figure 3**, we see how restaurants with low ratings have lower review comment lengths.

Figure 2

Top 25 Average Comment Length by Restaurant.

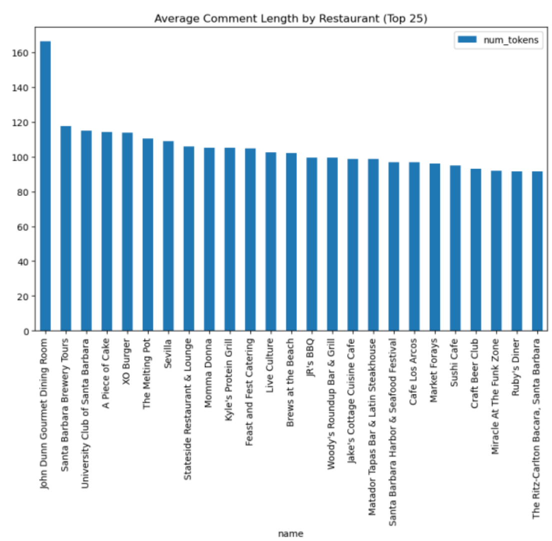
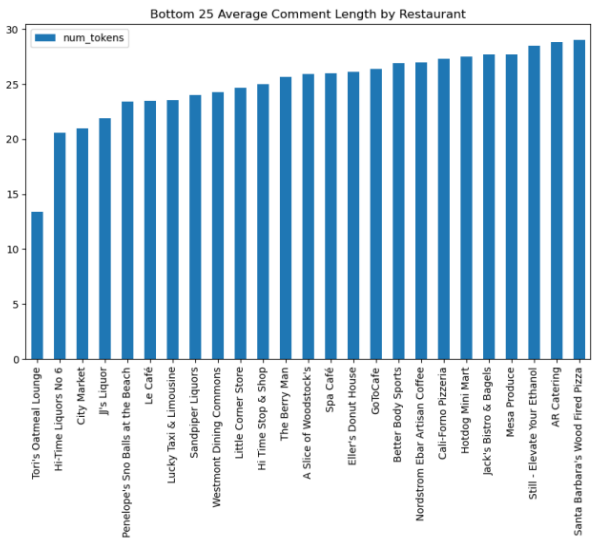


Figure 3

Bottom 25 Average Comment Lenght by Restaurant.

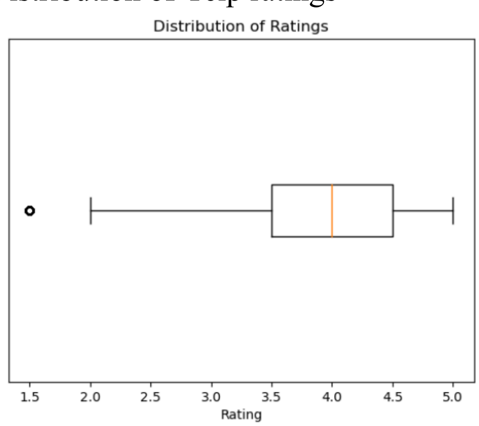


4.2.1 Distribution of Reviews

Understanding the overall distribution of Yelp reviews is essential for performing topic modeling and text sentiment analysis, as it can provide insight into the general sentiment of users towards different businesses and services. **Figure 4** describes the rating distribution found in Yelp reviews. A key takeaway is that 50% of Yelp reviews have a rating of 4 have a rating of 4 or higher, and the upper whisker shows that most yelp reviews fall on the positive side.

Figure 4

Distribution of Yelp ratings



4.2.2 Rating Groups Themes

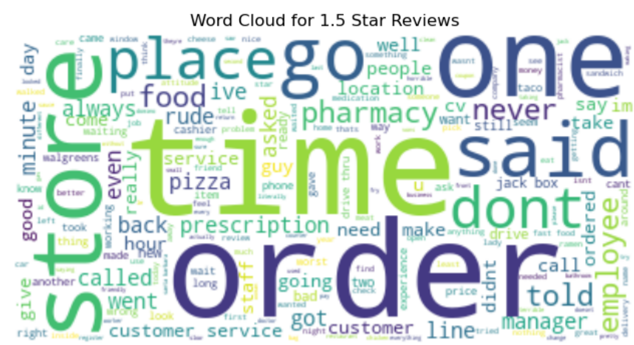
With reviews tokenized, we could look into key themes in different rating groups. In figure 5, we mapped the word cloud for restaurants with five-star ratings to look for language patterns, emphasis, and tone. In comparison, Figure 6 maps the lower rating to visualize the contrast in key words within the top and lower ratings.

Figure 5

Word Cloud for 5.0 Star Reviews

Figure 6

Word Cloud for 1.5 Star Reviews



4.3 Feature Engineering

The approach to feature engineering for sentiment analysis is based on leveraging both linguistic and semantic features of the text data. By splitting the text into phrases based on grammatical and conjunctional relationships, important aspects or topics of the text data can be identified that are relevant to the sentiment analysis task. After identifying these aspects, topic modeling techniques can be used to create new variables based on prevalent topics in the text data.

Creating new variables based on the topics allows the sentiment analysis to be focused on specific aspects of the text data that are most relevant to the domain. For example, in the case of restaurant reviews, new variables can be created based on topics such as food, service, atmosphere, price, alternative diet, and location. Analyzing the sentiment of phrases related to these topics separately provides a more granular understanding of the sentiment towards different aspects of the restaurant, which can be more informative than a single overall sentiment score.

The combined use of Top2vec and LDA facilitated the topic modeling process. LDA provided a preliminary outline of the topics and their respective associated words. Top2vec was utilized to further investigate word relationships, thereby establishing a foundational understanding of the words that belong to each topic. Consequently, a comprehensive list of words was generated for each of the following topics: location, food sentiment, service sentiment, atmosphere sentiment, price sentiment, and alternative diet sentiment. The significance of this approach lies in the algorithm's ability to accurately categorize phrases based on their inclusion of words related to specific topics.

In summary, the approach to feature engineering for sentiment analysis demonstrates the importance of considering domain-specific aspects of the text data and how they relate to the sentiment analysis task. By creating new variables based on relevant topics, more informative and actionable insights can be obtained from the sentiment analysis. This approach provides a powerful tool for understanding the sentiment towards specific aspects of the text data, which can be particularly useful in domains such as restaurant reviews where customers may have different opinions about different aspects of the restaurant.

4.4 **Modeling**

4.4.1 Selection of modeling techniques.

Two Python packages that were used for sentiment analysis are VADER (Valence Aware Dictionary and sEntiment Reasoner) and TextBlob. While both are capable of analyzing sentiment in text data, they differ in several key ways. VADER is a lexicon-based approach, using pre-built dictionaries of words and phrases with assigned sentiment scores, whereas TextBlob is a machine learning-based approach, having been trained on labeled data to recognize language patterns that correspond to positive or negative sentiment.

VADER was specifically designed to analyze sentiment in social media text, making it particularly useful for analyzing informal language, sarcasm, and other nuances often found in social media (Calderon, 2018). It provides not only an overall sentiment score, but also scores for positive, negative, and neutral sentiment, as well as a "compound" score that combines all three. In contrast, TextBlob provides only an overall sentiment score and is designed to work with general text.

When it comes to sentiment analysis in the context of restaurant reviews, VADER may be the more appropriate choice due to its focus on social media and fine-grained sentiment analysis capabilities. Reviews often contain informal language, sarcasm, and other nuances, and the ability to differentiate between positive, negative, and neutral sentiment can provide more insightful results. However, the choice between VADER and TextBlob ultimately depends on the specific needs and characteristics of the text data being analyzed (GrabNGoInfo, A, 2023).

5 Results and Findings

Overall the sentiment analysis of Yelp restaurant reviews using VADER and TextBlob with predefined topics yielded similar accuracy, F-1, and Recall. The results showed that VADER had higher marks across the board, but the difference was minimal. VADER achieved an overall difference of 8% across the board. The analysis also revealed that most restaurant reviews were positive in sentiment, with a mean sentiment score of 0.68 for VADER and 0.25 for TextBlob. However, the reviews also had a significant amount of negative sentiment, with a mean sentiment score of 0.12 for VADER and 0.10 for TextBlob. The difference is minimal, but we choose Vader because it fits our research more closely.

5.1 Evaluation of Results (Heading Level 2)

We created six restaurant recommendation categories combining topic analysis and domain knowledge to develop a more comprehensive understanding of the sentiment expressed in the Yelp reviews. These sections included location, food sentiment, service sentiment, atmosphere sentiment, price sentiment, and alternative diet sentiment. Figure 7 visually represents the sentiment expressed in each of the six restaurant categories in the dataset, resulting in a score ranging from 0-100 for each category. The scores were then aggregated and visualized in Figure 1, which provides a quick and easy way for potential customers to gain insight into the sentiment expressed in Yelp reviews for each restaurant. Potential customers can use these results to quickly gain insight into the sentiment expressed in Yelp reviews for each restaurant without having to read hundreds of reviews.

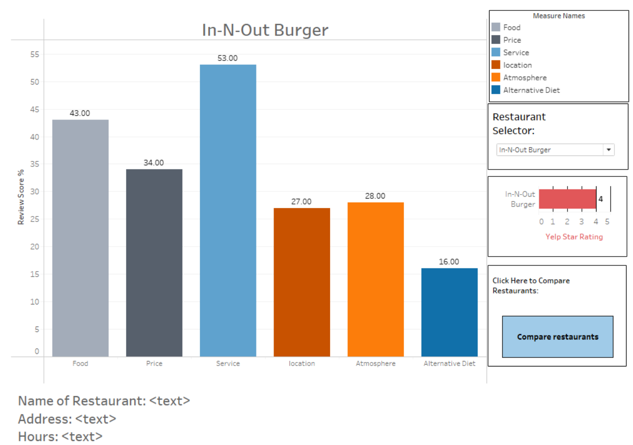
Figure 7

Sentiment Scores Per Restaurant

	location	food_sentiment	service_sentiment	atmosphere_sentiment	price_sentiment	alternative_diet_sentiment	locationCategory_sentiment
0	101 Deli	65.0	57.0	25.0	49.0	0.0	23.0
1	1114 Sports Bar & Games	51.0	64.0	49.0	36.0	16.0	32.0
2	212° Hotpot	40.0	57.0	44.0	38.0	15.0	24.0
3	4 Eggs & Pizza	62.0	46.0	37.0	26.0	17.0	22.0
4	6831 AA Hollister Ave	28.0	48.0	31.0	17.0	0.0	28.0
...	...	...	...	...	...	...	...
1388	duo events	58.0	69.0	27.0	0.0	0.0	0.0
1389	goa taco	50.0	51.0	29.0	35.0	20.0	23.0
1390	iGrill Korean BBQ	37.0	54.0	32.0	39.0	15.0	18.0
1391	il Fustino	30.0	63.0	45.0	61.0	16.0	26.0
1392	il fustino	29.0	56.0	42.0	0.0	0.0	44.0

Figure 8

Tableau Dashboard



6 Discussion

The current study indicates that sentiment analysis and topic modeling can be effective tools in developing a recommender system for restaurants. By analyzing Yelp reviews, specific features that customers appreciate or would like to see improved were identified. This information can be utilized by restaurant operators to enhance customer satisfaction and ultimately contribute to business success. The study's implications are significant for the restaurant industry, as leveraging sentiment analysis and topic modeling can provide valuable insights into customer preferences and enable data-driven decisions for business improvement. Moreover, the use of a recommender system can provide customers with more personalized recommendations, leading to a better dining experience overall.

To accomplish this task, we developed a Python script that categorizes reviews based on different aspects such as food, service, atmosphere, price, alternative diet, and location. We employed several libraries including NLTK and Vader to perform sentiment analysis on customer reviews. We also used topic modeling to investigate specific restaurant features in more detail than is currently available on Yelp.

The sentiment analysis of Yelp restaurant reviews using VADER and TextBlob with predefined topics yielded similar accuracy, F-1, and recall. VADER achieved an overall difference of 8% across the board, with a mean sentiment score of 0.68, compared to TextBlob's mean sentiment score of 0.25. However, the reviews also had a significant amount of negative sentiment, with a mean sentiment score of 0.12 for VADER and 0.10 for TextBlob. The difference is minimal, but VADER was chosen because it fits the research more closely.

In the evaluation of results, six restaurant recommendation categories were created, combining topic analysis and domain knowledge to develop a more comprehensive understanding of the sentiment expressed in the Yelp reviews. These sections included location, food sentiment, service sentiment, atmosphere sentiment, price sentiment, and alternative diet sentiment. Figure 7 visually represents the sentiment expressed in each of the six restaurant categories in the dataset, resulting in a score ranging from 0-100 for each category. The scores were then aggregated and visualized in Figure 1, which provides a quick and easy way for potential customers to gain insight into the sentiment expressed in Yelp reviews for each restaurant. This approach provides a powerful tool for understanding the sentiment towards specific aspects of the text data, which can be particularly useful in domains such as restaurant reviews where customers may have different opinions about different aspects of the restaurant.

Our study acknowledges several limitations, including a limited sample size from the Santa Barbara metropolitan area that may not represent other regions or demographics. Sentiment analysis is constrained by its reliance on text data, failing to consider the user's background and the social and cultural context of the reviews, which could lead to inaccurate sentiment interpretations. Additionally, subjectivity, small datasets, domain-specific language, unbalanced classes, and bias can all impact the accuracy of sentiment analysis.

It is crucial to consider these limitations when employing sentiment analysis on Yelp data, and to utilize supplementary methods and expert knowledge for validating and interpreting the results. Factors such as differing interpretations of reviews, the difficulty in detecting sarcasm or irony, and the influence of biases in data (e.g., reviewer demographics, review time period, and restaurant location) all play a role in the potential inaccuracy or bias of sentiment analysis outcomes.

In conclusion, our data science project, which aimed to analyze customer reviews of restaurants in downtown areas using sentiment analysis and topic modeling techniques, demonstrated that leveraging advanced techniques can yield valuable insights into customer reviews. These insights can be used to provide personalized recommendations to customers and improve the overall reputation of restaurants in popular areas. However, further research is needed to explore other factors that may influence customer feedback and to validate our findings in different contexts.

6.0.1 Study Comparison

When comparing studies that have utilized topic modeling, such as "Improving Restaurants by Extracting Subtopics from Yelp Reviews" by Haung (2014), it was observed that similar words based on the topics such as service, location, value, and food were identified. Although LDA is a traditional approach that can be useful in cases where the topics are unknown, we have improved on this method by utilizing Top2Vec, which is a more recent technique that incorporates clustering and neural networks to identify topics based on the semantic meaning of words. Unlike LDA, Top2Vec does not necessitate a pre-specified number of topics and can handle larger datasets more efficiently. Furthermore, Top2Vec can identify topics that are not explicitly represented in the text but are rather implied by context, making it particularly suitable for a dataset with a specific theme in mind, as in the case of restaurant reviews. Additionally, Top2Vec can work well when the topics are already known, which is the case with our study. To summarize, LDA is a well-established, interpretable, and traditional approach to topic modeling, while Top2Vec is a more flexible and contemporary approach that can handle larger datasets and identify implicit topics with greater ease.

Furthermore, the research conducted by New York University on Identifying Restaurant Features via Sentiment Analysis on Yelp Reviews utilized Support Vector Machines (SVM) to classify reviews as positive or negative, and also identified words commonly associated with restaurants such as "friendly" in positive reviews. Our study, however, took a different approach, utilizing Top2Vec to identify and categorize topics and themes present in the dataset, and subsequently computing sentiment scores using Vader. This methodology enabled us to provide consistent ratings across all types of restaurants, building upon their study by utilizing a different approach for sentiment analysis. Our final results were presented in a dashboard, enabling easy comparisons between restaurants. While their study focused on identifying the specific words associated with restaurants, our approach is more useful when comparing two restaurants.

6.1 **Conclusion**

(Discuss your conclusions in the order of the most important to the least important based on your technical findings of your data science project.)

6.2 **Recommend Next Steps/Future Studies**

Based on our findings, we recommend that future studies in the restaurant industry continue to leverage sentiment analysis and topic modeling techniques to gain valuable insights into customer reviews and preferences. We suggest exploring the use of other machine learning algorithms besides Textblob and Vader, such as boosting, random forest, neural networks, Naive Bayes,, BERT, LSTM, and SVM to possibly improve the accuracy of sentiment analysis and topic modeling. This can help restaurant operators better understand customer feedback and adapt accordingly.

Furthermore, there are measures that can be implemented to enhance the quality of the data. Fine-tuning and refining the topic modeling process would enable superior topic categorization, resulting in more precise outcomes during the sentiment analysis. Moreover, to establish a clear understanding of what constitutes a favorable or unfavorable rating, it is essential to examine the final results in combination with the restaurants' comments. This necessitates reviewing the comments related to each topic and leveraging expert knowledge to identify the range of scores that qualify as either positive or negative ratings. Another viable option is to categorize the scores based on letter grades, which would facilitate a more straightforward determination of whether a score is good or bad. Nonetheless, our current findings provide users with a comprehensive comparison of how various restaurants fare against one another. However, it may not be the most effective method for evaluating whether a particular restaurant is good or not solely based on the provided scores.

Additionally, we recommend that future studies focus on developing more personalized recommender systems for restaurants. Our study demonstrated that by leveraging advanced techniques such as sentiment analysis and topic modeling, personalized recommendations can be provided to customers based on their preferences. This can lead to increased customer satisfaction and ultimately drive business success.

Furthermore, we suggest exploring the use of social media platforms beyond Yelp in collecting customer feedback. By analyzing data from multiple sources, restaurant operators can gain a more comprehensive understanding of customer preferences.

In conclusion, our study highlights the importance of leveraging data-driven techniques such as sentiment analysis and topic modeling in the restaurant industry. By continuing to explore these techniques and developing more personalized recommender systems, restaurant operators can improve customer satisfaction and drive business success.

ACKNOWLEDGMENTS

References

References should be alphabetically ordered using APA 7.

Boya, Y., Jiaxu , Z., Yi, Z., & Yunong, C. (2017). Identifying Restaurant Features via Sentiment Analysis on Yelp Reviews. *ArXiv*. https://doi.org/10.48550/ARXIV.1709.08698

Calderon, P. (2018, March 31). *Vader sentiment analysis explained*. Medium. Retrieved April 3, 2023, from https://medium.com/@piocalderon/vader-sentiment-analysis-explained-f1c4f9101cd9

Huang, J., Rogers, S., & Joo, E. (2014). Improving Restaurants by Extracting Subtopics from Yelp Reviews. *ISchools*. Retrieved March 10, 2023, from https://www.ideals.illinois.edu/items/48852.

GrabNGoInfo, A. (2023, March 17). *Textblob vs. vader for sentiment analysis using Python*. Medium. Retrieved April 3, 2023, from https://pub.towardsai.net/textblob-vs-vader-for-sentiment-analysis-using-python-76883d40f9ae

Luca, M. (2011). Reviews, reputation, and revenue: The case of yelp.com. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.1928601

Meyer-Waarden, L., & Sabadie, W. (2023). Relationship quality matters: How restaurant businesses can optimize complaint management. *Tourism Management*, *96*, 104709. https://doi.org/10.1016/j.tourman.2022.104709

Salehi-Esfahani, S., & Kang, J. (2019). Why do you use yelp? analysis of factors influencing customers' website adoption and dining behavior. *International Journal of Hospitality Management*, *78*, 179–188. https://doi.org/10.1016/j.ijhm.2018.12.002

International Journal of Hospitality Management, *78*, 179–188. https://doi.org/10.1016/j.ijhm.2018.12.002

Meyer-Waarden, L., & Sabadie, W. (2023). Relationship quality matters: How restaurant businesses can optimize complaint management. *Tourism Management*, *96*, 104709. https://doi.org/10.1016/j.tourman.2022.104709

List order by date for keywords*

Luca, M. (2011). Reviews, reputation, and revenue: The case of yelp.com. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.1928601

Huang, J., Rogers, S., & Joo, E. (2014). Improving Restaurants by Extracting Subtopics from Yelp Reviews. *ISchools*. Retrieved March 10, 2023, from https://www.ideals.illinois.edu/items/48852.

Boya, Y., Jiaxu , Z., Yi, Z., & Yunong, C. (2017). Identifying Restaurant Features via Sentiment Analysis on Yelp Reviews. *ArXiv*. https://doi.org/10.48550/ARXIV.1709.08698

Salehi-Esfahani, S., & Kang, J. (2019). Why do you use yelp? analysis of factors influencing customers' website adoption and dining behavior.