Decoding Customer Sentiments: A Hierarchical Clustering Analysis of Pizza Hut Customer Reviews

[Abstract 3](#_Toc152663138)

[1. Research Rationale and Motivation 4](#_Toc152663139)

[1.1 Introduction 4](#_Toc152663140)

[1.2 Business Aim 4](#_Toc152663141)

[1.3 Scope of Research 4](#_Toc152663142)

[2. Literature Review 4](#_Toc152663143)

[3. Research Tools and Methods 5](#_Toc152663144)

[3.1 Tool Selection 5](#_Toc152663145)

[3.2 Web Scraping Process 6](#_Toc152663146)

[3.3 API Integration 6](#_Toc152663147)

[3.5 Dataset 6](#_Toc152663148)

[3.4 Data Preprocessing 6](#_Toc152663149)

[4. Results and Visualizations 17](#_Toc152663150)

[4.1 Reviews Count and Sentiment Analysis 17](#_Toc152663151)

[4.2 Keywords Analysis 19](#_Toc152663152)

[4.3 Common Words Analysis through Clustering 21](#_Toc152663153)

[5. Conclusion 23](#_Toc152663154)

[6. Limitations 23](#_Toc152663155)

[7. Recommendations 23](#_Toc152663156)

[5. Appendices 25](#_Toc152663157)

# Table of Figures

[Figure 1 Raw Dataset 9](#_Toc152664065)

[Figure 2 Code Segment 1 Sentiment Analysis and Keyword Analysis 10](#_Toc152664066)

[Figure 3 Result 1 Sentiment Analysis and Keyword Analysis 11](#_Toc152664067)

[Figure 4 Code Segment 2 Sentiment Analysis and Keyword Analysis 11](#_Toc152664068)

[Figure 5 Code Segment 3 Sentiment Analysis and Keyword Analysis 11](#_Toc152664069)

[Figure 6 Result 2 Sentiment Analysis and Keyword Analysis 12](#_Toc152664070)

[Figure 7 Code Segment 1 Month-Over-Month Positive and Negative Review Count Change 13](#_Toc152664071)

[Figure 8 Result 1 Month-Over-Month Positive and Negative Review Count Change 14](#_Toc152664072)

[Figure 9 Code Segment 1 Common Words Identification 15](#_Toc152664073)

[Figure 10 Result 1 Common Words Identification 16](#_Toc152664074)

[Figure 11 Result 2 Common Words Identification 16](#_Toc152664075)

[Figure 12 Code Segment 1 Clustering the Common Words with Word Embeddings and Hierarchical Clustering 17](#_Toc152664076)

[Figure 13 Result 1 Clustering the Common Words with Word Embeddings and Hierarchical Clustering 18](#_Toc152664077)

[Figure 14 Result 1 Cluster Analysis 18](#_Toc152664078)

[Figure 15 Code Segment 1 Cluster Analysis 20](#_Toc152664079)

[Figure 16 Result 2 Cluster Analysis 20](#_Toc152664080)

[Figure 17 Result 3 Cluster Analysis 21](#_Toc152664081)

[Figure 18 Reviews Count and Sentiment Analysis 21](#_Toc152664082)

[Figure 19 Reviews Count and Sentiment Analysis 22](#_Toc152664083)

[Figure 20 Reviews Count and Sentiment Analysis 23](#_Toc152664084)

[Figure 21 Keywords Analysis 24](#_Toc152664085)

[Figure 22 Keywords Analysis 25](#_Toc152664086)

[Figure 23 Common Words Analysis through Clustering 26](#_Toc152664087)

[Figure 24 Common Words Analysis through Clustering 26](#_Toc152664088)

# Abstract

This study utilizes complex hierarchical clustering techniques to analyze Pizza Hut customer reviews and accurately interpret customer sentiments. The study commences with an introductory section that emphasizes the crucial significance of customer attitudes in influencing business strategies, specifically within the food industry. By guaranteeing the dependability of subsequent studies, a clear explanation of the comprehensive process of collecting and preparing data can be found in the research tools and method section.

The result and visualization section reveals significant trends and patterns through informative visualizations. It offers a detailed comprehension of the total number of reviews, the analysis of sentiment over time, and the month-to-month variations. It highlights important insights that are essential for making strategic decisions. The keyword analysis part examines significant keywords such as staff, price, time, quality and online, providing valuable information into consumer preferences.

The core of the research is the analysis of common terms using hierarchical clustering, which identifies diverse customer experience clusters. The conclusion and recommendations consolidate the research findings, providing practical insights to improve customer satisfaction. This research enhances the field of social analytics by offering Pizza Hut a clear plan for making strategic enhancements and increasing customer involvement.

# 1. Research Rationale and Motivation

## 1.1 Introduction

In the modern era of digital technology, social media analytics is an essential tool for gaining a knowledge of the preferences, attitudes, and trends of customers. In order to accomplish the purpose of this study, a comprehensive assessment of Pizza Hut reviews that were obtained by web scraping from Trustpilot.com will be carried out. The objective of this project is to glean substantial insights that Pizza Hut could have the opportunity to utilize in the process of making strategic and tactical decisions. Our objective is to discover patterns that are associated with consumer satisfaction, to identify areas that require improvement, and to achieve a competitive edge in the food market, which is constantly evolving. This will be accomplished through the exploitation of data from customer reviews sites.

## 1.2 Business Aim

The primary objective of this study is to provide Pizza Hut with assistance in better comprehending the experiences and perspectives of their customers. The development of marketing strategies, the expansion of product offerings, and the growth of brand recognition are all areas in which this data can be incredibly beneficial. Through the utilization of the vast amount of information that is accessible internet, Pizza Hut may be able to acquire a profound comprehension of the preferences and expectations of their customers, which will ultimately result in an improvement in customer satisfaction and loyalty.

## 1.3 Scope of Research

This study's primary objective is to conduct an analysis of the reviews that Pizza Hut has received on Trustpilot.com, which is a well-known website that provides reviews of products and services. In order to simplify strategic decision-making, boost marketing initiatives, and enhance offers, the scope of this project includes the study of customer reviews. By appealing to the public's curiosity and harmonizing with consumer expectations, the study intends to provide significant viewpoints on the business approach that will be taken by Pizza Hut.

### 1.4 Report Layout

The study rationale, business goal, and scope are introduced in Section 1. The literature review in Section 2 reviews domain knowledge. Section 3 covers research tool selection, web scraping, API integration, dataset information, and data preprocessing. The report's heart is Section 4, with results and visualizations. Reviews count and sentiment analysis, keyword analysis, and clustering-based common word analysis are covered in subsections 4.1–4.3. Section 5 concludes with a summary, and Section 6 acknowledges study limitations. Section 7 concludes primary research with insights-leveraging suggestions. Section 8 appendices provide the codebase. This organized layout makes the research journey and its conclusions clear.

# 2. Literature Review

Through the development of a conceptual framework that takes into account the role of gender and visual attention in comments on online product evaluations and purchasing intention, this study investigated the impact that online product reviews have on the decisions that customers make regarding their purchases at retail stores(Chen *et al.*, 2022). Additionally, this provides a profound understanding of the fundamental mechanism by which internet reviews influence purchasing behavior, revealing the importance of gender in the process.

In order to better understand the influence that internet reviews have on the purchasing decisions of consumers, this study provides a conceptual framework(Le *et al.*, 2022). Written remarks, images that are pertinent to the review, and star ratings are the three primary components that could be found in internet reviews, according to the user. Furthermore, In order to improve reading comprehension of online reviews, an analysis of variance (ANOVA) test were utilized.

According to the findings of this study(Nowicki and Sikora, 2012), the key elements that influence purchasing decisions in the market for catering services are the quality and variety of dishes, the atmosphere both inside and outside of the location, and the pricing, which is similar to the factor that influences purchasing decisions in the food market as a whole.

According to the findings of the study(Singh *et al.*, 2019), customers placed the highest importance on the quality of the cuisine when they were dining at a restaurant. This was followed by the quality of the service, which they also considered to be very important. Consumers placed atmosphere in third place when it came to the importance of a product.

The findings of this study indicate that employees' skills, the atmosphere of the restaurant, the quality of the food, the cleanliness of the restaurant, and the amount of time spent waiting are all aspects that have a substantial impact on the online evaluations that customers leave for quick service restaurants(Norazha *et al.*, 2022).

This study makes use of a substantial amount of secondary research in addition to primary data obtained from a survey questionnaire that was delivered to walk-in customers at four Pizza Hut locations in Southeast London that were specifically chosen(Okobia, 2023). The findings indicate that there is a favorable association between the levels of customer satisfaction and the chance of acquiring complete brand loyalty towards Pizza Hut. The findings of this study establish a number of factors that have an effect on the degrees of satisfaction experienced by the respondents. The convenience, atmosphere, location of the outlet, quality of service and products, and promotional programs that Pizza Hut provides are some of the aspects that distinguish it from its competitors.

Identifying the factors that influence the purchasing decisions and overall satisfaction of Pizza Hut consumers, with the intention of providing the company with relevant recommendations that will assist management in formulating their strategies, is the goal of this study(Teoh *et al.*, 2021). A combination of primary and secondary data sources, such as research conducted on websites, were utilized in the course of this study. Primary data sources included questionnaires. According to the findings of the study, the majority of customers choose Pizza Hut because of its delicious flavor and satisfying experience, the freshness of its pizzas, recommendations from family and friends, cultural and habitual preferences, discounted prices, convenient online delivery services, excellent customer service, and the ease of payment and cashback options provided by the company.

# 3. Research Tools and Methods

## 3.1 Tool Selection

Data miner that enable efficient web scraping were utilised for data extraction(*Scrape data from any website with 1 Click | Data Miner*, 2023). The programming language Python was used for analysis because of its wide range of libraries, including Pandas, WordCloud, TextBlob, and nltk. In addition, Jupyter Notebooks were utilised for the purpose of organising and documenting the code. The selection of Python is in accordance with the project's technical specifications and the necessity for adaptability in managing various data formats.

## 3.2 Web Scraping Process

The process of web scraping involved using Data Miner to collect user reviews from Trustpilot.com. To make analysis easier, the dataset was then organised into a Pandas DataFrame. The ethical ramifications were thoroughly examined to guarantee compliance with Trustpilot.com's terms of service.

## 3.3 API Integration

Web scraping has shown to be a feasible substitute for data extraction, even though Trustpilot.com does not have a publicly accessible API. Using a technique called web scraping, DataMiner is a tool created to extract data from HTML web pages. This strategy made guaranteed that data was extracted ethically and efficiently.

## 3.5 Dataset

The following is the scrapped raw dataset.



Figure 1 Raw Dataset

There are three columns including the reviewer’s name, their review and the date that review was made.

## 3.4 Data Preprocessing

### 3.4.1 Sentiment Analysis and Keyword Analysis

This code segment organises the data by classifying reviews according to keywords and conducts sentiment analysis on each review using the TextBlob library. The pizza\_reviews DataFrame contains a new column called 'Sentiment' which stores the sentiment labels ('positive', 'neutral', 'negative').



Figure 2 Code Segment 1 Sentiment Analysis and Keyword Analysis

**Keyword-Based Review Classification**

The function passes through each key theme in the key\_themes array in order. A new boolean column is added to the pizza\_reviews DataFrame for each keyword. The column is named after the keyword, and its values indicate whether the related review incorporates the theme True or not False. This makes it easier to categorise each review based on the presence of keywords.

**Sentiment Analysis**

The code implements a method called analyze\_sentiment(review) which accepts a review as an argument and utilises TextBlob to determine its sentiment, categorising it as either 'positive', 'neutral', or 'negative'. For typical natural language processing (NLP) activities, like sentiment analysis, TextBlob offers a straightforward API and streamlines Python text processing. The function does sentiment analysis on every review in the 'Review' column of the pizza\_reviews DataFrame, and saves the outcomes in a newly created column called 'Sentiment'.

The result data frame is as follows.



Figure 3 Result 1 Sentiment Analysis and Keyword Analysis

The following code computes and presents the frequency counts for each keyword in the pizza\_reviews DataFrame. Additionally, it generates a summary DataFrame named "summary\_data" which includes the number of occurrences of True values for each theme and sentiment, as well as the total count of sentiments. The results are outputted for subsequent analysis and interpretation.

A white background with black text

Description automatically generated

Figure 4 Code Segment 2 Sentiment Analysis and Keyword Analysis

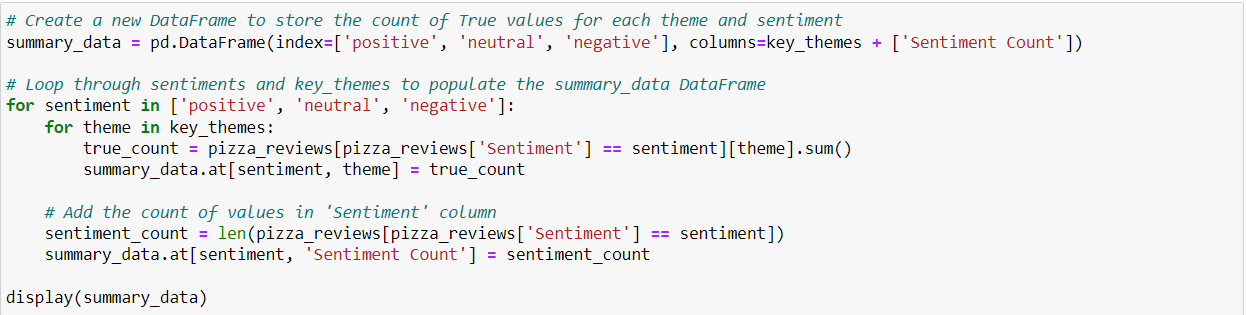


Figure 5 Code Segment 3 Sentiment Analysis and Keyword Analysis

**Calculation of the Frequency for Each Value in the Key\_Themes Column**

Instantiate a dictionary called frequency\_data with no initial values to hold frequency counts. Each keywords was iterated in the key\_themes collection. The frequency counts for each topic was computed and saved in the pizza\_reviews DataFrame using the value\_counts() function. the outcomes was saved in the frequency\_data dictionary.

**Generating a DataFrame Called "summary\_data" to Summarise the Data**

An empty DataFrame called summary\_data was instantiated, where each theme in key\_themes is represented as a column, along with an extra column for 'Sentiment Count'. The code iterates through several sentiments such as positive, neutral and negative and keywords in order to fill the summary\_data DataFrame. For every sentiment and keyword, The frequency of occurrences of the value "True" was calculated in the respective column representing themes for the specified sentiment. The count within the summary\_data DataFrame was stored. The aggregate count of reviews was calculated for each sentiment and was saved in the 'Sentiment Count' column.

The resulting dataframe is as follows.

A screenshot of a menu

Description automatically generated

Figure 6 Result 2 Sentiment Analysis and Keyword Analysis

### 3.4.2 Month-Over-Month Positive and Negative Review Count Change

The code examines sentiment data chronologically, primarily focusing on monthly patterns in favourable and unfavourable reviews, and computes the percentage changes in good and negative sentiments from one month to the next. The ultimate outcomes are displayed in a DataFrame to facilitate comprehension and subsequent examination.



Figure 7 Code Segment 1 Month-Over-Month Positive and Negative Review Count Change

**Performing Month-over-Month (MoM) Calculations**

Two additional columns were introduced, namely 'MoM\_Positive\_Increase' and 'MoM\_Negative\_Increase', that indicate the percentage variation in positive and negative reviews compared to the respective last month.

**Formatting and Presenting the Outcome**

The resulting DataFrame with columns for positive and negative reviews and their corresponding month-over-month percentage variations is displayed after rounding the percentage variations values.

The resulting data frame is as follows.

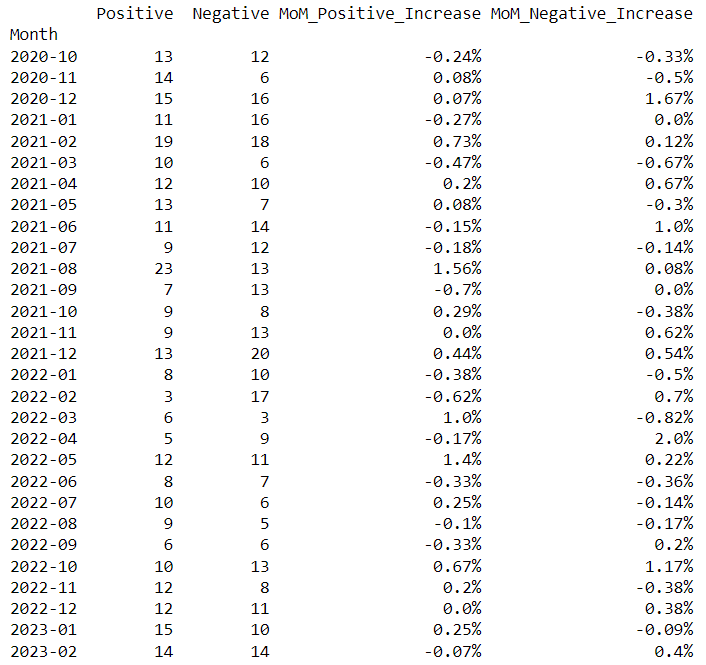


Figure 8 Result 1 Month-Over-Month Positive and Negative Review Count Change

### 3.4.3 Common Words Identification

This code segment does tokenization, and the frequency of words were calculated in positive, neutral, and negative evaluations. Subsequently, it detects frequently occurring words in the reviews that have a frequency exceeding 5 and the outcome was displayed. This result is valuable for comprehending commonly appearing words across various sentiment categories.



Figure 9 Code Segment 1 Common Words Identification

**Preparation of the Text**

The pizza\_reviews were filtered to categorise them into positive, neutral, and negative reviews. the reviews were aggregated into three distinct text strings: positive\_text, neutral\_text, and negative\_text.

**CountVectorizer**

The CountVectorizer module from scikit-learn was utilized to transform the textual input into a matrix of words. The vectorizer is initialized with the English stopwords obtained from NLTK. The fitting and transformation were applied to the positive text data and extracts the names of the features.

**Transform Word Matrices into DataFrames**

This function converts the word matrices representing positive, neutral, and negative reviews into DataFrames named positive\_word\_df, neutral\_word\_df, and negative\_word\_df, respectively.

**Identify Common Terms**

Terms that appear more than 5 times in any of the reviews were identified as given in the following sample.

The resulting list is as follows.



Figure 10 Result 1 Common Words Identification

A white rectangular sign with black text

Description automatically generated

Figure 11 Result 2 Common Words Identification

### 3.4.4 Clustering the Common Words with Word Embeddings and Hierarchical Clustering

This code is mostly performing clustering on common terms using word embeddings and hierarchical clustering. The clusters obtained demonstrate the presence of word groupings with comparable embeddings, indicating a certain degree of semantic similarity. The number of clusters is predetermined as 5, although it can be modified according to the desired level of detail in the clusters. Complete linkage and the distance measure of cosine similarity are used in the clustering process. The result displays each individual word along with its appropriate cluster assignment.



Figure 12 Code Segment 1 Clustering the Common Words with Word Embeddings and Hierarchical Clustering

Word Embeddings with spaCy: Uses spaCy (en\_core\_web\_md model) to obtain word vectors for each common word. Adds a new column ('Vector') to the DataFrame to store the word vectors.

**Word Embeddings with spaCy**

The spaCy library with the en\_core\_web\_md model was utilised to acquire word vectors for every frequently occurring word. A novel column ('Vector') was appended to the DataFrame in order to hold the word vectors.

**Hierarchical Clustering**

After filtering, if any viable word vectors remain, hierarchical clustering is performed. The word vectors were obtained, and Agglomerative Clustering was applied using five clusters. A novel column called Cluster was appended to the DataFrame.

The results are as follows:

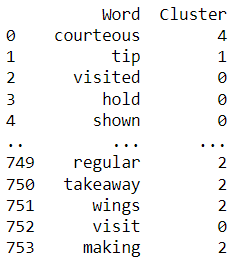


Figure 13 Result 1 Clustering the Common Words with Word Embeddings and Hierarchical Clustering

### 3.4.5 Cluster Analysis

The five clusters were carefully analyzed in order to determine the attributes by which they can be identified. Provided below is a compilation of words categorized under Cluster 1 for reference.



Figure 14 Result 1 Cluster Analysis

Each of those clusters was described using the following features, based on subjective criteria.

Cluster 1:

* Customer service and Customer experience.
  + Delay and Service Issues
  + Customer Interaction and Communication
  + Location and Timing
  + Customer Experience
  + Family and Celebration
  + Travel and Arrival

Cluster 2:

* The quality and taste of food, menu items, and overall dining experience
  + Food Quality and Taste
  + Menu and Items
  + Delivery and Packaging
  + Temperature and Freshness
  + Critique and Complaints
  + Specific Food Items

Cluster 3:

* Various aspects of the customer experience, transactions, and overall satisfaction
  + Customer Service and Transactions
  + Pricing and Deals
  + Ordering and Delivery Process
  + Customer Interaction
  + Quality and Satisfaction

Cluster 4:

* Numerical values and some location-related terms
  + Numerical Values
  + Location-Related Terms

Cluster 5:

* Expressing sentiments and opinions about experiences
  + Positive Expressions
  + Negative Expressions
  + Neutral Expressions

All clusters were considered except for Cluster 4, as it comprises numerical values and location-related phrases.

Requesting sentiment counts for each cluster, giving the frequency of words from the corresponding clusters in reviews of each sentiment. The counts of sentiment are determined by setting minimum threshold for occurrences. The following code segment calculates the frequencies of words in defined clusters within reviews of various moods such as positive, negative, neutral. Subsequently, a DataFrame named cluster\_counts\_df is generated to retain these counts.



Figure 15 Code Segment 1 Cluster Analysis

**Minimum Occurrence Criterion**

A threshold (min\_occurrences = 10) was established to determine the minimum number of occurrences needed in a review for a word within a cluster to be included in the count.

**Perform an Iterative Process to Examine and Analyze Clusters and Sentiments**

The nested loops was used to iterate through each cluster and sentiment, sequentially examining each one. The algorithm calculates the frequency of words within a specific cluster in reviews with a particular sentiment. The relevant entry was being modified in the cluster\_counts\_df DataFrame.

The result is as follows.

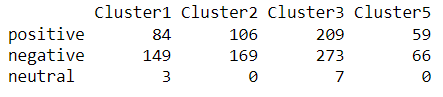


Figure 16 Result 2 Cluster Analysis

The value (Positive, Cluster1) = 84 indicates that there are 84 positive reviews that belong to Cluster 1, as these reviews contain more than 10 terms that are associated with Cluster 1. A review can contain many clusters.

The following presents the percentages of all these values in relation to the total number of reviews for each sentiment.

A number and numbers on a white background

Description automatically generated

Figure 17 Result 3 Cluster Analysis

The value 18.66% for (Positive, Cluster1) indicates that 18.66% of positive reviews belong to Cluster 1.

# 4. Results and Visualizations

## 4.1 Reviews Count and Sentiment Analysis

The following insights include the overall number of reviews, the number of reviews for each sentiment, and the corresponding percentage. This option enables users to promptly ascertain the total number of received reviews and their sentiment categorization.

A screenshot of a graph

Description automatically generated

Figure 18 Reviews Count and Sentiment Analysis

Based on the aforementioned statistics, it is evident that the number of positive reviews slightly exceeds the number of negative reviews, as indicated by the fact that 44.5% of the reviews are negative. However, it should be noted that this somewhat higher count of positive reviews does not necessarily indicate a definitive accomplishment. Furthermore, there is also a 10.5% portion of reviews that are classified as neutral.

The following line graph identifies the significant months with a high and low number of positive and negative reviews received.

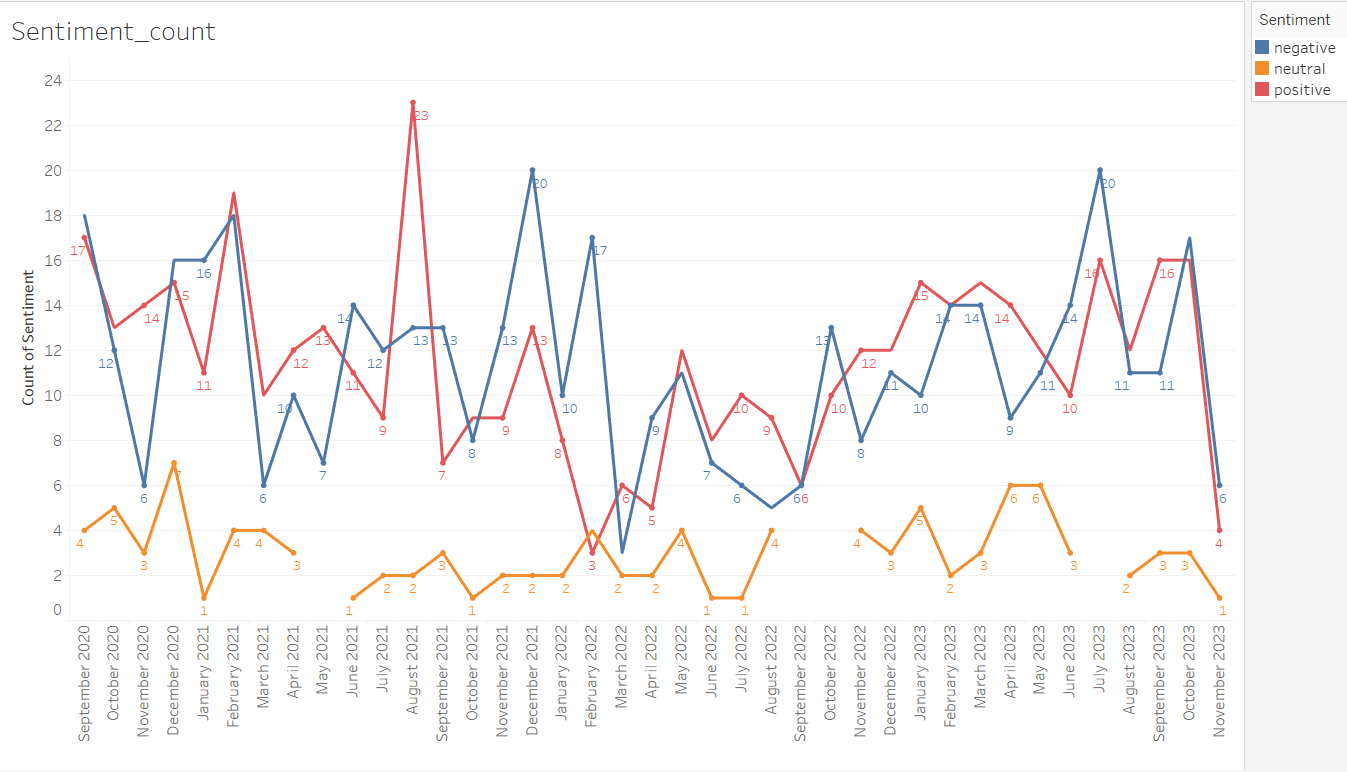


Figure 19 Reviews Count and Sentiment Analysis

The month of August 2021 recorded the maximum number of positive reviews, however it is evident that the count of positive reviews does not exceed 16 thereafter. That can be regarded as a drawback. December 2021 and July 2023 had the highest number of negative reviews. Given that one of the most recent months also had the highest number of negative reviews, it can be regarded as a negative aspect. In February 2022, the most significant difference between negative and positive ratings occurred, with 15 more negative reviews than positive reviews received during that month.

The following bar chart can be utilised to ascertain the change in the count of positive and negative reviews on a month-over-month basis. This feature enables the user to identify the months that experienced significant increases and decreases in the number of reviews, categorised by positive and negative sentiments.

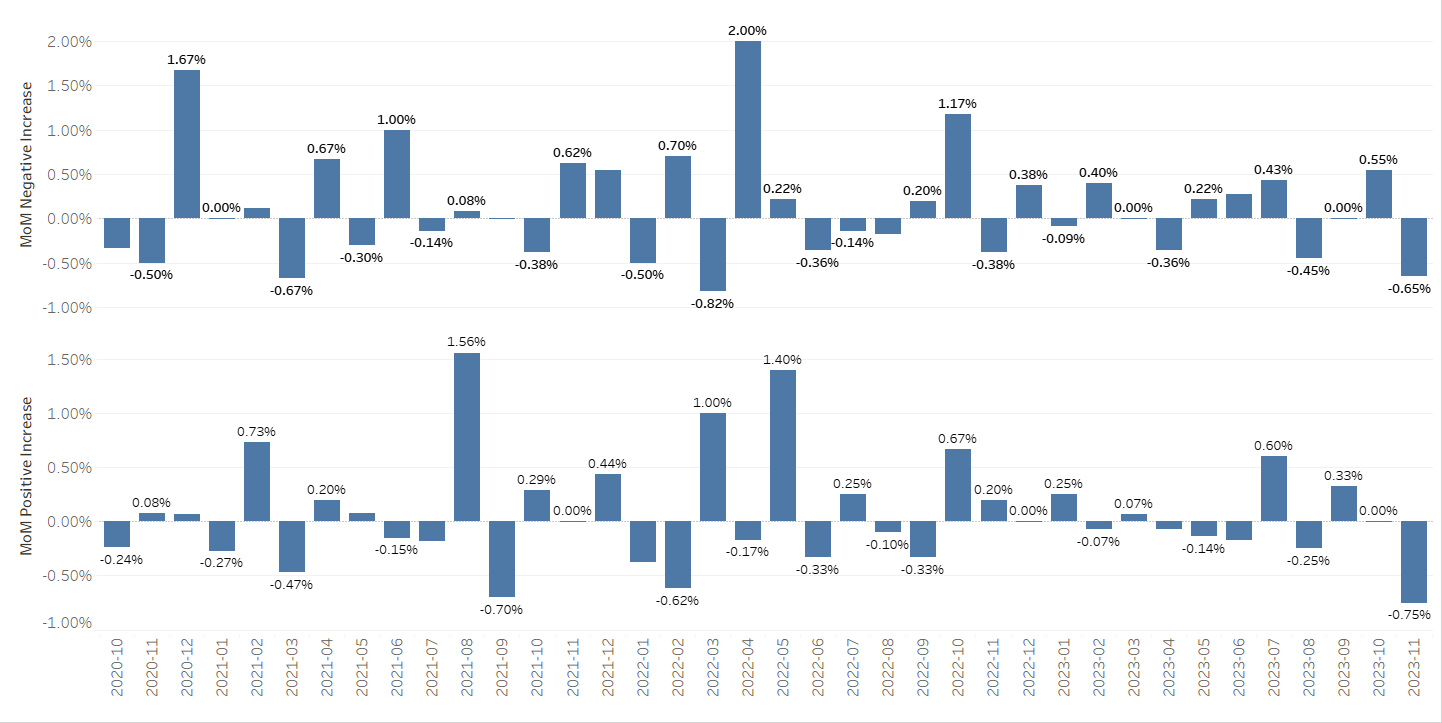


Figure 20 Reviews Count and Sentiment Analysis

The largest month-over-month (MoM) increase in the count of negative reviews occurred in April 2022. Prior to that, there was also a significant spike in November 2020. In March 2022, there was a 1% increase in positive reviews and a 0.82% decrease in negative reviews compared to the previous month. This can be seen as a productive month. The highest month-over-month (MoM) increase in positive review count occurred in August 2021, while the largest MoM decrease in positive review count occurred in the most recent month, November 2023.

## 4.2 Keywords Analysis

The bar charts shown below highlight the frequency of 5 specific terms - Package, Price, Quality, Staff, and Time - and their corresponding sentiment classes.

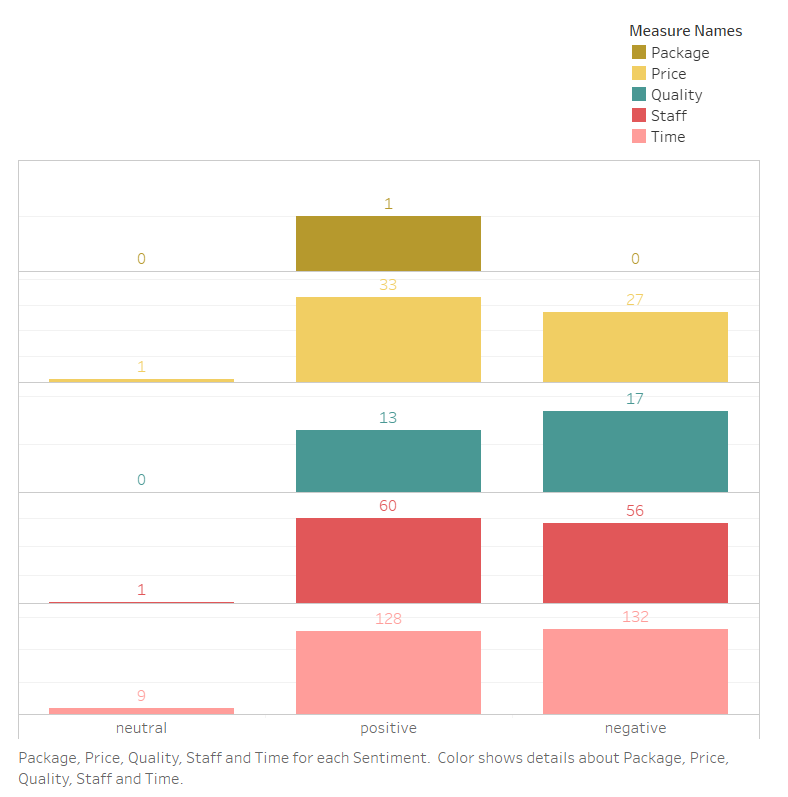


Figure 21 Keywords Analysis

Package, Price, and Staff keywords are more frequently found in positive reviews than in negative reviews. Even though this is a positive sign, Price and Staff have been mentioned considerable times in negative reviews as well. Quality and time are frequently mentioned in negative reviews, but they are also mentioned in positive reviews in considerable amount of times. When it comes to the amount of reviews, the Time keyword has 260 total reviews and 132 negative reviews. The Staff keyword has 116 total reviews and 60 positive reviews. Overall, these two factors are key factors on which consumers focus their attention. Pizzhut should evaluate these factors because both have a significant number of negative ratings.

The five keywords—customer service, delivery, menu, money, and online—as well as the number of each sentiment class are highlighted in the bar charts that follow.

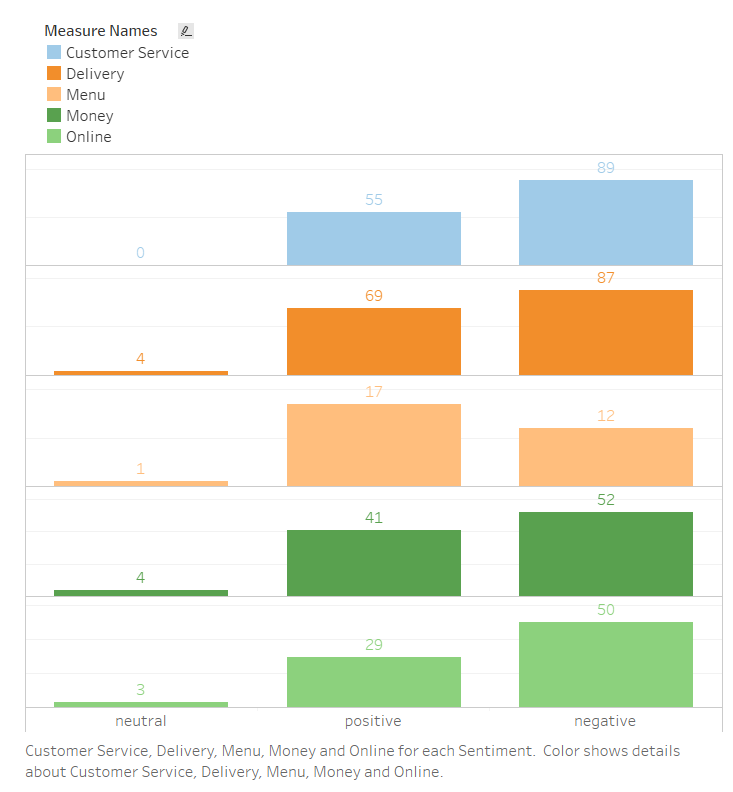


Figure 22 Keywords Analysis

The only keyword in this list with a higher positive review count is menu. Customer Service, Delivery, Money, and Online are the most often used terms in negative reviews. Customer Service and Delivery had the highest number of reviews, with 144 positive reviews and 89 bad reviews for Customer Service and 156 positive reviews and 87 negative reviews for Delivery. Pizzhut should take these factors into consideration. Despite the fact that the number of reviews for Online keyword is about 79, there are 50 negative reviews for Online keyword. It has the highest difference in the number of positive and negative ratings. There should be an issue with the online service.

## 4.3 Common Words Analysis through Clustering

The following bar charts represent the percentage of reviews under each sentiment class that was segmented under each cluster.

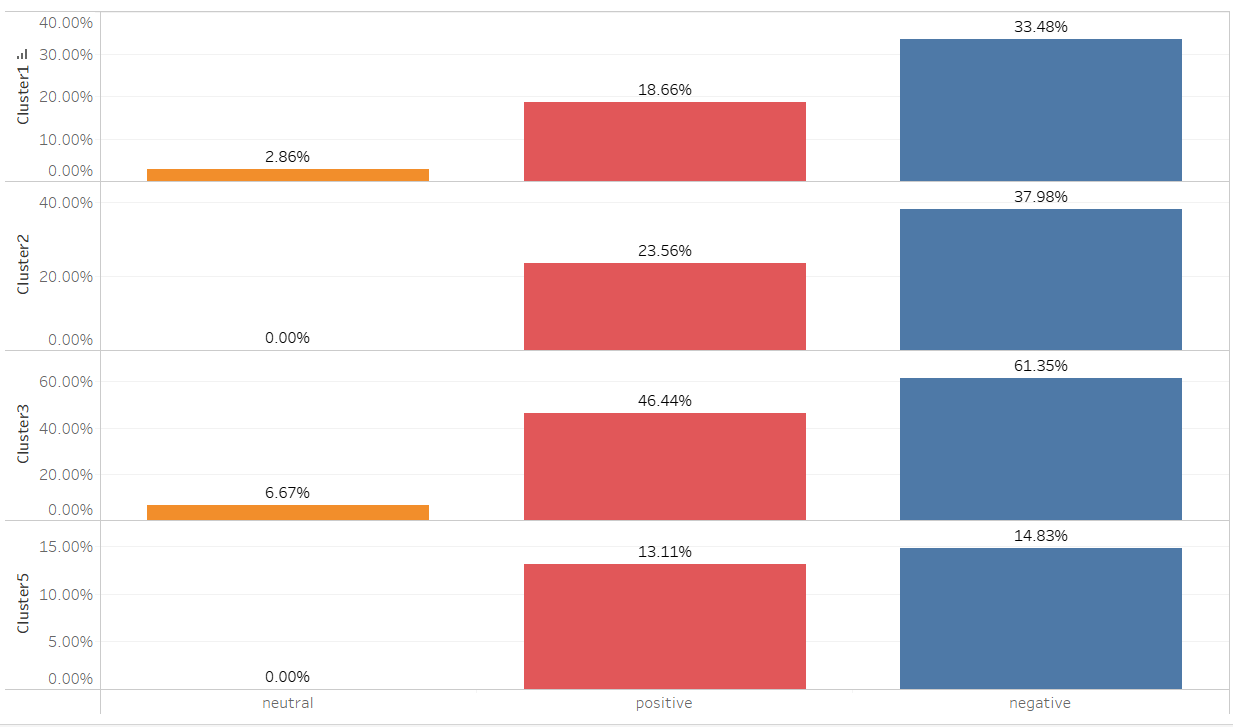


Figure 23 Common Words Analysis through Clustering

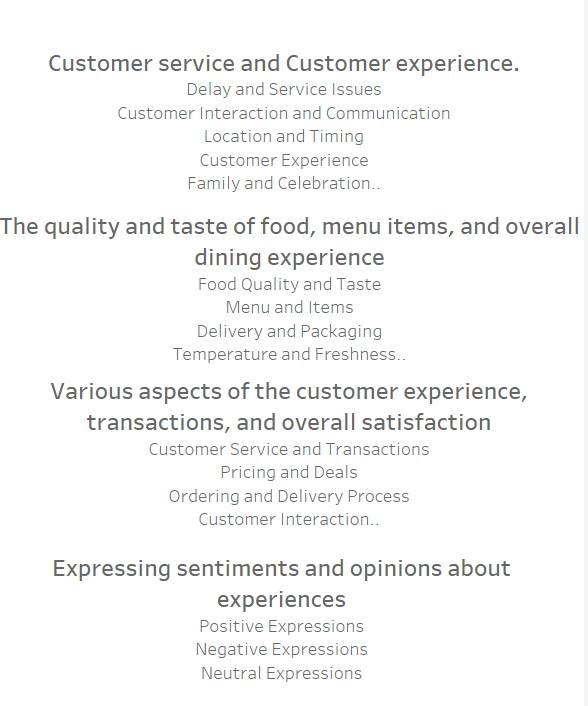


Figure 24 Common Words Analysis through Clustering

Cluster 3, which relates to customer experience, transactions, and overall satisfaction, has the highest percentage of reviews. Specifically, it accounts for 46.44% of positive reviews and 61.35% of negative reviews. These elements may be seen as the most salient features on which customers are concentrating. However, it is important to note that the proportion of negative reviews is greater than the proportion of positive ones. Cluster 1, which pertains to Customer Service and Customer Experience, and Cluster 2, which pertains to the quality and taste of food, menu items, and overall eating experience, both have greater percentages of negative ratings, precisely 33.48% and 37.98% respectively. The disparity in the percentages of positive and negative reviews, which is approximately 15%, is nearly same in both clusters. Given that 15% is a significant proportion, it is necessary to pay special attention to the issues addressed in Cluster 1 and Cluster 2 as well. Cluster 5 pertains to the expression of thoughts and ideas regarding experiences. The proportion of both negative and favourable ratings is lower, with a difference of approximately 1% between the two.

# 5. Conclusion

This study provides important results about Pizza Hut reviews, as well as practical facts for making informed strategic decisions. The usefulness of social media analytics is demonstrated in detecting important patterns from unstructured data by using Data Miner and Python for site scraping and analysis.

To sum up, analyzing PizzaHut's reviews offers insightful information about the opinions of customers and important issues. The overall review count reveals a significantly greater proportion of positive reviews, although the existence of a notable percentage of negative and neutral reviews implies areas that should be enhanced. The MoM analysis reveals variations in the number of positive and negative reviews, underscoring the significance of regularly addressing consumer reviews.

The keyword analysis indicates that specific factors, such as "Time" and "Staff," attract considerable customer attention, generating both positive and negative feedback. Moreover, the analysis of common words using clustering techniques reveals distinct groupings, such as Cluster 3 (related to customer experience and happiness), which are of utmost importance for PizzaHut.

# 6. Limitations

The limitations of this research include the potential for bias in assessments since people who hold strong opinions are more likely to comment. Moreover, web scraping—a technique vulnerable to changes in the website's architecture—was necessary due to the absence of a publicly accessible application programming interface (API) for Trustpilot.com.

# 7. Recommendations

Future research could concentrate on improving sentiment analysis methods and looking at additional data sources in order to address the constraints. Creating an approved API in association with Trustpilot.com would resolve ethical issues and enhance the reliability of data retrieval.

**Acknowledge and tackle negative aspects**

Attention should be directed to the aspects emphasized in Cluster 3, since it exhibits the highest proportion of reviews, but also a greater proportion of negative reviews.  Issues pertaining to transactions, client experience, and overall satisfaction should be identified and resolved.

**Enhance the quality of customer service and optimize the overall customer experience.**

To raise overall customer happiness, it is advisable to prioritise changes in customer service, experience, and food quality, as indicated by the greater negative review percentages in Clusters 1 and 2.

**Optimize Internet-based services**

A particular emphasis should be given to the term "Online" as it signifies possible complications with internet-based services. It is a need to conduct an inquiry and enhance the online ordering procedure in order to diminish negative reviews in this particular domain.

**Track and analyze monthly patterns**

Monthly patterns should consistently be observed to discern trends and comprehend client sentiments as they evolve. Any sudden surge in negative reviews should be responded to swiftly and consistently.

**Interact with customer feedback**

It is important to proactively interact with client input, including both positive and negative, on social media sites. Replies should be made to reviews as a means of demonstrating recognition and dedication to resolving any issues raised.

**Advertising and Publicity**

It is possible to utilize positive views, particularly from the most optimistic review month (August 2021), to guide marketing tactics. Implementing promos or campaigns can be considered as a means to uphold positive customer sentiments.

**Staff Development**

Supplementary training can be offered to employees, particularly in the specific areas that have been emphasized by customer feedback. This can lead to enhancements in the quality of service, leading to an increase in positive ratings.

**Standard Evaluation**

Perform frequent assessments of client reviews to remain knowledgeable about changing patterns and mindsets. By engaging in this continuous process, PizzaHut will be able to effectively adjust to evolving client expectations and preferences.

To summarize, incorporating these suggestions can enhance the entire consumer experience, boost positive opinions, and address any possible issues raised by customer feedback.

# References

Chen, T. *et al.* (2022) ‘The Impact of Online Reviews on Consumers’ Purchasing Decisions: Evidence From an Eye-Tracking Study’, *Frontiers in Psychology*, 13, p. 865702. Available at: https://doi.org/10.3389/fpsyg.2022.865702.

Le, L.T. *et al.* (2022) ‘Online reviews as a pacifying decision-making assistant’, *Journal of Retailing and Consumer Services*, 64, p. 102805. Available at: https://doi.org/10.1016/j.jretconser.2021.102805.

Norazha, N. *et al.* (2022) ‘Influence of quick-service restaurant’s service quality towards customer online review’, 14, pp. 97–129.

Nowicki, P. and Sikora, T. (2012) ‘CONSUMER BEHAVIOUR AT THE FOOD MARKET’, in.

Okobia, E. (2023) *Customer Satisfaction and Brand Loyalty: Pizza Hut Case Study | The International Journal of Business & Management*. Available at: https://www.internationaljournalcorner.com/index.php/theijbm/article/view/173292 (Accessed: 4 December 2023).

*Scrape data from any website with 1 Click | Data Miner* (2023). Available at: https://dataminer.io/ (Accessed: 4 December 2023).

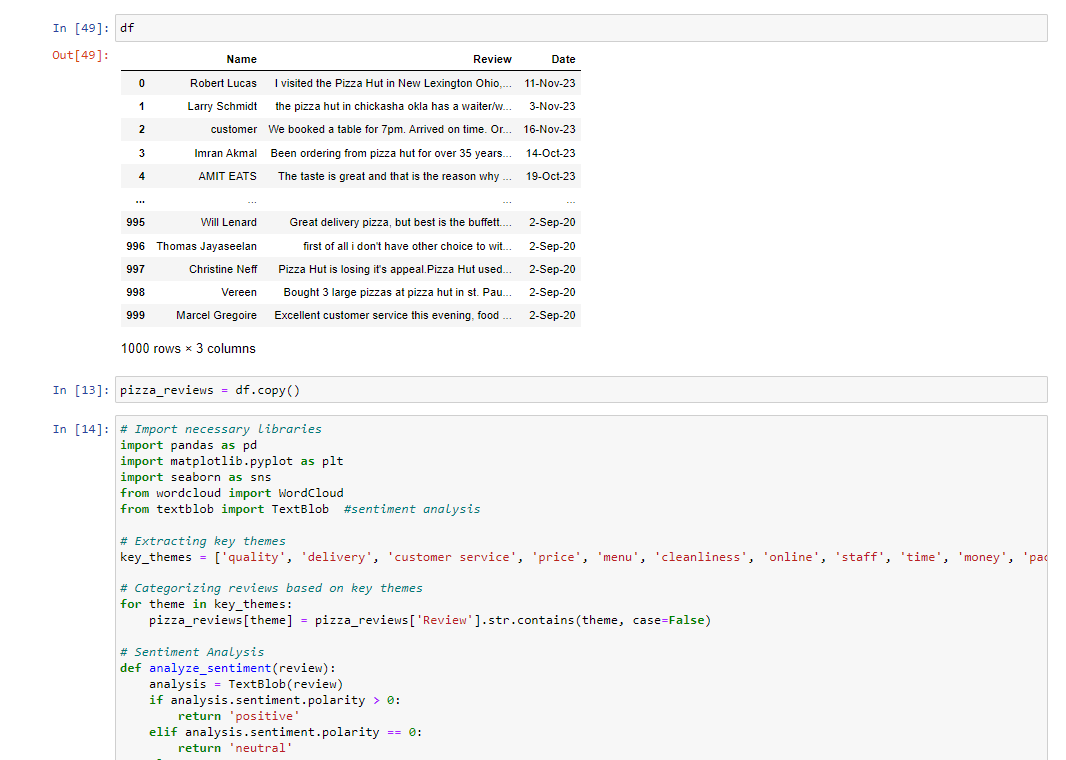
Singh, U. *et al.* (2019) ‘A study on Impact of Consumer Reviews on Consumer Behavior with reference to Restaurants in Bengaluru’, *International Journal of Multidisciplinary Research Review*, 4, pp. 532–538. Available at: https://doi.org/10.5281/zenodo.2580240.

Teoh, K. *et al.* (2021) ‘Factors influencing customers’ purchasing behavior and satisfaction in Pizza Hut’, 2, pp. 39–56.

# 5. Appendices

Tableau Dashboard Link:

Notebook:



A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer program

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer program

Description automatically generated

A screenshot of a computer screen

Description automatically generated

A screenshot of a computer screen

Description automatically generated

A screenshot of a computer program

Description automatically generated

A screenshot of a computer program

Description automatically generated

A screenshot of a computer

Description automatically generated