**Analyzing Customer Reviews of McDonald's Stores in the United States**

Kalaiarasi Kaliappan, Manoj Athreya Chatnalli Satish, and Sunil Hule

MPS – Data Science, University of Maryland, Baltimore County

DATA 690: NLP for Practitioners

Dr. Antonio Diana

**Abstract**

The project aims to examine the emotions expressed in more than 33,000 reviews of McDonald's stores. The main objective is to gain insights into customer sentiment, which can help make informed decisions to improve the brand image. The dataset includes store information, review ratings, texts, and timestamps. The approach involves preparing the data to ensure that the text is suitable for analysis, followed by categorizing sentiments as positive, negative, or neutral to understand satisfaction levels. Furthermore, a rating-based analysis will be performed to recognize any variations in sentiment. This project's expected outcomes include understanding customer sentiment and pinpointing improvement areas based on specific customer experiences. This project compares the sentiments based on the various rating levels. Overall, this project contributes towards McDonald's goal of identifying customer sentiments and strengthening its brand image by providing insights derived from detailed sentence-level sentiment analysis of customer reviews.

Keywords: sentiment analysis, machine learning, transformers, natural language processing

**Analyzing Customer Reviews of McDonald's Stores in the United States**

Sentiment mining holds significant importance within data mining, enabling the extraction of crucial information in the gathered data. Sentiment Analysis, also called Opinion Mining, involves the application of natural language processing, text analysis, and computational linguistics to discern and extract subjective information from source materials.

Though customer feedback in the stores will be appreciated for improvement (Wang et al., 2015), online comments or reviews might still leave the business room for improvement.

Sentiment analysis involves assessing the polarity of reviews, classifying them as positive, neutral, or negative based on the expressed opinions. Through polarity analysis, we seek to determine whether a review leans towards a positive, neutral, or negative stance. For instance:

* Positive Review Example:

Review: "The McDonald's staff at Central provided excellent customer service, and the food was delicious." In this example, the positive sentiment is evident in the favorable comments about the excellent customer service and delicious food.

* Neutral Review Example:

Review: "The mobile order at the drive-thru speaker was processed efficiently, and the food was as expected." The comments express a straightforward acknowledgment of an efficient mobile order process and the food meeting expectations without strong positive or negative sentiments.

* Negative Review Example:

Review: "Repeated my order three times in the drive-thru and still managed to mess it up." This review conveys a negative sentiment due to the frustration about repeating the order multiple times in the drive-thru.

In Sentiment Analysis, a notable challenge lies in the subjective nature of opinion words, wherein a term viewed positively in one context may be perceived negatively in another. Moreover, the intensity of positivity or negativity significantly influences opinions; for instance, distinguishing between "good" and "very good" is crucial. By examining the subtleties, we can distinguish between varying degrees of positivity or negativity, enhancing sentiment classification accuracy.

So, In this study, we employ VADER and BERT techniques for sentiment analysis on customer reviews, aiming to discern the intensity of the sentiments accurately. Our hypothesis asserts that 1-star ratings exhibit significant negative sentiment scores, substantiating the strong negativity associated with lower ratings. Additionally, we integrate topic modeling to unveil nuanced themes within the text, augmenting our sentiment analysis by providing a deeper contextual understanding of expressed positive, negative, or neutral sentiments. This Collective approach enhances the depth and precision of our sentiment analysis methodology.

**Literature Review**

Sentiment analysis has been used for a range of purposes such as tracking the popularity and desirability of a brand (Greco & Polli, 2019), identifying product opportunities ([Jeong et al., 2017](https://www.sciencedirect.com/science/article/pii/S0268401219311181?casa_token=ezbrDSZwPR0AAAAA:DeEfWA6TfB9fPRXF1kWdGR-z888WijMn5AH_joPbH-sZWR_OMoxmJsWiAZTssHIuhPDIrkBIVA#bib0130)), studying product launches (Rathore & Ilavarasan, 2020), predicting market movement of stocks (Maqsood et al., 2020).

Topic modeling, a sophisticated technique in natural language processing (NLP), unveils the hidden thematic structure within a collection of text documents. In this context, the dataset with meticulously cleaned review text, and rating presents a fertile ground for topic modeling to flourish (Kerwa & Bhansal, 2018). We are utilizing Topic Modelling to explore the topics

While lexicon-based models like SentiWordNet offer simplicity, (Alaparthi, S., Mishra, M, 2021) suggest their performance falls short compared to supervised learning approaches. This paper provides evidence, demonstrating the clear advantage of BERT, a pre-trained deep learning model, in sentiment analysis over both traditional and other deep learning methods. Building upon this strong foundation, our project also leverages BERT's capabilities to uncover valuable insights from customer reviews.

**Methodology**

**About the Data**

The dataset comprises over 33,000 anonymous reviews of McDonald's stores in the United States, sourced from Google reviews. It is a valuable resource for understanding customer experiences and opinions against ratings across various McDonald's locations (see Table 1).

**Table 1**

*Description of the dataset*

| Description | |
| --- | --- |
| Store Names | Identifiers for McDonald's locations. |
| Categories | Categorization information for each store. |
| Addresses | Physical locations of the stores. |
| Geographic Coordinates | Latitude and longitude data. |
| Review Ratings | Numeric ratings given by customers. |
| Review Texts | Actual textual reviews provided by customers. |
| Timestamps | Date and time information of the reviews. |

**Data Cleaning**

The raw data underwent a meticulous cleaning and modification process to ensure consistency and relevance, addressing extraneous words and text while also adapting the data types for enhanced analysis. This refinement aimed to improve the contextual integrity of the data without compromising its meaning. The cleaning and modification process included several essential steps:

***Removal of Special Characters and Symbols:*** Extraneous characters and symbols were eliminated to enhance text clarity and maintain uniformity.

***Whitespace Removal:*** Unnecessary white spaces were eliminated to streamline the text, ensuring a cleaner and more coherent dataset.

***Tokenization:*** The text was broken down into individual tokens, facilitating further analysis by organizing it into manageable units.

***Stop-word Removal:*** Common stop words were excluded to concentrate on the most meaningful content, eliminating frequently used but less informative words.

***Lemmatization:*** Words were reduced to their base or root form through lemmatization, ensuring consistency in language representation and enhancing the accuracy of subsequent analyses.

***Data Modification:*** The data underwent modifications, including changes in data types where necessary. This ensured that the data was appropriately formatted for the intended analyses.

**Exploratory Data Analysis**

After cleaning up the data, we wanted to know the distribution of customer ratings for McDonald's, for which we used a bar plot. As shown in Figure 1, the graph suggests that most customers have a positive experience at McDonald's. However, there is room for improvement as the number of 1-star ratings is also significantly higher.

**Figure 1**

*Sample Count For Every Star Rating*

**

The word cloud in Figure 2 presents a visual summary of the most frequent terms found in customer reviews of a fast-food chain, with the prominence of each word correlating to its frequency of occurrence. Notably, the brand name is centrally featured, suggesting it is a common reference point in reviews. At the same time, terms like "service," "staff," "clean," and "location" highlight critical areas of customer concern and satisfaction.

Following the high-level data analysis, we wanted to know the most frequently used bigram in each rating. The bar chart in Figure 3 illustrates the frequency of the top five bi-grams found in reviews for each star rating category, from 1 to 5 stars. The bi-grams indicate customer sentiment, with negative service-related terms such as "drive-thru" appearing more frequently in 1-star reviews and favorable terms such as "good service" being more prevalent in higher-rated reviews.

**Figure 2**

*Word Cloud for Frequency Analysis*



**Figure 3**

*Countplot for Top 5 Bi-grams with respect to Rating Stars*

## 

It is also evident that in some lower ratings, reviews mention the long wait time while the franchise is expected to serve the food fast. This frequency analysis also underscores the importance of service quality in customer satisfaction, as evidenced by the linguistic patterns in customer feedback.

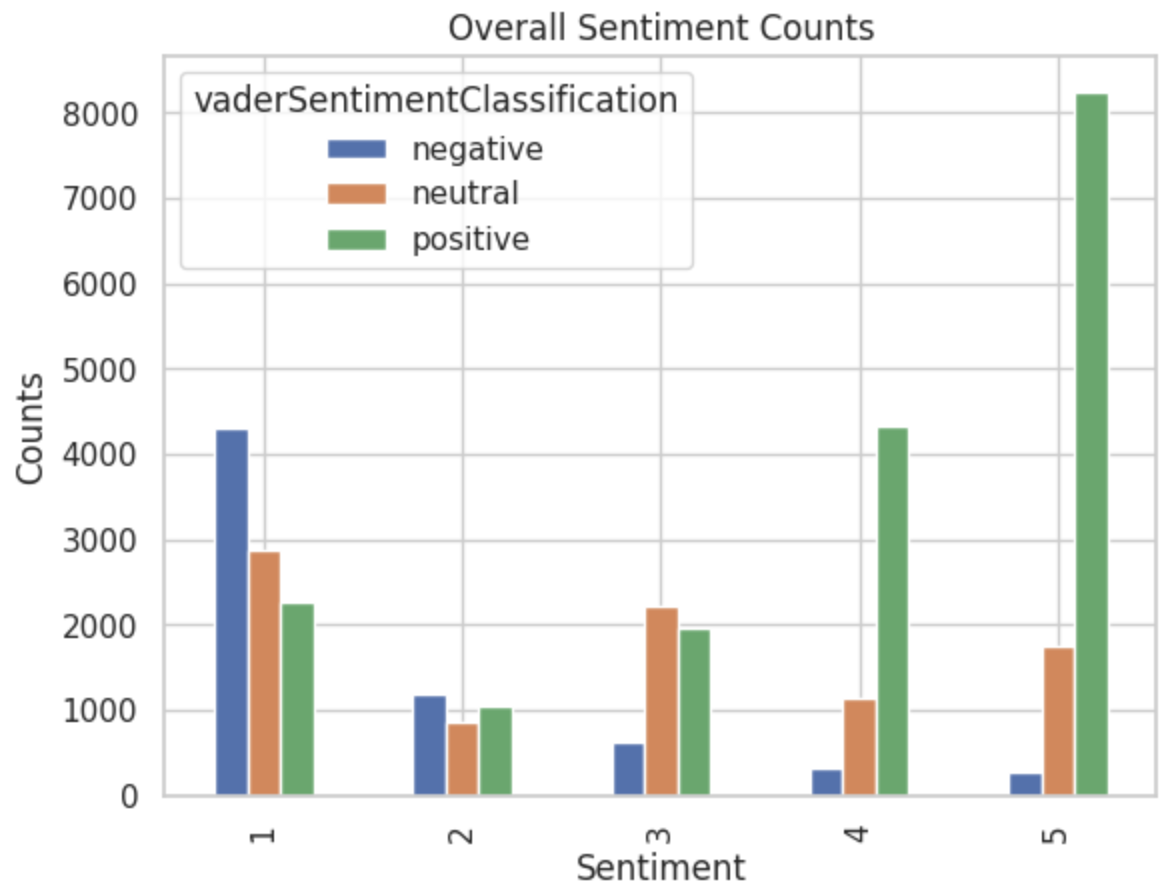
**Sentiment Analysis**

In sentiment analysis, we will assess the sentiments of the review statements and assign the appropriate sentiment, namely 'Negative', 'Positive', or 'Neutral'. We have employed techniques like VADER (Valence Aware Dictionary for Sentiment Reasoning), a lexicon-based model to categorize the reviews as positive, negative, or neutral. Output from the VADER model is a compound of a sentiment ranging from -1 to 1, where a value less than -0.05 is considered negative, values greater than 0.05 are considered positive, and the rest are considered neutral (Hutto & Gilbert, 2014). We have also used an Attention-based BERT model to get the sentiments for each review. The output of a softmax function from the BERT model gives the probability of the review being negative, neutral, or positive. However, there are significant differences in the sentiments assigned by both BERT and VADER.

In Figure 4, which represents the sentiment counts for each star by VADER, we can see the dominance of negative sentiment in a 1-star rating. As we move from 1 star to 5 stars, the negative sentiment decreases, and the positive sentiment increases quickly.

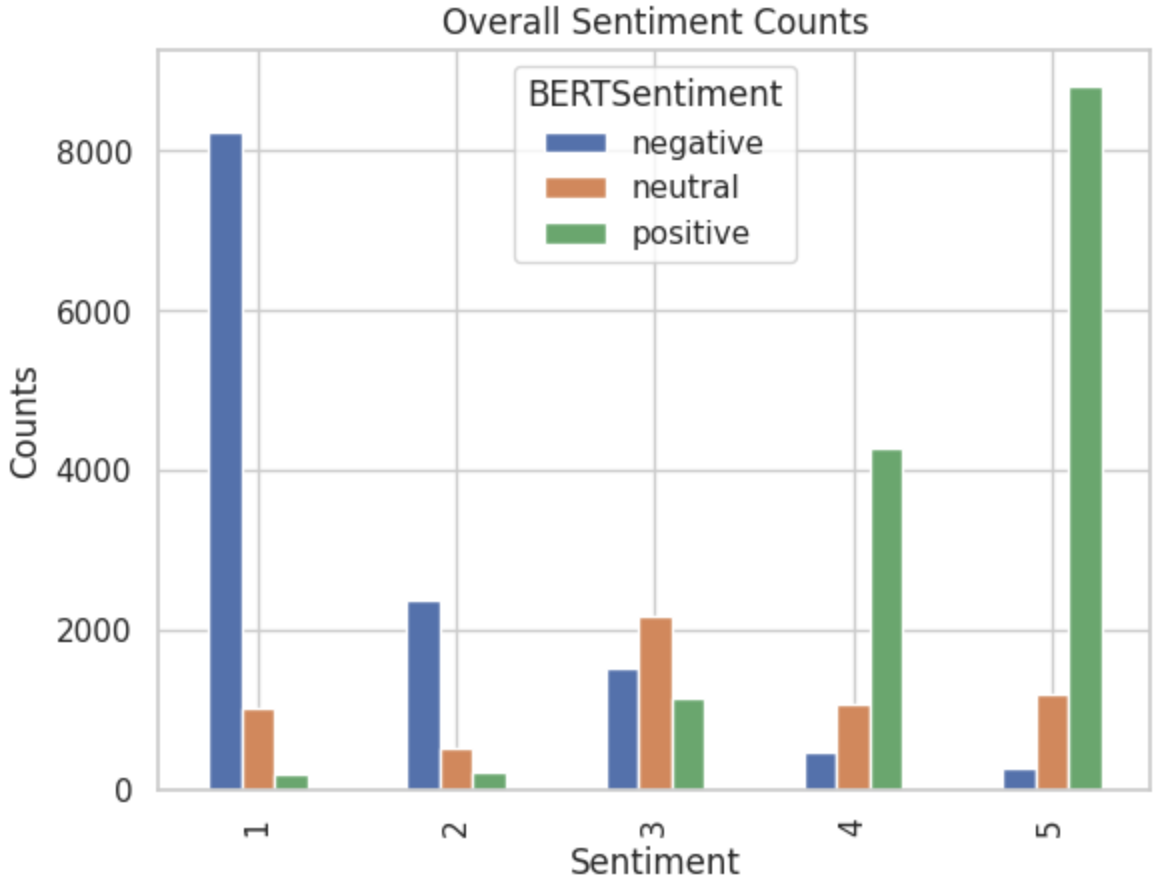
**Figure 4**

*Sentiment Counts of Each Star reviews by VADER*

****

**Figure 5**

*Sentiment Counts of Each Star reviews by BERT*



In a research comparing Attention-based and Lexicon-based sentiment classifiers (Subrata et al., 2023) researchers claim that BERT performs better than VADER in terms of sentiment classification. Figure 6 shows a few reviews and the corresponding classification results from BERT and VADER. From visual inspection and research findings (Subrata et al., 2023), BERT performs better.

**Table 2**

| Rating | Review | | | | | | VADER | BERT |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 1 | Always out of nearly everything regardless of the time of day no value menu items are ever available past noon and good luck getting anything but fries after 7 | | | | | | Positive | Negative |
| 1 | Love McDonalds but in this one I have had the worst service of the century | | | | | | Neutral | Negative |
| 1 | Its breakfast sandwiches are human meat | | | | | | Neutral | Negative |
| 1 | Not open 24 hours Only the drive thru Dining room actually closes very early at 10pm Pretty dirty location too | | | | | | Neutral | Negative |
| 1 | Terrible customer service and no juice for the kids If your kids don't drink soda go elsewear | | | | | | Neutral | Negative |

**Topic Modeling**

Topic modeling can shed light on the factors that contribute to "1 - Star", "2 - Star", "3 - Star", "4 - Star", and "5 - Star" ratings, enabling McDonald's to prioritize improvements in areas that matter most to its customers. By identifying the topics that consistently evoke positive or negative sentiments, McDonald's can focus on enhancing the aspects of its operations that drive customer satisfaction.

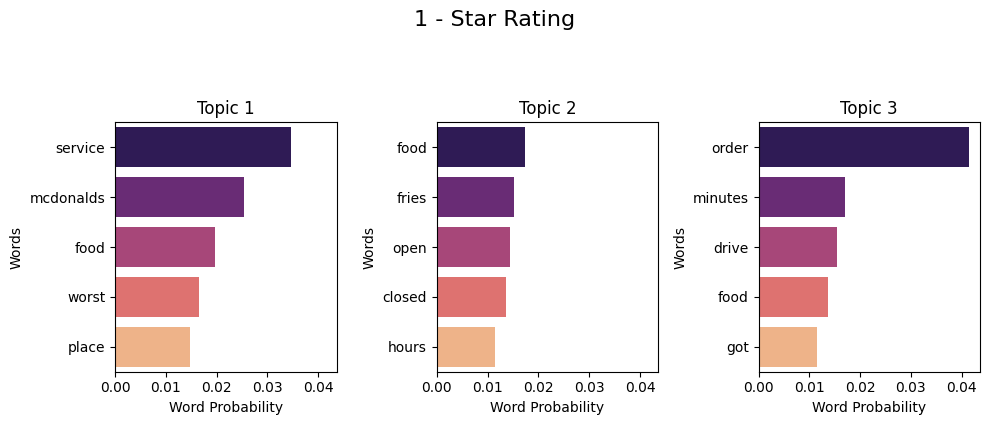
In essence, topic modeling is a powerful tool for extracting actionable insights from customer reviews, empowering McDonald's to make informed decisions that enhance customer satisfaction and strengthen its brand reputation (Investopedia, n.d.).

The topic modeling output for 1-star reviews of a McDonald's store in Figure 6 suggests three main areas of customer dissatisfaction. Topic 1 highlights issues with 'service', indicating that service quality is a primary concern among dissatisfied customers. Topic 2 emphasizes operational problems, with 'food', 'fries', and store 'hours' being significant, which may reflect issues with product availability and store timings. Lastly, Topic 3 focuses on the 'order' and 'drive' experience, suggesting that wait times and the drive-thru experience also contribute to negative reviews.

To address the issues from 1-star reviews, McDonald's should prioritize staff training for better service, improve inventory and staffing for product availability, and enhance the drive-thru system to decrease wait times. These targeted actions are likely to elevate the customer experience and mitigate the main concerns of dissatisfaction.

**Figure 6**

*Top 3 Topics of 1-star Ratings*

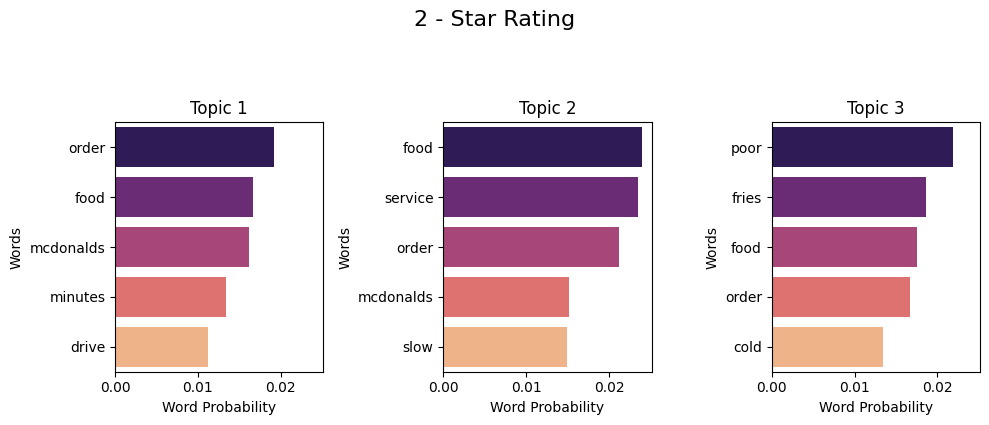
******

The topic modeling for 2-star reviews of McDonald's in Figure 7 reveals areas of customer dissatisfaction that are less severe than those in 1-star reviews. Topic 1 is focused on 'order' and 'food', suggesting that problems with food orders may not be as critical as those highlighted by 1-star reviews. Topic 2 includes 'food' and 'service' with an emphasis on 'slow' service, indicating delays as a significant concern. Topic 3 introduces 'poor' quality related to 'fries' and 'food' being 'cold', highlighting specific food temperature and quality issues. These insights can guide speed and quality control improvements to enhance customer experience.

To improve based on the 2-star reviews, McDonald's should focus on enhancing the accuracy and efficiency of the order process, addressing the causes of service delays, and implementing stricter quality controls to ensure the food served is hot and fresh.

**Figure 7**

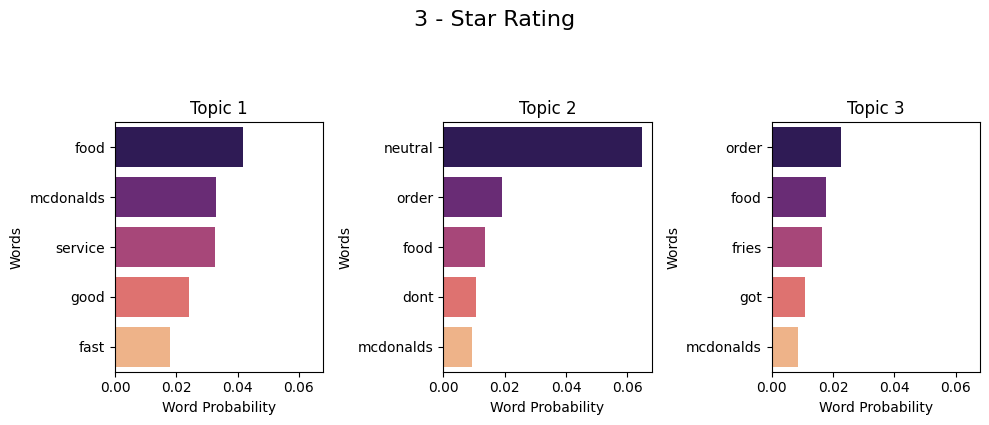
*Top 3 Topics of 2-star Ratings*

******

The topic modeling for 3-star reviews in Figure 8 displays in a nuanced view of customer feedback. Topic 1 suggests that some customers are satisfied with the 'service' and find the food 'good' and 'fast', pointing to positive aspects even in average-rated experiences. Topic 2 shows terms like 'neutral' and 'do not', indicating a level of indifference or a lack of solid sentiment, which is characteristic of middling ratings. Finally, Topic 3, with words like 'order', 'food', and 'fries', likely reflects common themes in fast-food dining experiences, with 'got' possibly indicating a satisfactory fulfillment of orders, which aligns with a moderate 3-star rating.

**Figure 8**

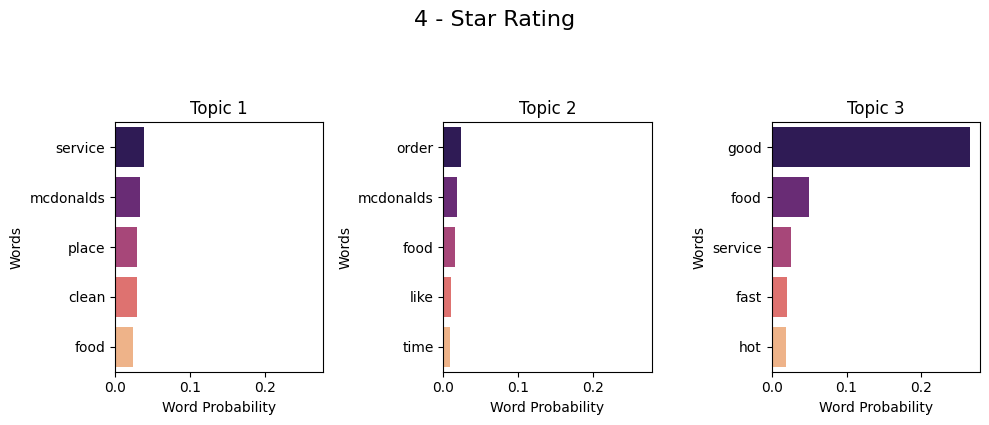
*Top 3 Topics of 3-star Ratings*

******

The topic modeling for 4-star reviews (in Figure 9) of a McDonald's store reflects more positive customer experiences. Topic 1 emphasizes 'service' and 'clean', suggesting that cleanliness and good service are notable factors in higher-rated experiences. 'Place' also features prominently, which could imply that the store's environment is a factor in customer satisfaction. Topic 2 focuses on the efficiency and quality of the 'order' process. In contrast, Topic 3 strongly highlights 'good' alongside 'food', 'service', 'fast', and 'hot', indicating that these are key attributes associated with satisfaction in the context of a 4-star review. These topics suggest that service quality, order efficiency, and food quality significantly positively affect customer reviews.

**Figure 9**

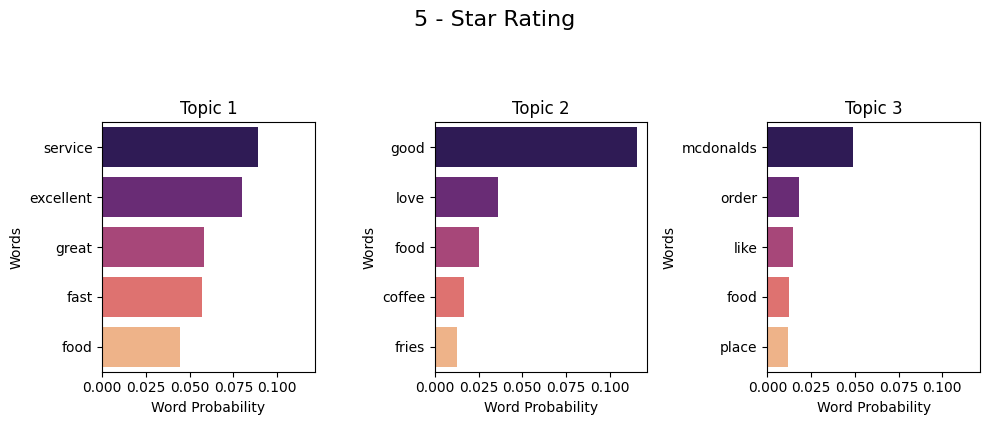
*Top 3 Topics of 4-star Ratings*

******

In Figure 10, the output of the Topic Modeling for rating 5-star is displayed. Topic 1 features words like 'service', 'excellent', 'great', and 'fast', which suggest that customers associate positive experiences with both the quality and efficiency of service. Topic 2 includes 'good', 'love', 'food', and 'coffee', implying a solid appreciation for the food and beverages served, with 'love' indicating a powerful positive sentiment. Finally, Topic 3 shows that even among the highest ratings, the brand ('mcdonald'), the 'order' process, and the 'place' are essential to customers, with 'like' reinforcing the positive sentiment.

**Figure 10**

*Top 3 Topics of 5-star Ratings*



**Hypothesis Testing**

Based on the Sentient Analysis and Topic Modelling results, it is evident that the number of negative reviews for 1-star is significantly larger than that for higher star ratings. From these results, we have formulated the null and alternate hypotheses as follows:

**Null Hypothesis (H0): The average of negative sentiment scores with a 1-star rating is equal to that of negative sentiment scores with a 2-star rating and above.**

**Alternate Hypothesis (H1): The average of negative sentiment scores with a 1-star rating is greater than that of negative sentiment scores with a 2-star rating and above.**

We performed a non-parametric (Mann-Whitney U) test, which resulted in a very small p-value. This makes us reject the null hypothesis and accept the alternate hypothesis. Additionally, we performed another hypothesis test to compare the average negative sentiment of 1-star reviews and the average negative sentiment of 2-star reviews.

**Null Hypothesis (H0): The average of negative sentiment scores with a 1-star rating is equal to the average of negative sentiment scores with a 2-star rating.**

**Alternate Hypothesis (H1): The average of negative sentiment scores with a 1-star rating is greater than the average of negative sentiment scores with a 2-star rating.**

Performing the Mann-Whitney U test, the p-value was extremely small. Therefore, we reject the null hypothesis and accept the alternate hypothesis. Both hypothesis testing results prove that the average negative sentiment of a 1-star rating is significantly higher than higher ratings, indicating that McDonald’s should consider the negative elements embedded in 1-star ratings to improve their service.

**Conclusion**

This project delved into the emotions expressed in over 33,000 McDonald's reviews. By employing VADER and BERT sentiment analysis techniques, we uncovered a strong correlation between lower star ratings and more negative sentiment, with a specific focus on service, food quality, freshness, and timely delivery. Topic modeling further revealed underlying themes within the reviews, highlighting customer frustrations with unhelpful staff, incorrect orders, and lengthy wait times. Notably, BERT outperformed VADER in its ability to identify subtle nuances and provide actionable insights. Based on these findings, we recommend that McDonald's prioritize improving staff training, enhancing order accuracy, maintaining clean store environments, and optimizing drive-thru and lobby operations. By addressing these key areas, McDonald's can effectively address customer concerns, improve its brand image, and ultimately achieve its customer-centric goals.

**References**

A.K. Rathore, P.V. Ilavarasan, 2020. Pre-and post-launch emotions in new product development: Insights from twitter analytics of three products.<https://www.sciencedirect.com/science/article/pii/S0268401219311181?casa_token=ezbrDSZwPR0AAAAA:DeEfWA6TfB9fPRXF1kWdGR-z888WijMn5AH_joPbH-sZWR_OMoxmJsWiAZTssHIuhPDIrkBIVA#bib0255>

Alaparthi, S., Mishra, M. BERT: a sentiment analysis odyssey. J Market Anal 9, 118–126 (2021).<https://doi.org/10.1057/s41270-021-00109-8>

Alexander Hogenboom, Ferry Boon, and Flavius Frasincar, 2016. A Statistical Approach to Star Rating Classification of Sentiment.<https://personal.eur.nl/hogenboom/files/ismis12-sentistar.pdf>

B. Jeong, J. Yoon, J.-M. Lee. Social media mining for product planning: A product opportunity mining approach based on topic modeling and sentiment analysis. International Journal of Information Management, 48 (2017), pp. 280-290

Boost Labs. (n.d.). Word Clouds & the Value of Simple Visualizations. Boost Labs. <https://boostlabs.com/what-are-word-clouds-value-simple-visualizations/>

Greco & Polli, 2019, F. Greco, A. Polli. Emotional text mining: Customer profiling in brand management (International Journal of Information Management (2019).<https://www.sciencedirect.com/science/article/pii/S0268401219311181?casa_token=ezbrDSZwPR0AAAAA:DeEfWA6TfB9fPRXF1kWdGR-z888WijMn5AH_joPbH-sZWR_OMoxmJsWiAZTssHIuhPDIrkBIVA#bib0105>

H. Maqsood, I. Mehmood, M. Maqsood, M. Yasir, S. Afzal, F. Aadil, M. Selim, K. Muhammad A local and global event sentiment based efficient stock exchange forecasting using deep learning. International Journal of Information Management, 50 (2020), pp. 432-451. <https://www.sciencedirect.com/science/article/pii/S0268401219311181?casa_token=ezbrDSZwPR0AAAAA:DeEfWA6TfB9fPRXF1kWdGR-z888WijMn5AH_joPbH-sZWR_OMoxmJsWiAZTssHIuhPDIrkBIVA#bib0170>

Hutto, C., & Gilbert, E. (2014). VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text.<https://ojs.aaai.org/index.php/ICWSM/article/view/14550>

Investopedia. (n.d.). Topic modeling: definition, benefits and use cases. [https://www.qualtrics.com/experience-management/research/topic-modeling/#:~:text=Wh at%20is%20topic%20modeling%3F,within%20a%20collection%20of%20documents.](https://www.qualtrics.com/experience-management/research/topic-modeling/#:~:text=Wh%C2%A0%20at%20is%20topic%20modeling%3F,within%20a%20collection%20of%20documents.)

Kauffmann, E., Peral, J., Gil, D., Ferrández, A., Sellers, R., & Mora, H. (2020). A framework for big data analytics in commercial social networks: A case study on sentiment analysis and fake review detection for marketing decision-making. Industrial Marketing Management, 90, 523-537.<https://doi.org/10.1016/j.indmarman.2019.08.003>

Kherwa, Pooja., Bhansal, Poonam. (2018). Topic Modeling: A Comprehensive Review.<https://www.researchgate.net/publication/334667298_Topic_Modeling_A_Comprehensive_Review>

Penedo, G., Malartic, Q., Hesslow, D., Cojocaru, R., Cappelli, A., Alobeidli, H., Pannier, B., Almazrouei, E., Launay, J. (2023, June 1). The Refined Web Dataset for Falcon LLM: Outperforming Curated Corpora with Web Data, and Web Data Only.<https://arxiv.org/pdf/2306.01116.pdf>

Plaue, Matthias. M.A.P.E.G.Y. Tech (2023) discusses large-scale language models for innovation and technology intelligence, focusing on sentiment analysis of news articles.<https://medium.com/mapegy-tech/large-scale-language-models-for-innovation-and-technology-intelligence-sentiment-analysis-on-news-2c1ed1f6f2ad>

Saha, Subrata & Showrov, Md. Imran Hossain & Rahman, Md & Majumder, Md. (2023). VADER vs. BERT: A Comparative Performance Analysis for Sentiment on Coronavirus Outbreak. 371-385. 10.1007/978-3-031-34619-4\_30.

W. Wang, W. Ning and H. Wang,. (2015). "Online negative public sentiment does not matter? — Empirical evidence from social media and movie industry," International Conference on Computing, Networking and Communications (ICNC).<https://ieeexplore.ieee.org/document/7069507/>

Zhang, Wenxuan., Deng, Yue., Liu, Bing., Jialin Pan, Sinno., & Bing, Lidong. (2023, May 24). Sentiment Analysis in the Era of Large Language Models: A Reality Check.<https://arxiv.org/pdf/2305.15005.pdf>