INSY669 Text Analytics Final Project

McDonald's stands as the most renowned and successful fast-food chain globally. In a time when customer feedback plays a pivotal role in improvement and success, our aim is to glean insights into customer priorities through text analysis. Using a dataset consisting over 33,000 anonymized reviews from McDonald's locations across the United States collected from Google reviews, we aim to bridge the gap between customer expectations and the current offerings, providing insights that could guide improvements in both product and service domains.

Our objective is to decipher customer preferences from both product-focused and store-specific standpoints. The expected outcome of our analysis is to provide McDonald's with strategic insights that can drive enhancements in customer satisfaction, potentially leading to increased loyalty, better store performance, and ultimately, greater business value.

In data preprocessing, we began by removing stopwords and rows including missing value in latitude and longitude. Since all rows with missing values were related to a store in Hawaii and the proportion of missing records was small, removing the missing values did not affect the quality of overall data. Subsequent analysis of rating distribution revealed a significant polarization in customer satisfaction, with a majority of ratings concentrated at the scale's extremes (1 or 5). We compiled reviews from 40 store locations across 11 states, focusing on the top ten with the highest volume of feedback, particularly the flagship store in Orlando, FL. To illustrate the distribution of ratings at each location. A word cloud was created to succinctly visualize the most frequent terms used in all reviews, with 'order,' 'food,' 'service,' and 'staff' being the most prevalent. Additionally, on average, the ratings hover between 3 and 4, with the state of UL registering the lowest average.

One pipeline of the analysis concentrated on sentiment analysis. Several methods were explored by integrating various tools, with accuracy scores compared against customer ratings as the benchmark for truth. To facilitate accuracy calculation, numeric ratings were initially converted into "Positive", "Negative", or "Neutral" sentiments, employing 3-star ratings as the threshold for categorization. The first and second methods employed the NLTK package for tokenization and lemmatization, and spaCy for tokenization and stopword removal, respectively. Subsequently, VADER was utilized to assign text sentiments to positive, negative, or neutral categories, additionally providing a numerical score to measure sentiment intensity. The remaining approaches engaged different Part of Speech taggings, focusing on either both adjectives and adverbs from the corpus or exclusively adjectives. The accuracy of the first two approaches was found to be remarkably similar. Although spaCy's preprocessing of the corpus yielded a slightly higher accuracy, NLTK proved to be more effective for subsequent tasks. Therefore, NLTK was selected for use throughout the analysis to maintain consistency.

To get store-specific standpoints, we made detailed store analysis based on sentiment. We first calculated the average compound score for each McDonald's store and rank accordingly. Using the specific stores' latitude and longitude for top and bottom rating stores, we drew the map of the compound sentiment score by location and found stores at either polarity located at the east side of the coast. We extract different combinations of bigrams and trigrams, and count the occurrence of these phrases. Secondly, to gain the underlying reasons for sentiment differences between top-performing and lowest-performing stores, we applied MDS to visualize cosine similarities between customer comments. In order to have a balanced number of records, we compare the top three stores with the bottom one. Although the MDS visualization did not

reveal distinct clusters due to the limitation of the data, it offered insightful visual cues into where customer experiences diverged. Interactive LDA visualization was also created based on reviews from the three top-rated stores. It provided insights into top rate store customer preferences and expectations, highlighting key factors driving satisfaction.

Topic modeling was also performed using all reviews and merged as the second pipeline of the analysis. Using LDA, we identified five distinct topics and extracted the top 20 words from each topic for arbitrary interpretation. We decided to focus on analyzing Topic 1, which seems to be food review related. We conducted the product analysis specifically on reviews falling under this topic. Firstly, we identified the mentions of specific products and examined the context of these mentions through the most frequent monogram associated with each product. Then we utilized bigrams to extract and identify descriptive phrases associated with specific products. In addition, we draw the bar charts to visualize each of the most frequently mentioned unigram and bigrams, allowing us to quantify and compare the prominence of each product and issue. Bigram provided more information with “big mac”, “chicken nugget” and “chicken sandwich” as the top 3 frequently mentioned phrases.

Based on the text analysis results, here are the impacts on business strategy and economics. Firstly, customer service was improved by helping the management team identify issues such as long wait times and staff interactions for specific stores, based on the analysis of reviews. Additionally, analyzing feedback on food products informed product development, allowing for the refinement of existing items and the innovation of new ones that met customer preferences. Location management was also enhanced through the analysis of feedback, identifying unique opportunities for each store to guide investment and marketing strategies. From an economic perspective, these strategies contributed to customer retention by improving service quality and offering tailored products, leading to repeat business and increased sales. Operational improvements, such as more efficient order processing and enhanced staff training, reduced costs and increased profitability. Moreover, insights into store performance supported strategic decisions regarding store renovations, openings, and closures, optimizing market presence and contributing to the overall success of the business.

For future steps, exploring additional LDA topics would be beneficial, specifically focusing on environmental aspects (topic 3) to enhance store cleanliness, as well as service-related issues (topic 4) to inform targeted training and improvements. Refining sentiment analysis by incorporating Peter Turney’s algorithm aims to boost accuracy in identifying customer emotions. To gain more comprehensive insights, expanding the database with an increased number of reviews and feedback from social media is planned. Additionally, enhancing dynamic map visualization tools is intended, providing McDonald's leadership with intuitive, real-time support for decision-making.

In conclusion, customer feedback analysis is essential for McDonald’s. It facilitates decision-making, enabling McDonald's to adapt to stay attuned to consumer preferences and sustain a competitive edge. Proactive feedback management enhances service and product quality, fortifying the brand's reputation and fostering customer trust. Overall, the text analysis reveals customer preferences from the perspectives of both product offerings and individual store experiences, guaranteeing McDonald's continues to be a favored brand for future generations.