Documentation: Sentiment Analysis of McDonald's Store Reviews

Seanne Cañete - ITD 111 Final Project

Dataset Description

This dataset contains over 33,000 anonymized reviews of McDonald's stores in the United States, scraped from Google Reviews. It provides valuable insights into customer experiences and opinions about various McDonald's locations across the country. The dataset includes the following key attributes:

- **Store Name**: The name of the McDonald's location.
- **Category**: Categorization of the review, such as food quality, service, cleanliness, etc.
- **Address**: Physical address of the store.
- **Geographic Coordinates**: Latitude and longitude of the store for geospatial analysis.
- **Review Ratings**: Customer ratings on a scale of 1 to 5.
- **Review Texts**: Written feedback provided by customers.
- **Timestamps**: Date and time when the reviews were posted.

Objectives

The primary objectives of this project are:

1. Analyze customer sentiments expressed in the reviews to determine whether they are positive, neutral, or negative.
2. Categorize user reviews into predefined aspects, such as food quality, service, cleanliness, and other relevant topics, to facilitate targeted insights and actionable feedback analysis.

Analysis & Methodology

Tools and Technologies

The analysis was conducted using the following tools and libraries:

- **Python**:
    - **Pandas**: For data cleaning, manipulation, and exploration.
    - **NLTK**: For text preprocessing and sentiment analysis.
    - **VADER**: To compute sentiment scores for review texts.
- **Orange Data Mining**:
    - For visualizations and advanced data mining techniques.
- **Geocoding Tools**:
    - Used to enrich data with geographic insights.

Steps Performed

1. **Data Preprocessing**:
    - Removed duplicates and irrelevant information.
    - Handled missing values.
    - Tokenized and normalized text data for analysis.
2. **Sentiment Analysis**:
    - Applied VADER sentiment analyzer to calculate polarity scores.
    - Classified reviews as positive, neutral, or negative based on thresholds.
3. **Aspect Categorization**:
    - Developed rules and models to tag reviews under categories such as food quality, service, cleanliness, etc.
4. **Geographic Analysis**:
    - Mapped reviews to store locations using latitude and longitude.
    - Analyzed state-level variations in customer experiences.
5. **Visualization**:
    - Used Orange Data Mining to create charts and graphs illustrating sentiment distributions and geographic trends.

Results and Insights

Key Findings

Overall Sentiment Analysis:
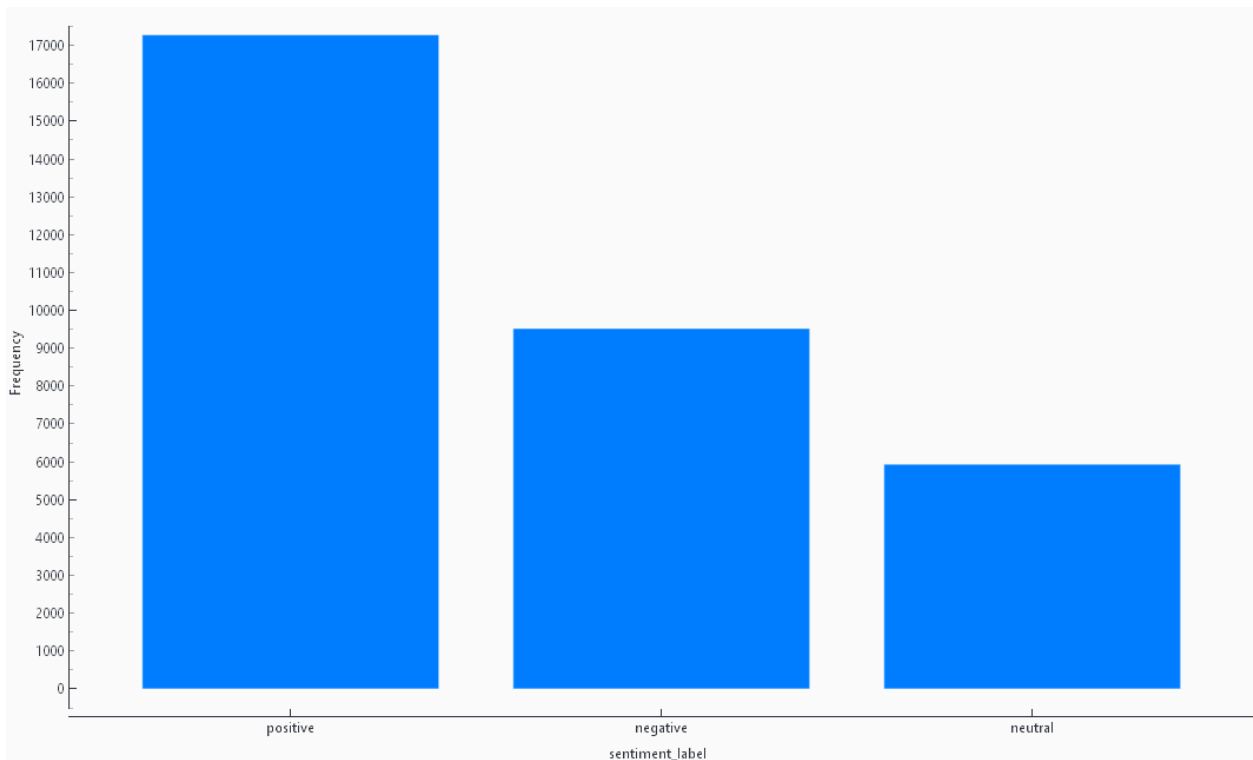
The dataset reveals that the majority of reviews for McDonald's locations across the United States are **positive**, accounting for a significant portion of the feedback. Negative reviews, while fewer, highlight specific areas of concern such as **long wait times** and **inconsistent food quality**. Neutral reviews serve as a middle ground, reflecting experiences that were neither overwhelmingly positive nor negative.

Key Findings:

- **Positive Sentiments:** Over 50% of the reviews were classified as positive, showcasing a generally favorable perception of McDonald's service and offerings.
- **Negative Sentiments:** Negative reviews emphasize operational inefficiencies like delays and occasional quality concerns, pointing to opportunities for improvement.
- **Neutral Sentiments:** Neutral reviews provide a balanced view, often commenting on average experiences without strong opinions.

Sentiment Distribution by State:
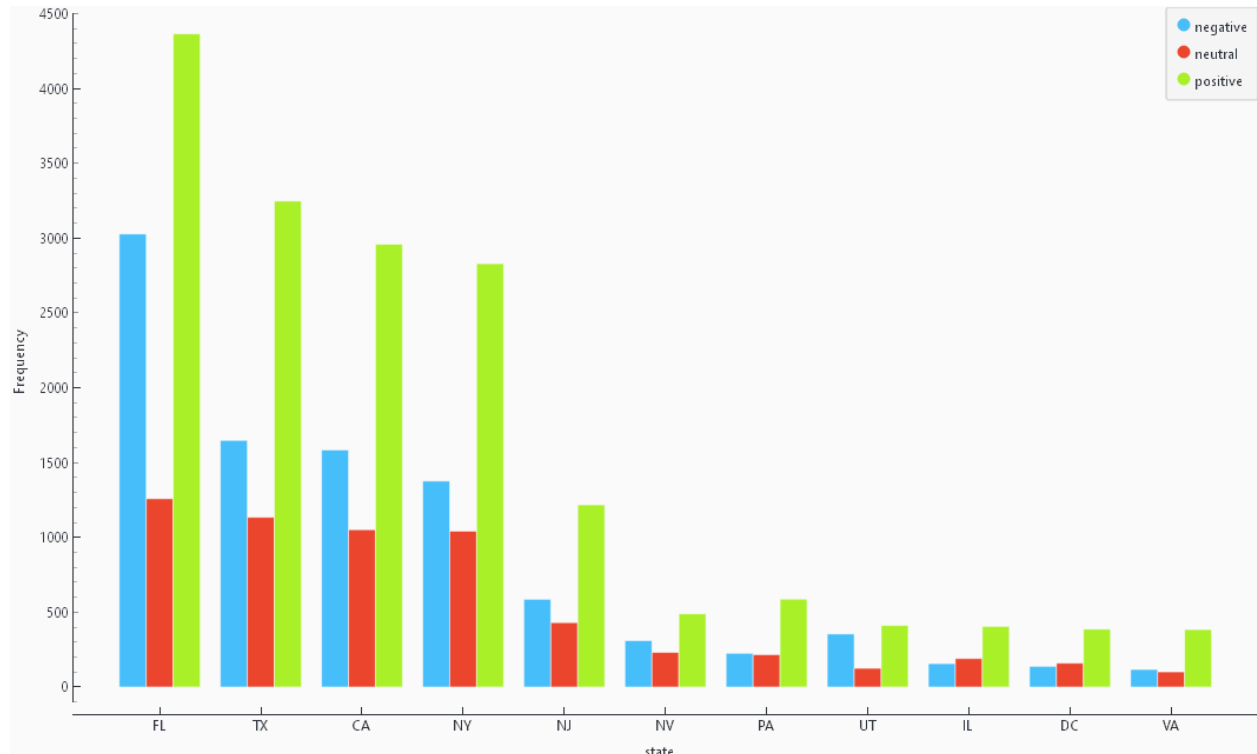
Insights from the State-Level Analysis:

- **High Positive Sentiments:** States such as **Florida (FL)**, **Texas (TX)**, and **California (CA)** demonstrate a higher proportion of positive reviews, suggesting strong customer satisfaction in these regions.
- **Negative Sentiments:** Negative reviews were more pronounced in states like **New York (NY)** and **New Jersey (NJ)**, indicating potential challenges in these areas, such as urban operational constraints or higher customer expectations.
- **Neutral Reviews:** States with a notable percentage of neutral reviews include **Nevada (NV)** and **Pennsylvania (PA)**, which may reflect variability in service quality or localized issues.

Aspect-Based Sentiment Analysis:

The sentiment analysis categorized reviews into various aspects to better understand customer feedback. The following insights were observed:

Key Findings:

1. **Food Quality**:
   - Food quality is the most frequently mentioned aspect after general comments categorized as "Other."
   - Feedback about food quality is **mixed**, with a significant share of both positive and negative sentiments. Positive reviews often praise the consistency of taste, while negative reviews point to freshness issues or inaccuracies in orders.
2. **Service**:
   - Service is the second most frequently mentioned aspect.
   - Sentiments regarding service are **mostly positive**, highlighting polite staff and efficient service. Negative reviews focus on wait times and occasional lapses in customer service.
3. **Speed**:
   - Speed is a critical factor for fast food establishments and receives mixed sentiments. While many appreciate quick service, delays during peak hours remain a common complaint.
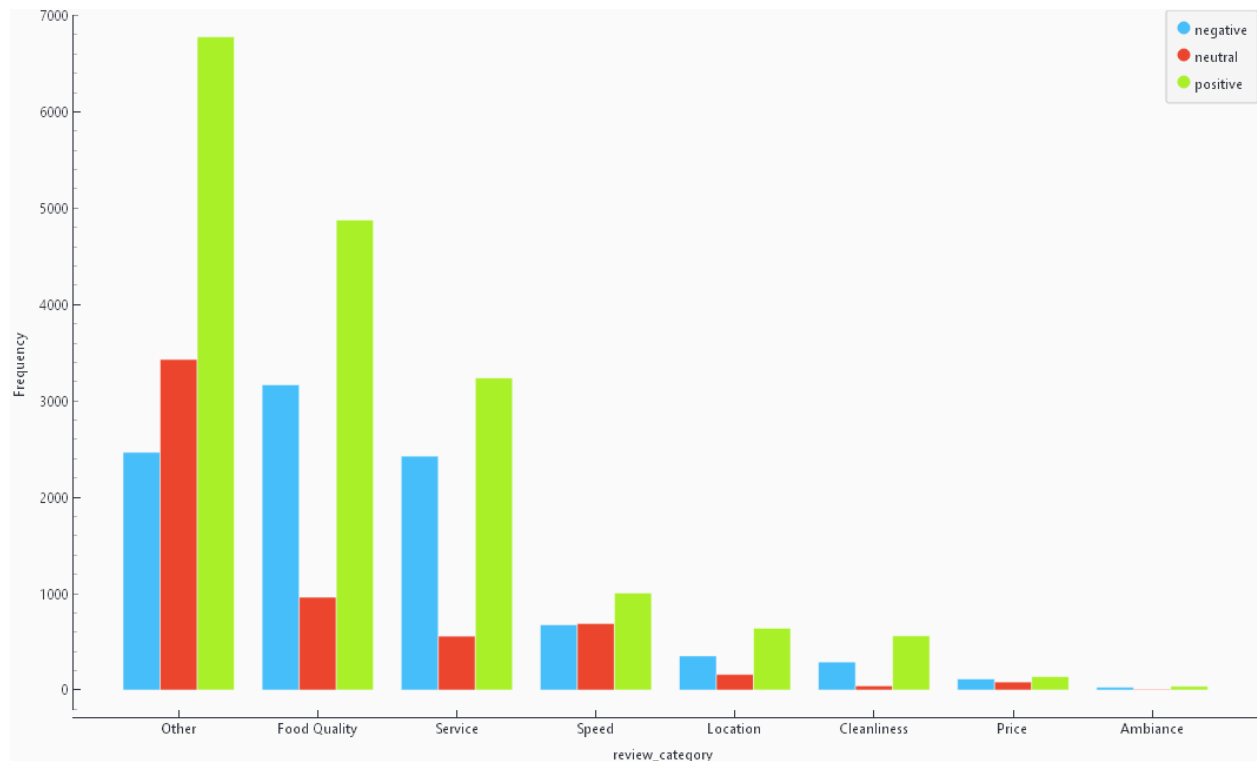4. **Cleanliness**:
   - Reviews about cleanliness are generally **positive**, with customers appreciating tidy premises. However, occasional complaints about dirty tables or restrooms were noted.
5. **Location and Ambiance**:
   - These categories are less frequently mentioned but are generally associated with **positive experiences**.

Sentiment Distribution by Review Categories:

- **Food Quality and Service** dominate the review categories, with service showing a higher proportion of positive feedback.
- Negative sentiments are more prevalent in the **Speed** and **Cleanliness** categories, indicating areas for operational improvements.
- Less frequent categories such as **Price** and **Ambiance** receive predominantly neutral or positive sentiments, reflecting customer expectations being met in these areas.

Business Implications

Targeted Improvements:

Focus on enhancing food quality and reducing service times, especially in areas with frequent complaints, to meet customer expectations and improve overall satisfaction.

Customer Feedback Systems:

Implement robust feedback loops to identify and address recurring issues promptly, ensuring a continuous improvement cycle that enhances customer experiences.

Appendix

United States' McDonald Reviews Sentimental Analysis

Feature Statistics

Mean, Median, Mode and etc.
(metadata of variables)

Feature Statistics

Data Table

Sentiment Distribution on each Category

Select Columns

Filtering of columns

Box Plot

Aspect Based Sentiment Analysis

Uses predefined categories (e.g, Food Quality, Service, Cleanliness, Speed, Price, Ambiance, Location).

Reviews are assigned to one or more categories based on keyword matching.

File

Geocoding

Coded Data → Data

Edit Domain

Distributions

Box Plot (1)

Location Based Sentiment Analysis

Location-based aspect-based sentiment analysis focuses on understanding customer feedback by analyzing sentiments about specific aspects (e.g, food quality, service, ambiance) in different locations across the United States.

Scatter Plot

Scatter Plot on Reviews Score correlation with other columns

Choropleth Map

Choropleth Map for Sentiment Scores Across the States

Each state is color-coded based on aggregated sentiment scores derived from customer reviews, providing a clear, visual representation of regional sentiment trends.