Data 620

Summer 2023

Final Project Proposal

Group Members: Khyati Naik; Laura Pueblo

Title: Analyzing Customer Sentiments towards Food Products in Amazon Fine Food Reviews

Research Questions:

- How does the sentiment towards food products change over time in Amazon Fine Food Reviews?
- How do customer sentiments differ between organic and non-organic food products based on Amazon Fine Food Reviews?

Motivation:

In this project, we aim to analyze customer sentiments towards food products based on Amazon Fine Food Reviews. Understanding how sentiments evolve over time and the differences between organic and non-organic food products can provide valuable insights into customer preferences and help businesses make informed decisions. By conducting sentiment analysis on customer reviews, we can gain a deeper understanding of customer sentiments towards food products, identify trends, and uncover potential areas for improvement.

Data Sources:

We will utilize the Amazon Fine Food Reviews dataset, which consists of customer reviews and ratings for various food products available on Amazon. The dataset includes information such as product reviews, scores, timestamps, and user demographics. This dataset provides a rich source of customer feedback that can be analyzed to understand the changing sentiments towards food products over time and the differences in sentiments between organic and non-organic food products.

Plan and Approach:

1. *Data Collection*: Download the Amazon Fine Food Reviews dataset from a reliable source, such as Kaggle or a public repository. Load the dataset into a suitable data analysis platform.
2. *Data Preprocessing*: Perform necessary data preprocessing steps, including removing duplicates, handling missing values, and cleaning the text by removing special characters, numbers, and converting to lowercase.

3. *Sentiment Analysis*: Use a sentiment analysis tool, such as VADER (Valence Aware Dictionary and Sentiment Reasoner), to calculate sentiment scores for each review. Measure the polarity and intensity of customer sentiments.
4. *Statistical Analysis*: Conduct statistical analysis to analyze the sentiment trends over time and compare the sentiments between organic and non-organic food products. Calculate average sentiment scores for different time intervals and compare the sentiment distributions between the two categories using appropriate statistical tests.
5. *Data Visualization and Results Interpretation*: Visualize the results using line charts, bar charts, or other suitable visualizations to showcase the sentiment trends over time and the differences in sentiments between organic and non-organic food products. Interpret the results and discuss the implications of the findings.

Roles and Responsibilities:

Laura Pueblo: Data collection, data preprocessing, results interpretation

Khyati Naik: Sentiment analysis, data visualization, statistical analysis, results interpretation, project presentation.

Concerns:

- The dataset may contain biased or subjective reviews that could affect the analysis results.
- The sentiment analysis approach may not capture the full context or nuances of customer sentiments.
- The dataset's size and complexity may pose challenges in terms of processing power and computational resources.

By analyzing customer sentiments towards food products in Amazon Fine Food Reviews, we aim to uncover valuable insights into customer preferences and sentiments. This analysis can assist consumers in making informed choices and support businesses in product development, marketing strategies, and addressing customer concerns related to organic and non-organic food products.