Proposal: Impact of Online Reviews on Restaurant Foot Traffic in Los Angeles

Zixi Wang (ID: 2854187591)

Chenyi Weng (ID: 3769237784)

DSCI510 - Principle of Programming for Data Science

Professor Amandeep Singh

Nov 9, 2024

1. Problem Statement

What problem are you trying to solve?

The question this project seeks to solve is to understand how the evolution of Los Angeles restaurant service models (dine-in, takeout, delivery) after the pandemic affects customer behavior related to online reviews. Specifically,

- **Service Model Evolution:** To explore how different service options (e.g., dine-in, takeout, delivery) have evolved in Los Angeles restaurants during the post-pandemic era.
- **Impact of Online Reviews:** Analyze how online reviews (e.g., Google ratings, Yelp reviews) influence restaurant foot traffic and sales, specifically focusing on the effects of positive and negative reviews.
- **Service-Specific Insights:** The focus is on understanding the relationship between ratings for dine-in, takeout, and delivery services and changes in customer traffic.

We aim to uncover actionable insights that help restaurants optimize their service models and use online reviews to drive customer satisfaction and engagement.

2. Data Collection Methods and Sources

- **Yelp API and Google Places API (Purpose: Collect detailed information on ratings, reviews, and service types of restaurants in Los Angeles.)**
    - **Data Types:**
        i. Collect ratings, review counts, review content, and service types (takeout, delivery, dine-in) for restaurants in Los Angeles.
        ii. Restaurant name, ratings, number of reviews, review content, restaurant type (e.g., Chinese, Western, fast food), and service options provided.
    - **Method:**
        i. Retrieve data about restaurants using Yelp API and Google Places API.
        ii. Leverage Natural Language Processing (NLP) tools to analyze review content, extract customer sentiment, and identify service-specific feedback.
- **Food Delivery Platform APIs (e.g., UberEats or DoorDash)(Purpose: Collect delivery-specific data to analyze post-pandemic trends and customer behavior)**
    - **Data Types:**
        i. Collect data on delivery orders, delivery times, and order prices for Los Angeles restaurants, Estimated preparation and delivery times, popular menu items
    - **Methods:**
        i. Web scraping using tools like BeautifulSoup for website data

    ii. API requests to delivery platforms to extract details (such as availability, pricing, and reviews)
- **SafeGraph or Placer.ai for foot traffic data (Purpose: Analyze restaurant traffic data and compare it with online review patterns.)**
    - **Data Types:**
        i. Number of restaurant visitors, Visitors' flow trends
    - **Methods:**
        i. Access traffic data through subscription platforms
        ii. Collect data using web scraping or API integration

3. Data Cleaning and Processing

- **Data Cleaning**
    - **Remove Invalid Reviews**: Filter out advertisements, duplicate reviews, and irrelevant content to ensure high data quality.
    - **Merge Data from Different Sources**: Combine Yelp and delivery platform data, match the same restaurant's records, and standardize naming conventions.
    - **Convert Categorical Data to Numeric:** such as Convert service types (e.g., "dine-in," "takeout," "delivery") into numeric codes. Or apply on-hot encoding for multi-class features such as restaurant type.
    - **Handle Missing Values:** For numeric data, fill missing values using the mean, median, or interpolation techniques.
- **Data Transformation and Processing**
    - Use NLP (Natural Language Processing) techniques to label the sentiment of reviews (positive, negative, neutral) and extract key phrases from the reviews.
    - Standardize timestamps to facilitate time series analysis, especially to observe customer traffic changes during different phases (e.g., post-lockdown).

4. Data Analysis

Correlation Analysis Between Reviews and Customer Traffic

Perform regression analysis to study the relationship between restaurant ratings and changes in customer traffic, specifically whether high ratings are correlated with increased foot traffic and delivery orders.

Separate ratings for different service types (dine-in, takeout, delivery) to analyze their individual impact on customer traffic.

Sentiment Analysis

Use NLP to analyze the sentiment of reviews, quantify positive and negative sentiments, and explore the relationship between sentiment, order quantity, and restaurant ratings.

5. Data Visualization

- **Time Series Visualization**
  - Use line charts to show changes in customer traffic at different industry phases (e.g., lockdown, reopening), providing a clear picture of the pandemic's impact on restaurants.
- **Sentiment Analysis Visualization**
  - Create word clouds to show key phrases from reviews, reflecting customer focus points regarding restaurant services.
  - Use bar charts to display sentiment distribution across different service types, such as comparing the proportion of positive sentiment between delivery and dine-in services.
- **Correlation Between Ratings and Customer Traffic**
  - Use scatter plots to visualize the relationship between restaurant ratings and customer traffic or order numbers, observing if a significant positive correlation exists.
- **Impact Analysis of Delivery Performance‘**
  - Use box-plot to compare customer satisfaction scores within different delivery time ranges

6. Conclusion and Future Directions

Conclusion

The analysis will help understand the development trends of the restaurant industry in Los Angeles during the post-pandemic period, providing insights for restaurant owners to adjust strategies based on customer feedback to attract more traffic.

Future Directions

Expand the analysis to other cities to test the generalizability of the findings.

Explore other factors that impact customer traffic, such as social media promotions or restaurant marketing campaigns.