Project Progress Report

Havish Rani: hrani2, Acchindra Thev: Athev2, Nishit Samarth samarth9, Harsha Gajja: hgajja2, Vinayak Manepalli: vmanep2

Sentiment Analysis of Restaurants in Champaign and Chicago

Introduction
Our project focuses on analyzing the sentiments of restaurant reviews in a college town (Champaign) and a large city (Chicago) using Python to scrape Reddit subreddits. We aim to understand public opinions and regional variations in sentiments towards restaurants in these two areas. This analysis is crucial for providing insights to restaurant owners, tourists, and residents and combines web scraping, natural language processing (NLP), and data analytics.

Tasks Completed

1. Web Scraping: Successfully utilized the Reddit API and Python to scrape posts and comments from restaurant-related subreddits in both Champaign and Chicago. This task involved understanding the API and developing scripts to efficiently gather data while adhering to ethical guidelines, as well as figuring out the proper subreddit to use per each city.

2. Sentiment Analysis Implemented sentiment analysis using Python libraries such as NLTK and VADER. This involved processing the scraped text data to determine sentiment scores, allowing us to gauge the overall sentiment (positive, negative, neutral) towards the restaurants in the two regions.

Pending Tasks

1. Data Preprocessing and Management: The next step involves cleaning and structuring the scraped data for analysis. This will include removing irrelevant or redundant information and organizing the data into a usable format.

2. Data Analysis and Comparison: After preprocessing, we need to analyze and compare the sentiment scores between Champaign and Chicago to identify significant trends and differences.

3.Data Visualization: Creating visual representations of our findings is crucial for effectively communicating the results. This will involve using tools and libraries for graphical representation of data.

4.Project Evaluation: We will evaluate the project based on its ability to compare sentiment scores effectively, generate clear visualizations, and assess the impact on decision-making for restaurant owners. Adherence to ethical guidelines will also be a key evaluation metric.

Challenges Faced

- Data Complexity: The scraped data from Reddit may contain complex and varied expressions of sentiment, posing challenges in accurately determining sentiment scores. So in large part figuring out an effective means to standardize the data, so make it easily usable is a huge part as well.
- Ethical Considerations: Ensuring the ethical use of data, especially in terms of privacy and adherence to Reddit's terms of service, is a constant consideration.
- Framework Learning and Implementation: A significant part of the project involves learning and effectively implementing the appropriate frameworks and tools. This is crucial for the successful execution of tasks like data scraping, sentiment analysis, and visualization. Managing time to accommodate this learning curve is a key challenge.
- Time Management: Balancing the extensive workload within the project timeline is challenging, considering the depth of analysis required in the subsequent stages.

Conclusion

So far, our project is progressing well with the completion of web scraping and the initial stage of sentiment analysis. We are on track to move into the more advanced stages of data preprocessing, analysis, comparison, and visualization. The challenges faced, including mastering new frameworks and effective time management, are being addressed proactively to ensure the successful completion of the project.