Time Series

Scholar

One of the methods I explored is the Prophet model, a popular tool for forecasting time series data that accounts for seasonality, holidays, and trends, which I thought would be helpful for studying consumers' cultural consumption behavior behind just seeking for food. Prophet has been extensively used in various fields, including financial forecasting, marketing, and consumer behavior analysis.

In recent literature, Prophet has been applied to improve prediction accuracy in contexts where time series data exhibits seasonal patterns. A paper by Chen et al. (2021) used Prophet to predict user engagement in social media, modeling weekly and monthly cycles while accounting for special events like product launches and seasonal variations. This method's ability to incorporate holidays as special events allowed the authors to capture fluctuations in engagement tied to these occurrences, which can be directly applied to my project where sentiment in reviews may change around cultural events like Chinese New Year or Japanese holidays. Luo and Xu (2021) also explored sentiment analysis of restaurant reviews using deep learning models, such as Bidirectional LSTMs. While their work is more advanced in modeling sentiment directly, it emphasizes how time series tools like Prophet can complement qualitative methods to uncover patterns over time. Together, these papers show how Prophet can integrate sentiment patterns with contextual factors like holidays.

Process

For testing Prophet, I grouped restaurant reviews by month and calculated the average sentiment and ratings for Japanese and Chinese restaurants across all years. Prophet was used to analyze recurring monthly trends and forecast future patterns. Initially, I configured the model with yearly, weekly, and holiday seasonality to capture recurring patterns, but the dataset's seasonal variations were not fully captured. Sentiment peaks for certain cuisines, such as those around Chinese New Year, highlighted this limitation.

I faced significant challenges during installation due to compatibility issues, which delayed progress. Prophet failed to install using pip, conda, or other sources, forcing me to explore alternative configurations. After resolving this on Midway 3, I was able to run the analysis but encountered inconsistencies caused by missing data and outliers. Prophet's robustness helped address uneven review distributions, but balancing review numbers revealed another issue—the importance of accounting for the total number of restaurants, as my scraping method could not capture every establishment.

Despite these challenges, I attempted to use Prophet to analyze monthly trends by aggregating data across all years. The model provided smoother forecasted lines, but several issues arose, including erratic observed data due to noise and insufficient smoothing. This misalignment between observed and forecasted trends indicated a need for more advanced preprocessing and handling of seasonal fluctuations.

One of the biggest challenges I faced in the analysis was the erratic behavior of the observed data, which I think was due to insufficient smoothing and noise in the aggregated dataset. Grouping data by month without considering how it varies across different years seemed to make inconsistencies stand out even more. For example, I saw strange fluctuations in sentiment and ratings for certain months that didn't seem to follow any clear pattern. I spent a lot of time debugging this issue by printing out the data I passed to Prophet and checking that the datasets were valid. Most of the time, they were, with plenty of rows. But even then, Prophet would either say there weren't enough data points or would fail entirely if I skipped handling missing values. I tried many different ways to fix this, but nothing worked, and I was left with data errors I couldn't fully understand.

Finding

For Chinese restaurants, I did notice possible seasonal patterns in sentiment and ratings, with peaks during January or February, which might relate to Chinese New Year. This was interesting because it shows how cultural events could influence consumer behavior, especially in how people perceive and engage with these restaurants during specific times of the year. Prophet was helpful in capturing some trends despite the noise, but the difference between observed and forecasted results made it clear that better preprocessing is needed. If I had applied rolling averages or been more careful with handling outliers, the trends might have been easier to interpret, and the forecasts might have aligned better with what the data was showing.

For future research, pairing sentiment trend and rating trends with geographical data, using latitude and longitude information can add a valuable dimension to the analysis, particularly to investigate if trends are more prominent in culturally significant locations like Chinatown. These areas could be helpful investigating cultural hunting as a reason behind the discrepancy of review sentiment, where consumers engage more deeply with the cuisine due to its perceived authenticity or cultural relevance.