DO YOU REALLY KNOW WHAT YOUR CUSTOMERS ARE TALKING ABOUT??

Yelp Review Analysis

Oct 03, 2024

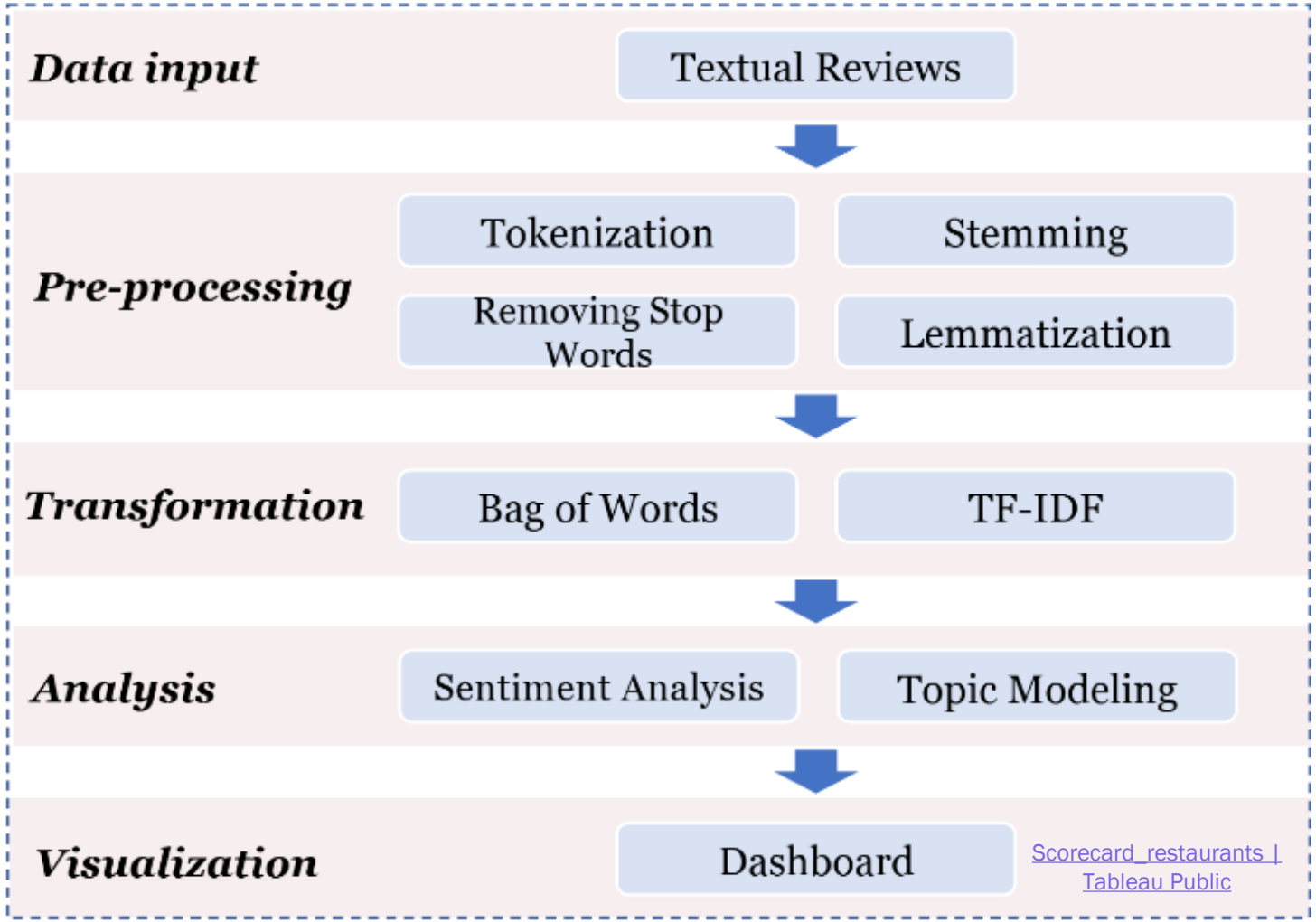
By Rashmi Chauhan

Flatiron School Capstone Project

Agenda

- Process Flow
- LDA Model for Topic Analysis
- Sentiment Analysis with VADER
- Conclusions from Analysis
- Future Enhancements
- References

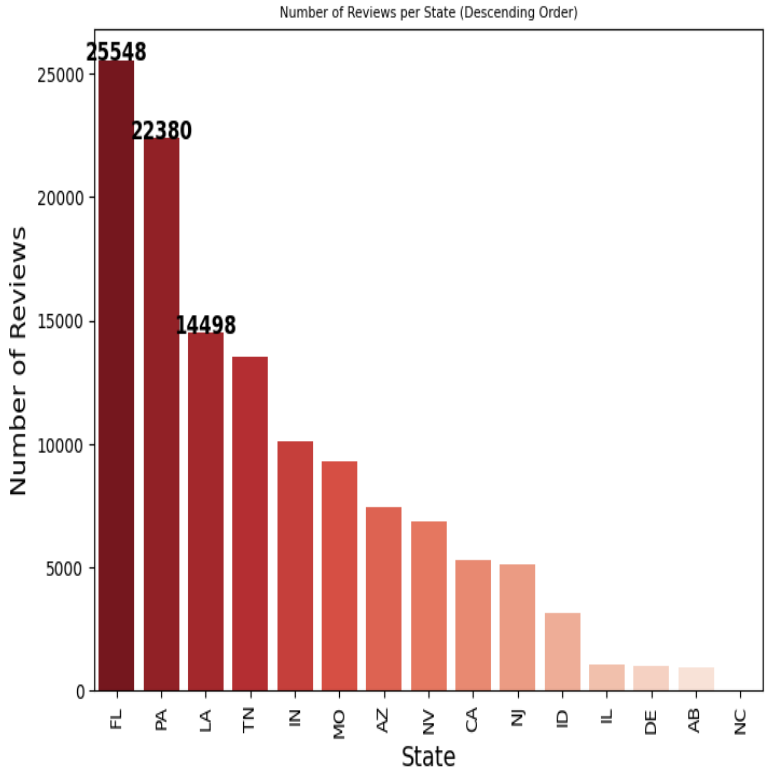
Process Flow



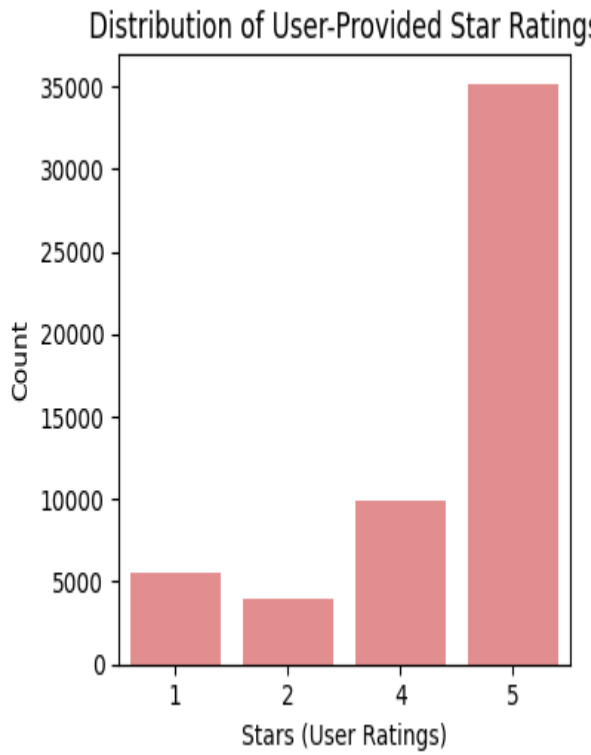
Scorecard_restaurants | Tableau Public

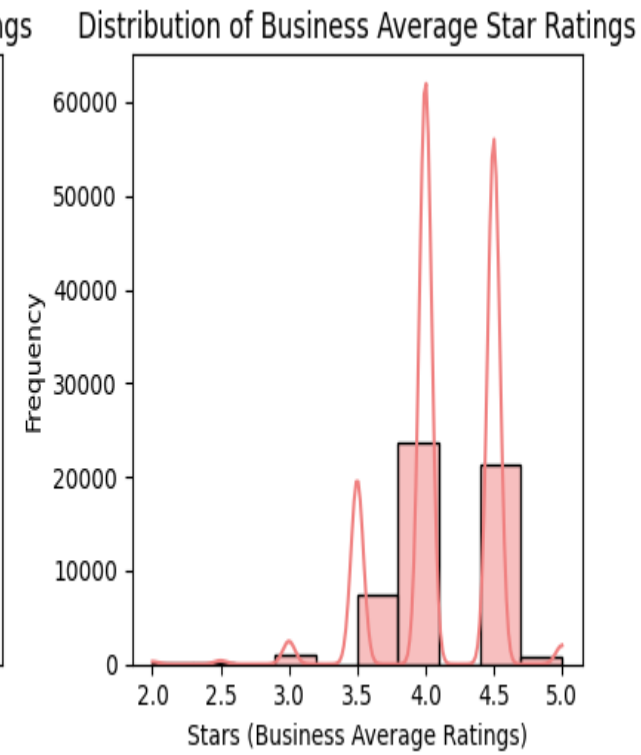
About Dataset

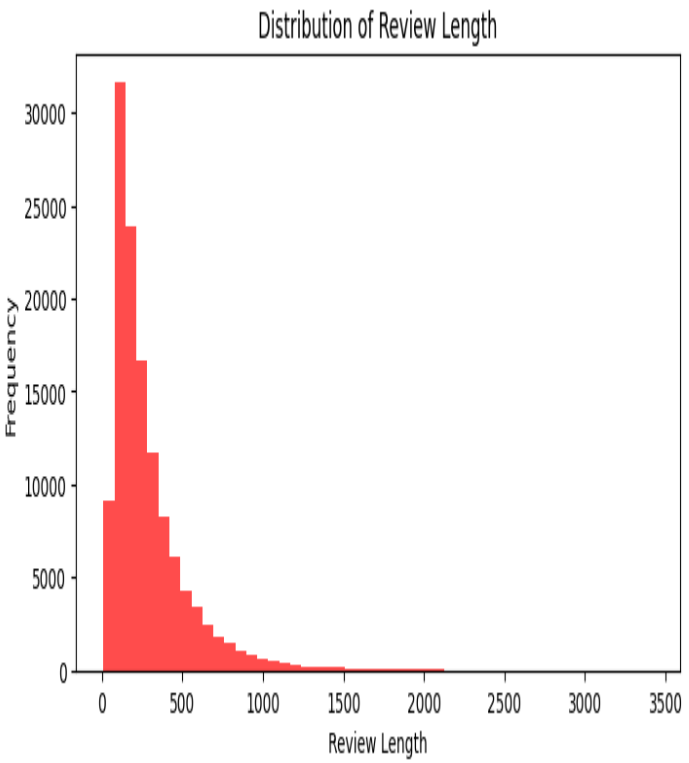
- The dataset for this project is taken form Yelp's open dataset. It has 6,990,280 reviews about over 150,346 businesses. Restaurants account for the most number of reviews on the site.
- Each row in the reviews dataframe consists of reviews data: business_id of the restaurant reviewed, star rating granted, number of [useful, funny, cool] votes, review text, and date.
- Each row in the business dataframe consists of business data: business_id, business name, state, overall star rating, number of reviews, categories, and business status (open or closed).
- The data is further filtered to show analysis on the restaurant business in Florida state, and the time period is from 2020 to 2024. The business with minimum 100 revies is taken for text analysis.



Exploratory Data Analysis







Insight from Exploratory & Inferential Analytics

- The distribution of user-provided star ratings shows that the most common ratings are on the higher end (4 and 5 stars).
- There is no clear linear relationship between the number of reviews and the average business rating.
- The most common categories are American, Bars, Seafood, and Breakfast & Brunch.
- Reviews that are either very positive (5 stars) or very negative (1 star) tend to be longer, indicating that customers may feel more inclined to elaborate when they have strong opinions.
- There is a noticeable increase in the number of reviews over time, especially in the last few years. This could indicate that the platform or business has grown in popularity.
- The noticeable dips in review patterns is during the pandemic period, potentially reflecting shifts in consumer behavior.
- The higher frequency of 4 and 5-star reviews could indicate
  reviewer bias, where people who are more likely to leave reviews tend to have positive experiences.

Sentiment Analysis with LDA & VADER

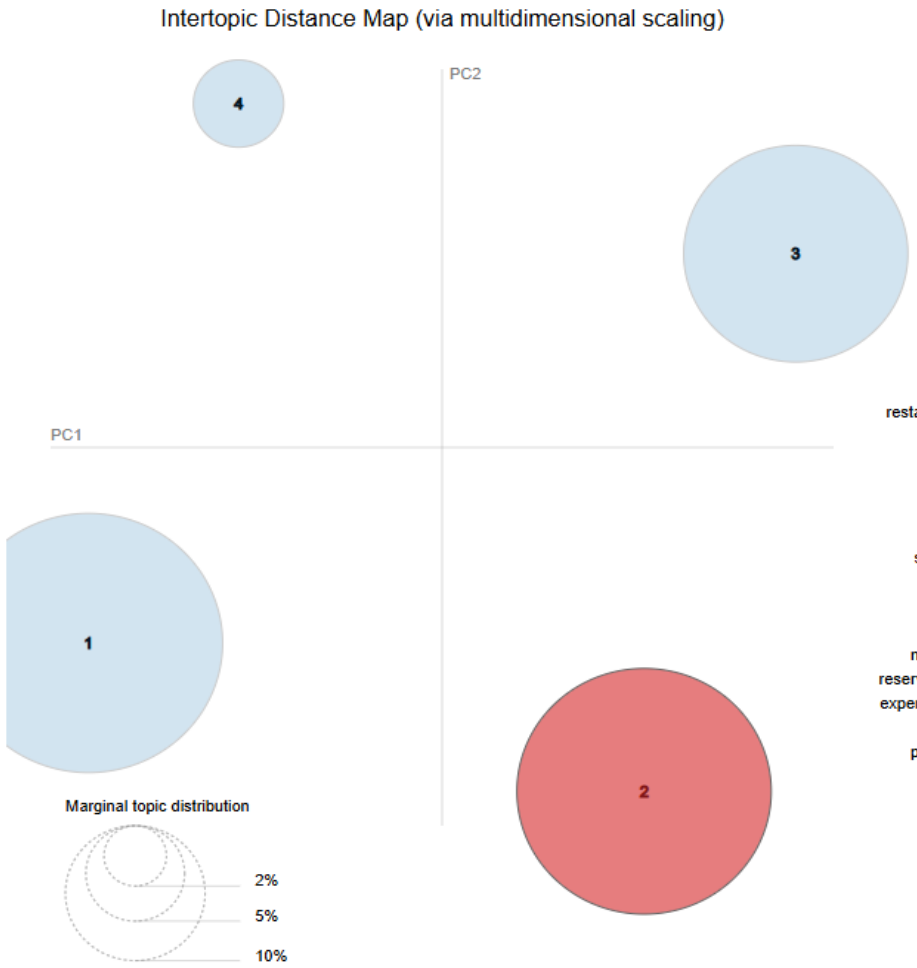
- This code performs topic modeling and sentiment analysis on the reviews text data.

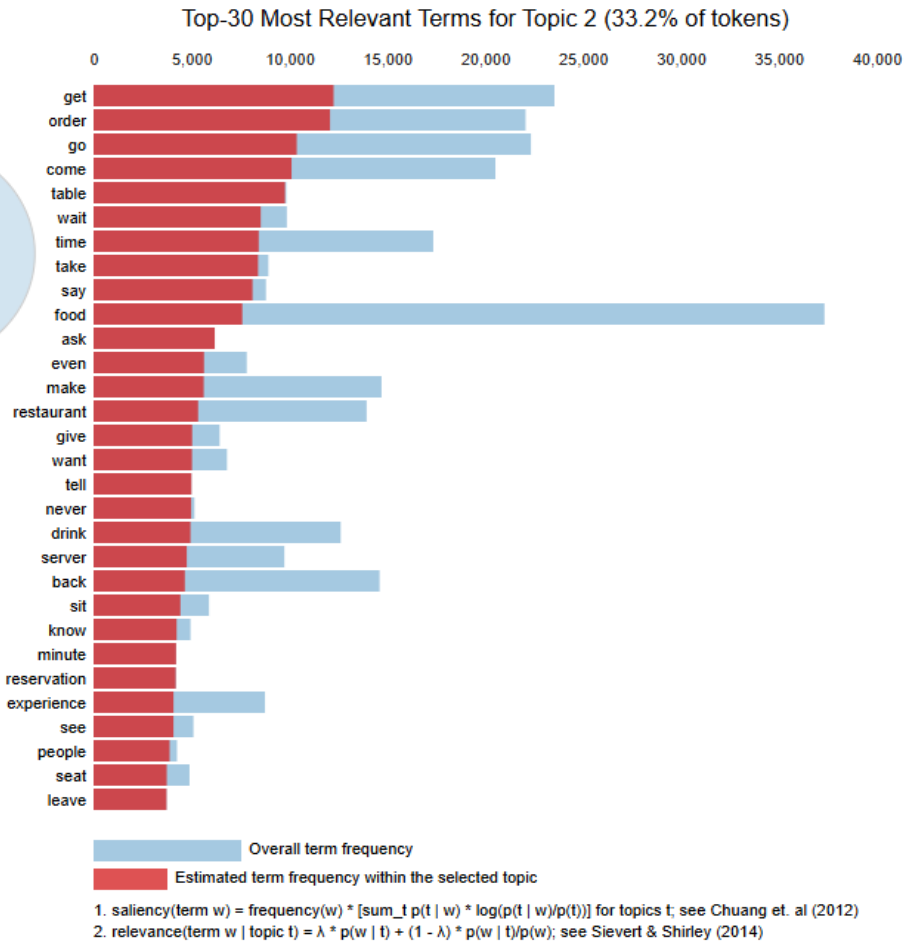
- It identifies the dominant topic of each review (e.g., food, serive, ambience).

- It computes the sentiment score for each review using VADER (positive, negative, or neutral).

- The results are stored in a new DataFrame column (Topic_Sentiment) for further analysis, such as visualizing how sentiment varies by topic.

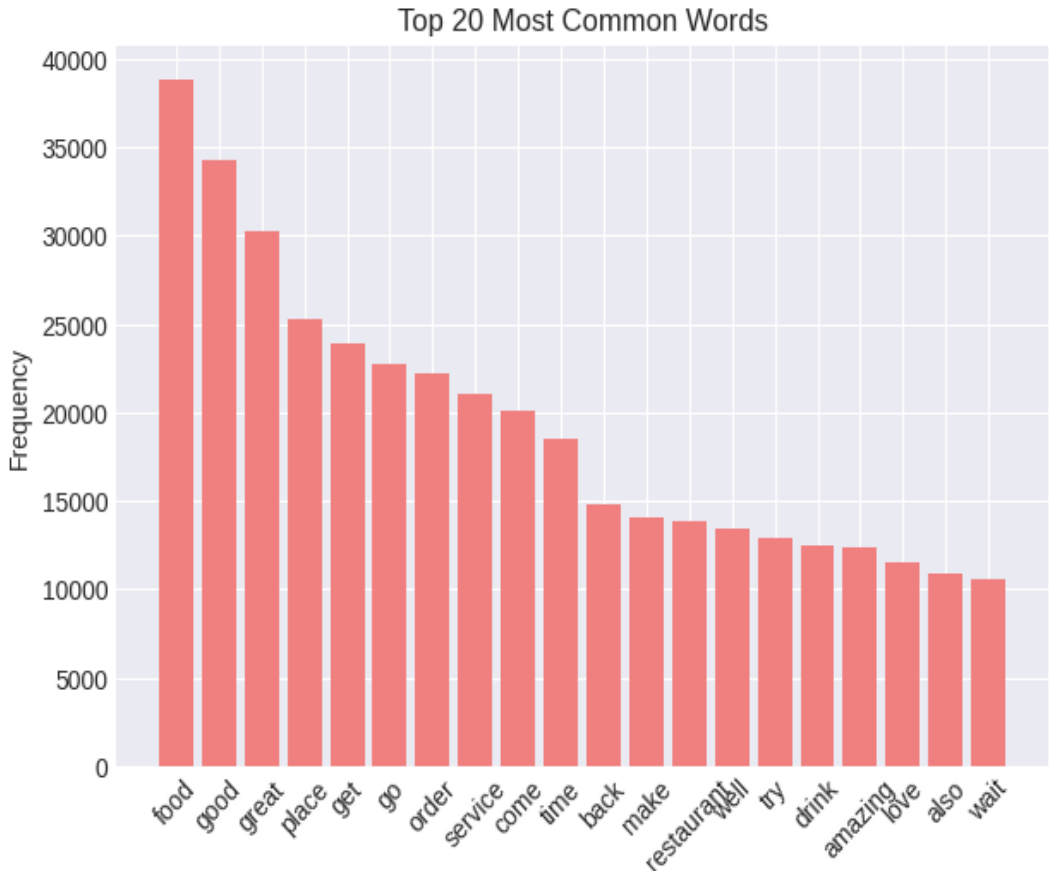
Topic/Aspect Modeling with LDA

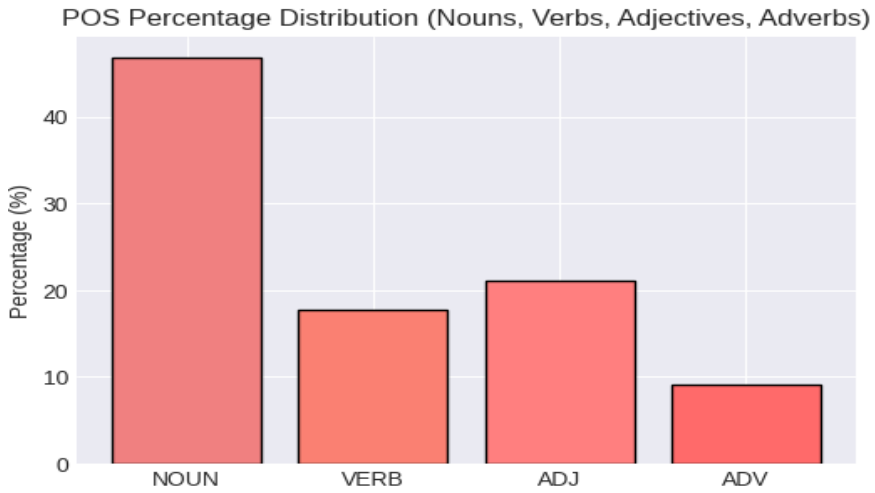


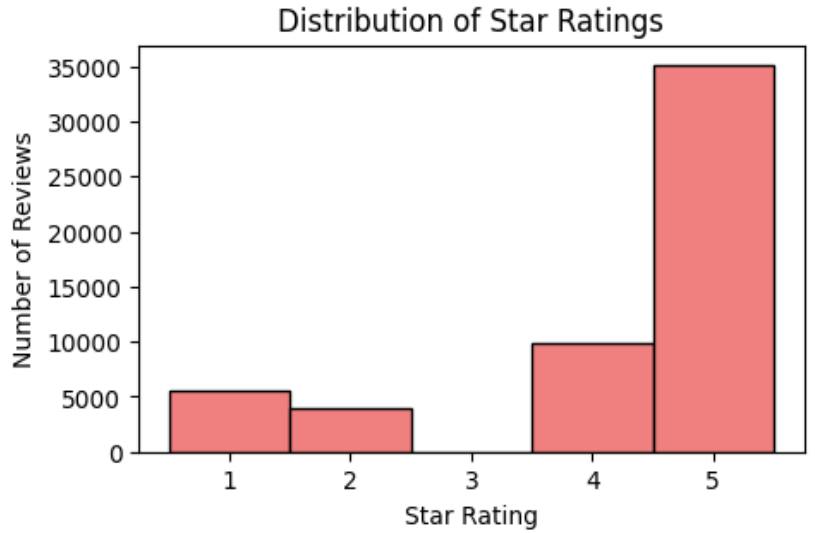


Topics from LDA: (1) Food Quality & Taste (2) Order & Dining, (3) Overall Experience (4) Ambiance & Location

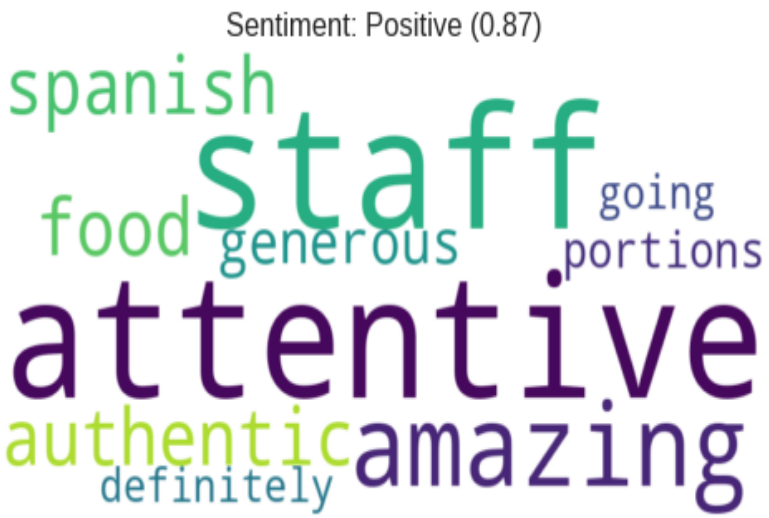
Text Analysis

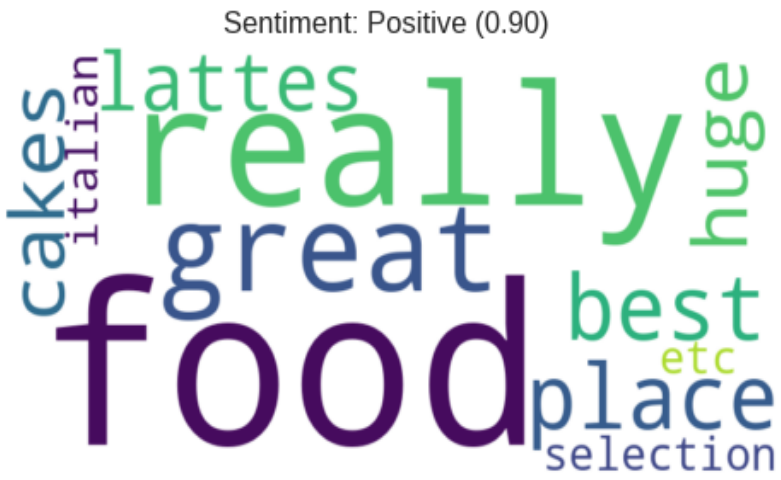






Most frequent words in reviews





Inference from Text Analytics

1. The word "food" is the most frequent, indicating that the primary reason people are reviewing the restaurant revolves around the quality and experience of the food itself. This suggests that food quality is the dominant factor that influences customer reviews.
2. Words like "good," "great," "amazing," and "love" indicate that a significant portion of the reviews contains positive sentiment.
3. The word "wait" also appears in the top 20, indicating that wait times are a frequent topic of discussion. It's an operational area that need attention.
4. The presence of the word "back" suggests that customers often mention their desire to return or share their previous dining experiences. This indicates a level of customer loyalty, where diners are willing to come back if they had a positive experience.
5. The words "service" and "place" appear frequently, this suggests that customers are not only focused on the food but also place importance on how they are treated by staff and the ambiance of the location.
6. Neutral sentiment may be a signal of early-stage dissatisfaction, giving the restaurant an opportunity to address concerns before they lead to negative reviews.

Conclusion

- From the sentiment analysis, it is evident that restaurants should prioritize delivering exceptional food while minimizing negative customer service experiences.

- Interestingly, being located in a popular area doesn't necessarily guarantee higher ratings. For aspiring restaurant owners, it is crucial to carefully evaluate the category they choose to enter, as this has a significant impact on the ratings they are likely to receive.

- By focusing on delivering great food, improving service, and creating a balanced overall experience, restaurants can enhance customer satisfaction and achieve better ratings.

Future Work

- In continuation of this project, the use of more deep learning NLP methods, like LSTM and BERT on the reviews would be helpful in extracting features from the language data.

- Geospatial Analysis can be done to review trends and sentiments based on geographic locations to identify patterns.

References

- Sentiment Analysis Using VADER - Analytics Vidhya
- ijaerv13n21_28.pdf (ripublication.com)
- YelpDatasetChallengeWinner_ImprovingRestaurants.pdf
- blei03a.dvi (jmlr.org)
- 2004.13851v1 (arxiv.org)

Thank You ☺