1 Introduction/Motivation: The landscape of breakfast and brunch establishments have undergone a significant transformation to diversify aspects of their operations to stay competitive in the rapidly growing restaurant industry, given that breakfast and brunch places have become the fastest-growing mealtime segment [1]. Our report focuses on analyzing factors that influence customer satisfaction in breakfast and brunch restaurants in Indianapolis due to its abundance of data. Our analysis aims to understand the impact of geographic location, operating hours, and specific menu items on Yelp average ratings. On a smaller scale we focus on other attributes such as pet policies, child-friendliness, smoking, and ambience. We hypothesize that a combination of these factors affects customer ratings and can be leveraged by restaurant owners to improve their service and popularity.

2 Background Information/Data Processing: We collected our data from two main sources, Yelp (business.json, review.json) and Trips by Distance.csv to examine the relationship between location and the average ratings of breakfast and brunch places in Indianapolis. For the 'business.json' dataset, we excluded closed restaurants and those not categorized as 'Breakfast & Brunch.' We used one-hot encoding for business attributes, where missing attributes, and not present were represented as 0's and present attributes as 1.

For the review.json our analysis concentrated on businesses with over 124 reviews, representing the 25th percentile of all reviews per business. This yielded a dataset of 41,300 reviews across 337 breakfast and brunch businesses in Indianapolis, totaling around 2.8 million words. To analyze attitudes regarding business attributes and food items, we used advanced Natural Language Processing methods like BERT and TextBlob on the 'reviews.json' data. Additionally, we excluded reviews that were too brief (e.g., consisting of only 2 words) or excessively long (e.g., 619 words) to ensure the quality and relevance of the textual data.

To explore the connection between location and average ratings, we integrated data from 'trips.csv' into our Yelp datasets. This integration was based on dates, zip codes, business IDs, and counties. This comprehensive approach allowed us to uncover insights into how various factors impact restaurant ratings in this specific category.

3 Methods & Discussion:

3.1 Geographic Analysis: Our geographic analysis of breakfast/brunch restaurants in Indianapolis utilized K-means clustering to explore how location impacts Yelp ratings. By plotting these establishments on a map, we observed distinct clusters forming, leading us to apply a K=10 clustering solution based on the Elbow Plot method. We categorized restaurants by zip code, ensuring that each cluster was distinct based on the average number of trips in that area. In which red and orange clusters had more average trips, but lower ratings in comparison to the blue and green clusters that had higher ratings and were more centered towards the central city. This is possibly due to the comprehensive services available in central areas, reducing the need for extensive travel. Conversely, some popular places with frequent visits had lower ratings, hinting at diverse customer tastes and potentially high expectations, leading to lower ratings if not met. This nuanced understanding of different clusters and their unique preferences provides valuable insights for restaurant owners looking to optimize their appeal and success in the Indianapolis market.

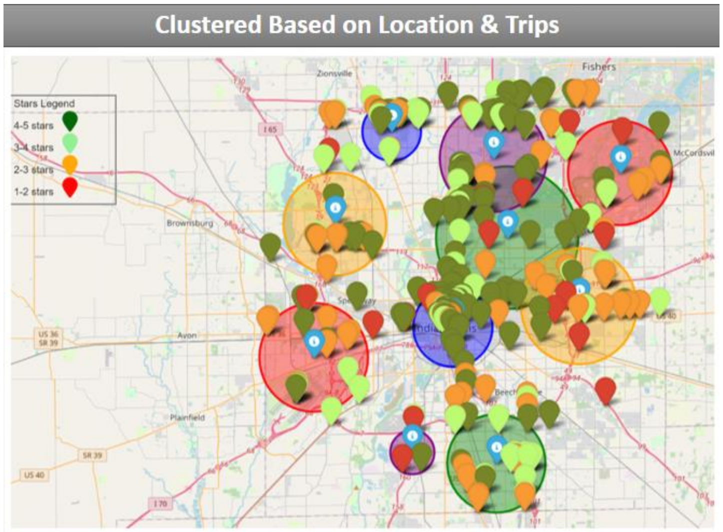

Figure 1: Geographic Clusters Based on Location and Number of Trips

3.2 Operating Hours Analysis: Operating hours significantly influence customer satisfaction and Yelp ratings for breakfast and brunch restaurants in Indianapolis. Our analysis highlights two key findings, restaurants operating from 7:00 AM to 2:00 PM receive higher average ratings, indicating a preference for places focusing on quality morning and early afternoon services. In addition, establishments operating around the clock tend to have lower ratings, possibly due to challenges in maintaining consistent quality. Other factors like location, menu, and ambience also affect customer satisfaction. A tested correlation between total operating hours and average ratings showed a weak correlation (Coefficient: 0.091, P-value: 0.67), suggesting other factors also play a role in ratings.

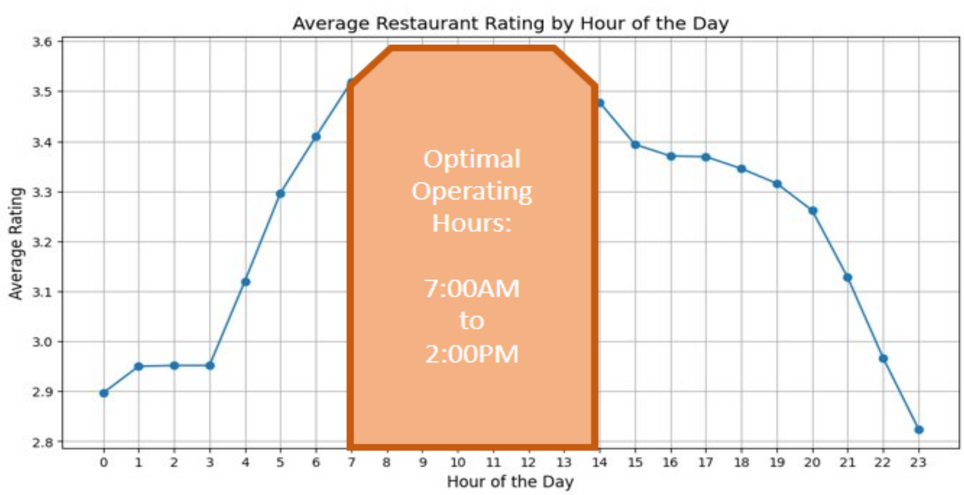

Figure 2: Average Ratings for Restaurants Open at Each Hour of the Day

3.3 Food Menu Items Analysis:

We created a bar plot analyzing the top 40 breakfast and brunch foods from a list of 100, focusing on their association with average ratings. Ratings above 3.5 were deemed positive, 3 neutral, and below 3 negatives. Despite a weak correlation coefficient (0.2) and insignificant p-value, these findings are noteworthy. Our sentiment analysis model, using a pre-trained BERT, showed limitations; for instance, it only assigned sentiment to 300 out of 3000 mentions of 'pancakes'. However, in cases where over 60% of sentiment data was captured, a clearer relationship between average rating and sentiment emerged. This gap indicates a potential limitation of our model in capturing the complete sentiment picture which our nuances in pre-trained models. Overall, the trend suggests that positive sentiment correlates with higher average ratings, even if the sentiment capture wasn't significant. There are foods that had opposite impacts regardless of sentiment which is because they are not popular foods.

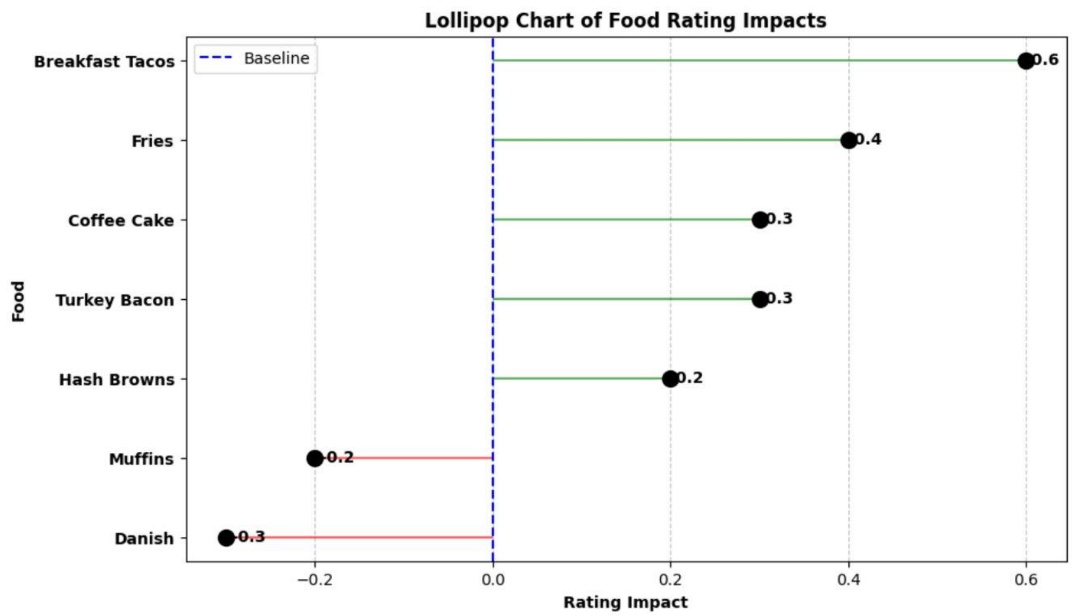

Figure 3: The Impact of Sentiments on Restaurants Ratings by Top Food Items

3.3 Other Attributes: Inspired by the business attributes columns we decided to conduct sentiment analysis to see if when mentioned in the text they were associated with average ratings. The sentiment was conducted by finding a dictionary of synonyms to the features, parsing through the text to see words mentioned before and after it, and taking the most common contexts and splitting them into 2 groups. We focused on the ones with the highest number of mentions in the text. For example, when pets were allowed the average rating decreased by 0.6 points. A weakness of this approach is it does not account for other sentiments in the text related to other external factors. However, the t-test based off the attributes columns also showed significance.

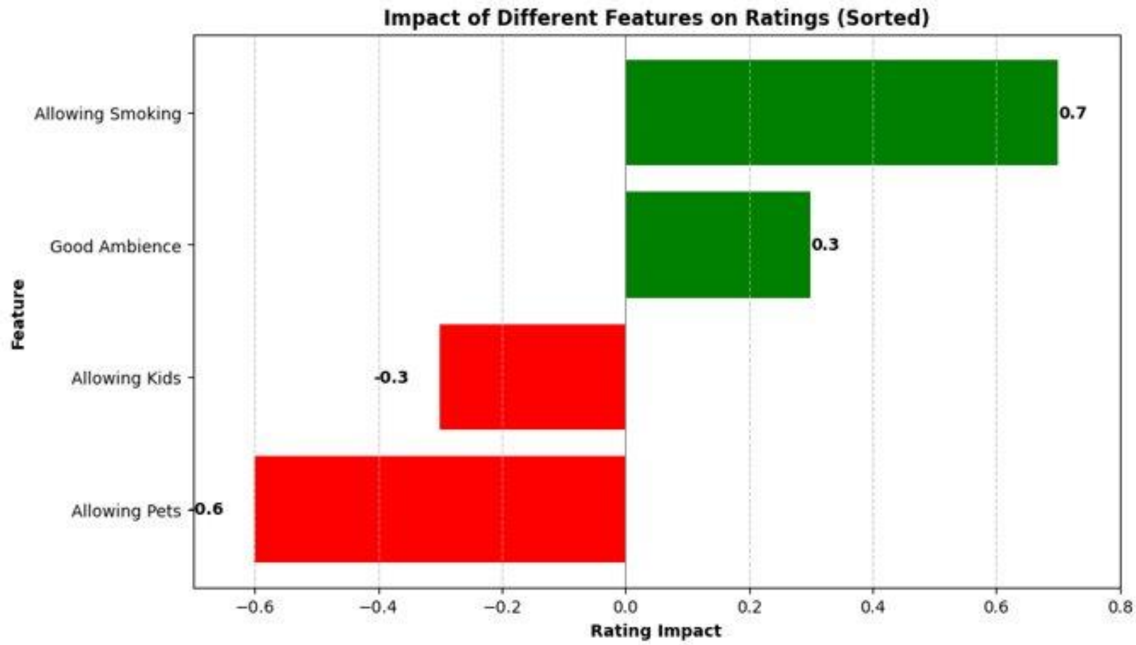

Figure 4: The Impact of Sentiments on Restaurants Ratings by Other Restaurant Attributes/Features

Conclusion: In conclusion, our analysis offers actionable insights for breakfast and brunch restaurant owners in Indianapolis to enhance their Yelp ratings and customer satisfaction. Key findings highlight the importance of strategic location selection, particularly in high foot traffic areas, which tends to attract more positive reviews. Additionally, optimizing operating hours, with a preference for 7:00 AM to 2:00 PM, aligns with customer preferences and generally results in better reviews. The influence of menu items is also notable, with certain items like scones and breakfast tacos frequently mentioned in positive reviews, while items like muffins and danish are less favored. Other attributes, such as allowing smoking and providing a good ambiance, positively impact customer experiences, whereas permitting kids and pets tends to be viewed less favorably. Despite some limitations, including low correlation coefficients, these findings provide a comprehensive guide for restaurant owners to strategically improve their services and popularity.

Contributions and References:

Contributions	Safa Eltgani	Osama Kheshaifaty	Chixu Ni
Presentation 1	Responsible for introduction and data cleaning. Reviewed/edited slides	Responsible for menu and hours EDA. Reviewed/edited slides.	Responsible for geographic EDA. Reviewed/edited and provided feedback on all slides.
Presentation 2	Responsible for introduction, data cleaning, menu.	Responsible for Shiny app, and hours. Reviewed/edited and provided feedback on all slides.	Responsible for geographic clustering. Reviewed all slides
Summary	Responsible for introduction, data cleaning, conclusion, and references. Reviewed/edited data analysis section.	Responsible for hours and figures 2,3,4. Reviewed/edited and provided feedback on whole document.	Responsible for attributes and K-means. Reviewed/edited the introduction, data cleaning, and conclusion
Code	Responsible for data cleaning code. Reviewed code for analysis section.	Responsible for methods/results for final models and code to replicate.	Responsible for methods/results under different models. Reviewed data cleaning code.
Shiny App	Reviewed/edited and provided feedback on Shiny app	Co-responsible for Shiny app	Co-responsible for Shiny app

[1] https://www.npd.com/news/press-releases/2022/breakfast-was-u-s-restaurant-industry-bright-spot-in-august/