Diane Lu
July 31, 2023

Tasty Trail – Restaurant Recommender System

Executive Summary

The primary objective of the system is to deliver personalized and data-driven restaurant recommendations, improve users' dining experience and enhance customer loyalty Using recommendation techniques, the system prepares recommendations based on each user's preferences, thereby facilitating the process of discovering new and interesting dining places.

1. Project Problem Statement

In today's fast-paced world, restaurants are constantly evolving, with new restaurants popping up all the time. It was a challenge to keep working and finding places that really matched our interests. While the desire to explore every option is strong, it is simply not practical. That's where the Tasty Trail comes to the rescue, taking you on an exciting foodie adventure. Our goal is to create a user-friendly restaurant recommendation system that understands individual preferences, resulting in delightful dining experiences and empowered customer loyalty.

- **Attract New Leads**: By using targeted recommendations, restaurants can attract potential customers who are more likely to be interested in what they offer. This personalized approach enhances the dining experience and makes it easier to attract new people who feel satisfied and valued.
- **Improved User Engagement**: Personalized recipes encourage consumers to explore new options and foods they may not have considered before. This interactive experience increases user engagement, leading to increased customer satisfaction and loyalty.
- **Deeper Insights into Customer Behavior**: Using personalized recommendation systems allows restaurants to gather valuable data on customer behavior and food preferences. By analyzing this data, restaurants can gain insights into trends, popular dishes and customer preferences, allowing them to make informed business decisions

2. Background on the Subject Matter Area

The problem area of restaurant recommendations is highly relevant in today's digital age. A data-driven restaurant recommendation system can transform the way people explore and choose places to eat, providing personalized and targeted suggestions that cater to each user's unique tastes. By leveraging machine learning algorithms, the system can analyze vast amounts of user-generated data, including reviews, ratings, and past dining experiences.

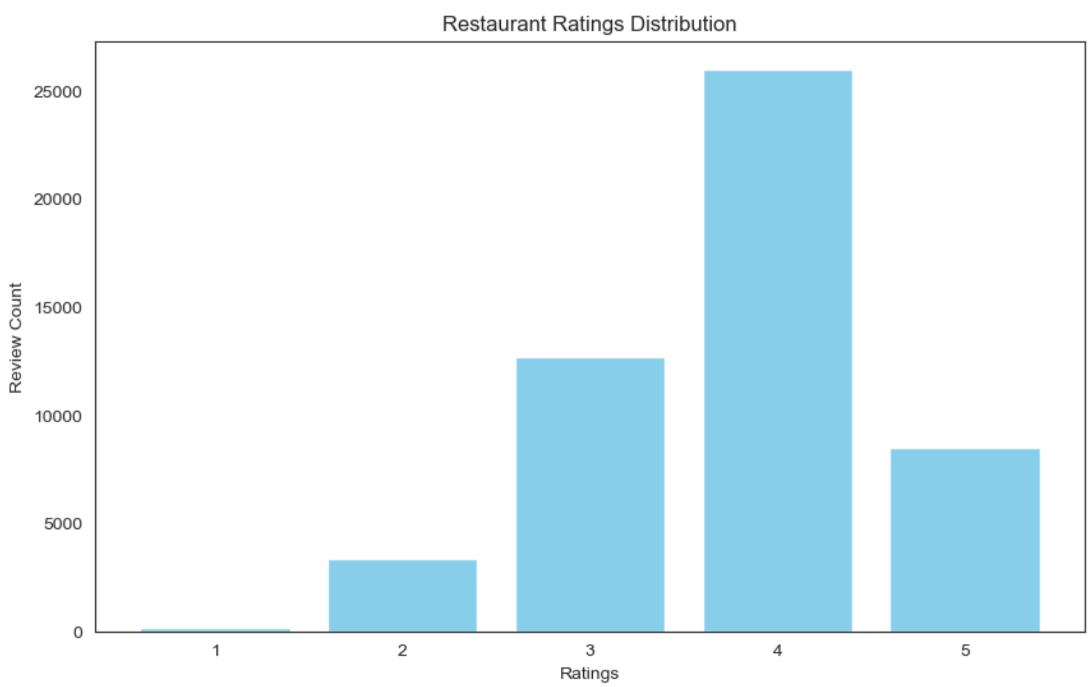
3. Details on the Dataset

The dataset used in this context was obtained from Kaggle and originally sourced from Yelp, one of the most prominent restaurant review platforms. The dataset consists of three separate datasets, each containing information related to businesses, reviews, and users. The primary purpose behind collecting this data is to provide a valuable service to users by offering insights, reviews, and ratings that assist them in making informed decisions about where to eat and what businesses to patronize.

4. Summary of Cleaning and Preprocessing

To begin the data processing, we start by verifying that all entries in our dataset pertain to restaurants, considering that Yelp contains diverse types of businesses. Unimportant columns that do not significantly contribute to our data analysis have been removed. To improve clarity and ease of use, we utilized label encoding for the IDs. Moving forward, we proceeded to filter the restaurants with a minimum of 100 reviews and users with at least 100 reviews. Finally, we will merge these refined datasets to form the final datasets, which will be used for exploratory data analysis (EDA) and modeling purposes.

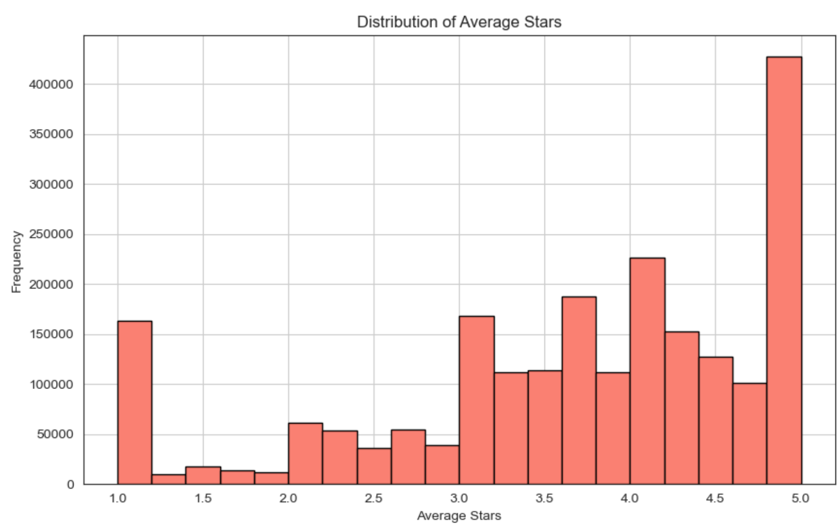
5. Insights, Modeling, and Results



The bar graph shows the review counts for different restaurant ratings, revealing the distribution of reviews across the rating scale. It's evident that a considerable number of reviews are on the higher end, with more 4s and 5s.

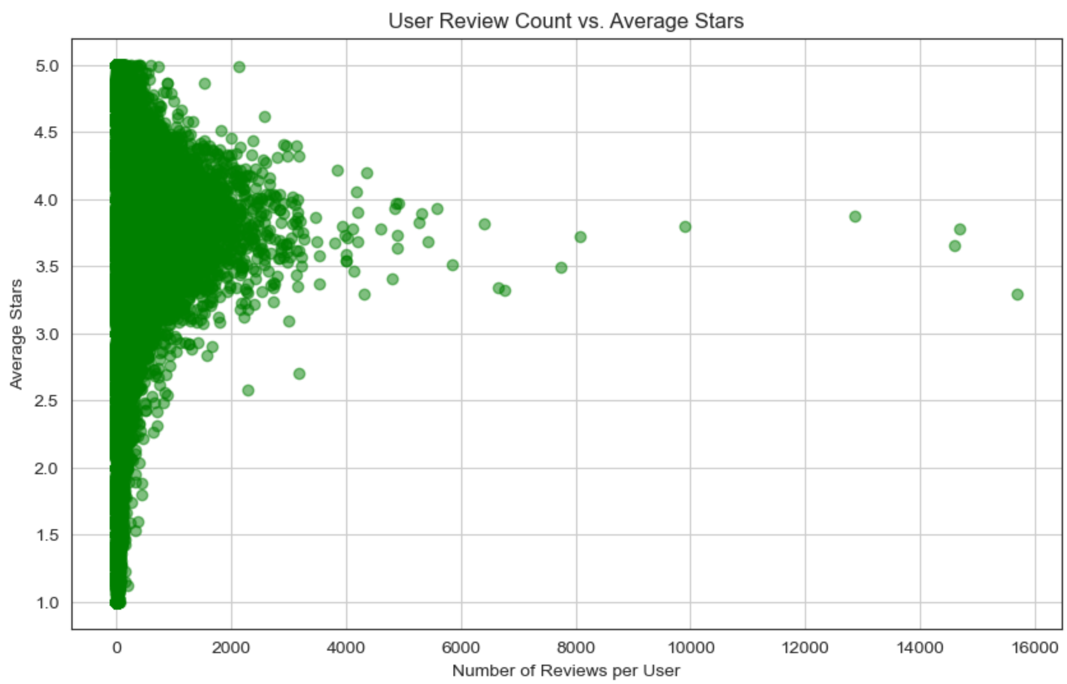
Diane Lu
July 31, 2023

The plot emphasizes that most reviews are positive and tend to be on the higher end. As the rating increases, we observe higher review counts. This relationship provides valuable insights for businesses to assess restaurant performance and customer satisfaction. By analyzing this plot, businesses can identify popular restaurants with high ratings and numerous reviews, helping potential customers make informed dining choices based on public feedback.



The "Distribution of Average Stars" histogram presents the distribution of average star ratings in a user dataset. It is left-skewed, indicating that higher average star ratings are more common. The histogram showcases a tendency towards 4s and 5s, suggesting users generally rate positively. This visualization provides valuable insights into user preferences and satisfaction levels, enabling businesses to understand common average rating ranges and how users perceive their products or services.

Diane Lu
July 31, 2023



Based on the number of reviews made by users, it's evident that most people do not write a large number of reviews. While there are outliers who have written over 6000 reviews, the majority falls within the 0 to 2000 review range. This plot offers valuable insights for businesses to understand their user base and customize strategies to enhance customer satisfaction based on user engagement and feedback.

In our recommendation system, we'll use two main models: content-based and matrix recommendation. The content-based model is perfect for suggesting restaurants based on their types. If you're into Korean cuisine, it can recommend new Korean restaurants for you to explore. On the other hand, the matrix factorization model aims to predict how users would rate restaurants they haven't been to yet.

Here, you can observe the content-based recommender in action. It recommends other Japanese restaurants similar to Suika based on their attributes. However, as we mentioned earlier, the lack of location evaluation results in recommendations from random locations. On the other hand, our matrix model achieved around 1% accurate ratings prediction and reached a 50% success rate when allowing a margin of 1 point.

	restaurant	similarity
78485	Suika	1.000000
39556	Izakaya Amu	0.781389
49712	Lola 42	0.760751
88189	Imperium Food & Wine	0.640402
29483	Savin Bar & Kitchen	0.622994
22038	Kyoto Sushi	0.620734
157113	Yui Japanese Bistro	0.620734
3344	Sushi Junai 2	0.620734
121566	Sushi Hurray	0.620734
38504	Sushi Town	0.620734

6. Findings and Conclusions

Our primary goal was to develop a recommendation system that effectively suggests restaurants customers would genuinely enjoy. During the development process, our content-based recommender demonstrated success in recommending similar restaurants based on category attributes. However, it encountered a limitation as it failed to account for location or radius, leading to recommendations from different cities. To address this, we recognize the need to incorporate cities or latitude and longitude as factors in future iterations to ensure more relevant and localized recommendations.

Another model we implemented was the matrix factorization recommender, designed to predict how users would rate restaurants they haven't visited. While achieving an exact prediction rate of approximately 1% proved challenging, we observed a more promising success rate of around 50% by allowing a margin of 1 point. Although the matrix recommender exhibits potential, it still requires refinement to enhance accuracy and reliability.

During the development process, we encountered significant computational challenges due to the sheer size of the dataset, which included millions of reviews. Our existing machines struggled to process such vast amounts of information effectively. To overcome this obstacle, future improvements should focus on acquiring computational capabilities that can handle large-scale data processing. By doing so, we can expect improved results and higher accuracy in our recommendation system.

In summary, our pursuit of an effective recommendation system led to successful content-based recommendations and promising outcomes with matrix factorization. However, addressing location-related issues and enhancing computational capabilities are essential steps for refining and maximizing the system's performance to deliver truly enjoyable and relevant restaurant recommendations to our customers.

7. Practical Applications

Most restaurant recommendation applications today rely on simplistic metrics like highest ratings, which may not be truly reflective of individual preferences. However, Tasty Trail aims to redefine this approach by offering a personalized app experience. Instead of presenting generic popular choices, Tasty Trail utilizes cutting-edge algorithms to understand each user's unique tastes and preferences. By analyzing the restaurants, a user already enjoys, the app can accurately predict and recommend new establishments that align closely with their culinary interests. This personalized approach ensures that users receive tailored and relevant recommendations, leading to more delightful dining experiences that resonate with their individual palates.

The impact of Tasty Trail's personalized recommendations extends beyond satisfying individual customers. By guiding users to discover new and exciting restaurants that match their preferences, the app plays a pivotal role in supporting local businesses. Lesser-known establishments gain exposure to a wider audience, attracting new customers who are genuinely interested in their offerings. This mutually beneficial relationship between customers and businesses creates a positive feedback loop, enhancing the dining experience for all involved. As Tasty Trail continues to facilitate meaningful connections between diners and restaurants, it contributes to a thriving and diverse culinary ecosystem, making it a win-win solution for both the app's users and the establishments it promotes.

8. Future Directions

When looking for restaurants, it is crucial to receive recommendations within a specific range or area rather than random suggestions. Unfortunately, the current implementation lacks consideration for location-based filtering, which should be taken into account to improve the recommendations.

Moving forward, we should explore different models, such as a hybrid recommender system that combines Content-Based Filtering and Collaborative Filtering. Integrating these two approaches into a single system offers significant advantages, as they can compensate for each other's limitations. By leveraging both methods, we can achieve a more diverse and precise set of recommendations, enhancing the overall user experience.

Conclusion

In conclusion, Tasty Trail, our restaurant recommendation system, aims to provide personalized suggestions for a delightful dining experience. While our content-based recommender works well with category-based recommendations, we need to improve location filtering. The matrix factorization recommender shows potential but requires refinement for better accuracy. Overcoming data processing challenges and exploring hybrid models will lead us to a more diverse and precise recommendation system in the future.