Introduction to SQL-Powered Business Problem Solving

Unlock the insights hidden within your data to drive strategic decision-making and uncover new opportunities for growth. Explore how SQL analytics can transform your understanding of the restaurant industry and provide a competitive edge.

Aa **by [Aaditya Pitroda](#)**

Data Source: Yelp Dataset

About the Yelp Dataset

Yelp is a web and mobile platform that functions as a crowd-sourced local business review site. Users can submit reviews, photos and tips about business, while also browsing information and ratings left by others.

Analytical Opportunities

By leveraging the Yelp Dataset, this project can uncover valuable insights into the relationship between user engagement metrics, such as reviews, tips, and check-ins, and business success indicators, like review count and ratings. These insights can inform strategic decision-making for restaurant owners and operators.

Agenda

- Problem Statement
- Research Objectives
- Hypothesis
- Data Overview
- Analysis and Findings
- Recommendation

Problem Statement

In a competitive market like the restaurant industry, understanding the factors that influence business success is crucial for stakeholders. Utilizing the Yelp dataset, this project aims to investigate the relationship between user engagement (reviews, tips, and check-ins) and business success metrices (review count, ratings) for restaurant.

Research Objective

1. Quantify the correlation between user engagement (reviews, tips, check-ins) and review count/average star rating:
2. Analyze the impact of sentiment on review count and average star ratings:
3. Time trends in User engagement:

Data Exploration and Cleaning

1 Data Overview

This dataset is a subset of Yelp and has information about businesses across 8 metropolitan areas in the USA and Canada, the original data is shared by Yelp as JSON files, this five JSON files are business, review, user, tip and check-in. The JSON files are stored in the database for easy retrieval of data.

2 Understand Data Structure

Examine the schema of the Yelp dataset to understand the different tables and their relationships. This will help identify the key entities and attributes relevant to the analysis.

3 Access Data Quality

Inspect the data for any missing values, duplicates, or inconsistencies that may need to be addressed. This step ensures the data is clean and reliable for further analysis.

Exploratory Data Analysis

To gain a deeper understanding of the factors influencing restaurant success, we will conduct a comprehensive exploratory data analysis (EDA). We will examine the distribution of key variables such as review count, rating, number of tips, and check-ins. This will help us identify patterns, outliers, and potential relationships between user engagement metrics and business performance metrics.

We will also investigate the geographic distribution of the data, looking for any regional or locational trends that may impact a restaurant's success. Additionally, we will analyze the temporal aspects of the data, such as the frequency of reviews, tips, and check-ins over time, to uncover any seasonal or time-based insights.

Hypothesis Testing

Correlation Analysis

We'll examine the correlations between user engagement metrics like reviews, tips, and check-ins, and business success metrics like review count and ratings. This will help us uncover the key relationships driving success in the restaurant industry.

Significance Testing

Using statistical techniques like t-tests and ANOVA, we'll assess the significance of the observed relationships. This will allow us to determine which factors have the strongest impact on business outcomes and are worth further exploration.

Deeper Insights

By delving into the nuances of user engagement and business performance, we aim to uncover insights that can guide strategic decision-making for restaurant owners and managers. Our hypothesis testing will shed light on the key drivers of success in this competitive market.

Predictive Modeling

1

Regression Analysis

Predict business success metrics

2

Classification Models

Identify factors for high/low ratings

3

Time Series Forecasting

Predict future review trends

To uncover the key drivers of business success, we will employ various predictive modeling techniques. Regression analysis will allow us to forecast important metrics like review count and overall ratings based on user engagement data. Classification models will help identify the factors that lead to high or low ratings, providing valuable insights. Additionally, time series forecasting will enable us to predict future trends in user reviews, empowering businesses to proactively address potential issues.

Analysis and Findings

- Out of 150k businesses, 35k are restaurants business and are open.
- Table showing distribution of business success metrics (review count and average rating):

average_review_count	55.975426
min_review_count	5
max_review_count	248
median_review_count	15
avg_star_count	3.477281
min_star_count	1
max_star_count	5
median_star_rating	3.5

Highest Rating

name	review_count	avg_rating
ā café	48	5.0
two birds cafe	77	5.0
the brewers cabinet production	13	5.0
taqueria la cañada	17	5.0
la bamba	44	5.0
la 5th av tacos	24	5.0
el sabor mexican and chinese food	21	5.0
eat.drink.Om...YOGA CAFE	7	5.0
d4 Tabletop Gaming Cafe	8	5.0
cabbage vegetarian cafe	12	5.0

Highest Review Count

name	review_count	avg_rating
McDonald's	16490	1.868702
Chipotle Mexican Grill	9071	2.381757
Taco Bell	8017	2.141813
Chick-fil-A	7687	3.377419
First Watch	6761	3.875
Panera Bread	6613	2.661905
Buffalo Wild Wings	6483	2.344828
Domino's Pizza	6091	2.29021
Wendy's	5930	2.030159
Chili's	5744	2.514706

- Higher ratings do no guarantee a higher review count, or vice versa.
- Success of Restaurants is not solely determined by ratings or review counts.
- Review count reflects user engagement but not necessarily overall customer satisfaction or business performance.

Do restaurants with higher engagement tend to have higher ratings?

- Data shows a general increase in average review, check-in, and tip counts as ratings improve from 1 to 4 stars.
- Restaurants rated 4 stars exhibit he highest engagement and shows a downward trend for rating above 4.
- The drop in engagement at 5.0 stars might suggest either a saturation point where fewer customers feel compelled to add their reviews, or a selectively where only a small, satisfied audience frequents these establishments.
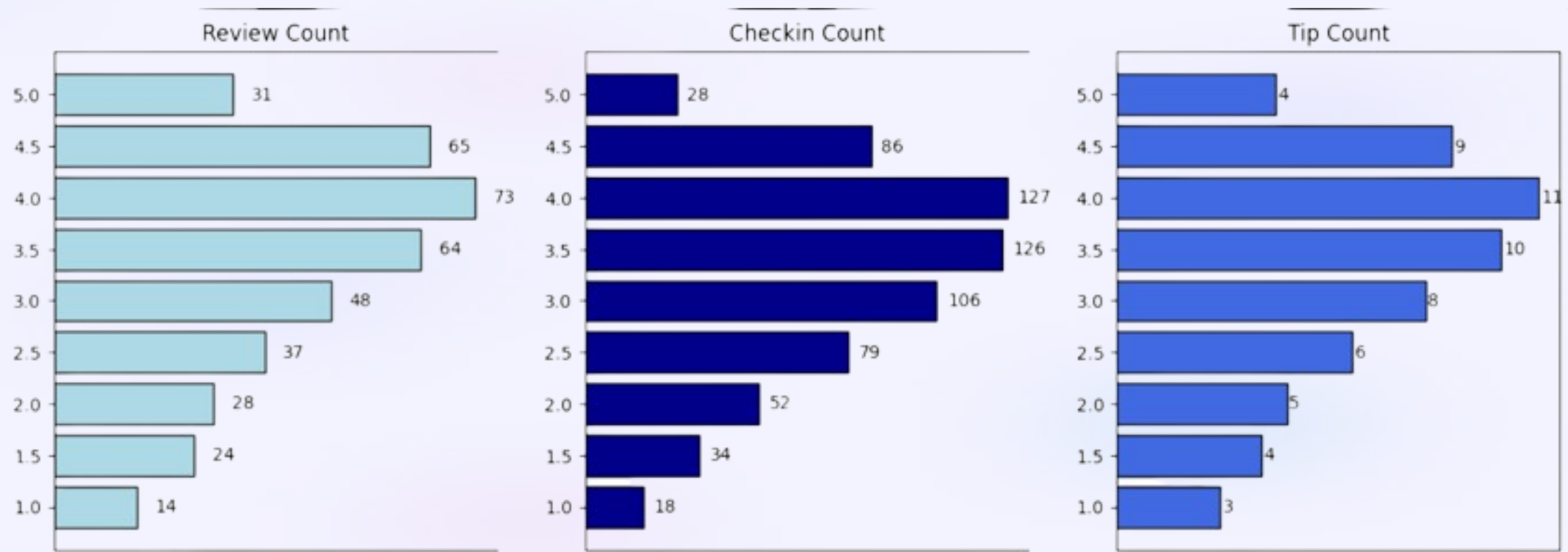
Is there a correlation between the number of Reviews, tips and check-ins for a business?

- Data shows a general increase in average review, check-in, and tip counts as ratings improve from 1 to 4 stars.
- These correlations suggest that user engagement across different platforms (review, tips and check-ins) is interlinked; higher activity in one area tends to be associated with higher activity in others.
- Business should focus on strategies that boost all types of user engagement are likely to drive increases in others, enhancing overall visibility and interaction with customers.
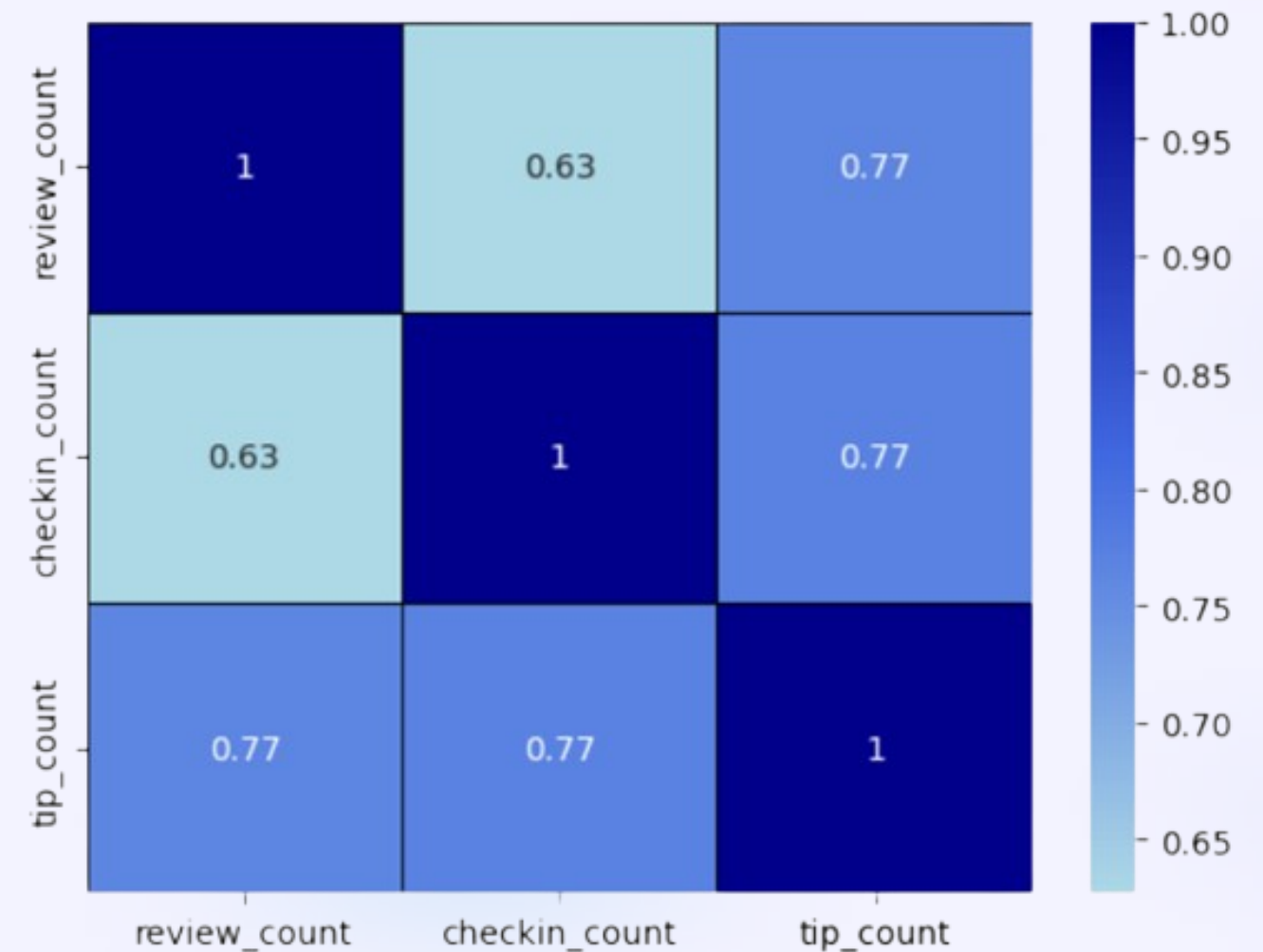
Is there a difference in the user engagement Between high-rated and low-rated businesses?

- Data indicates a clear correlation between higher ratings and increased user engagement across reviews, tips, and check-ins.
- This pattern underscores the importance of maintaining high service and quality standards, as these appear to drive more reviews, check-ins and tips, which are critical metrics of customer engagement and satisfaction

| category	review_count	tip_count	checkin_count
High-Rated	72.291062	10.162766	122.066641
Low_Rated	42.12342	6.541689	88.880828

How do the success metrics of restaurants vary across different states and cities?

- Philadelphia emerges as the top city with highest success score, indicating a combination of high ratings and active user engagement.
- Following Philadelphia, Tampa, Indianapolis and Tucson rank among the top cities with significant success scores, suggesting thriving restaurant scenes in these areas.
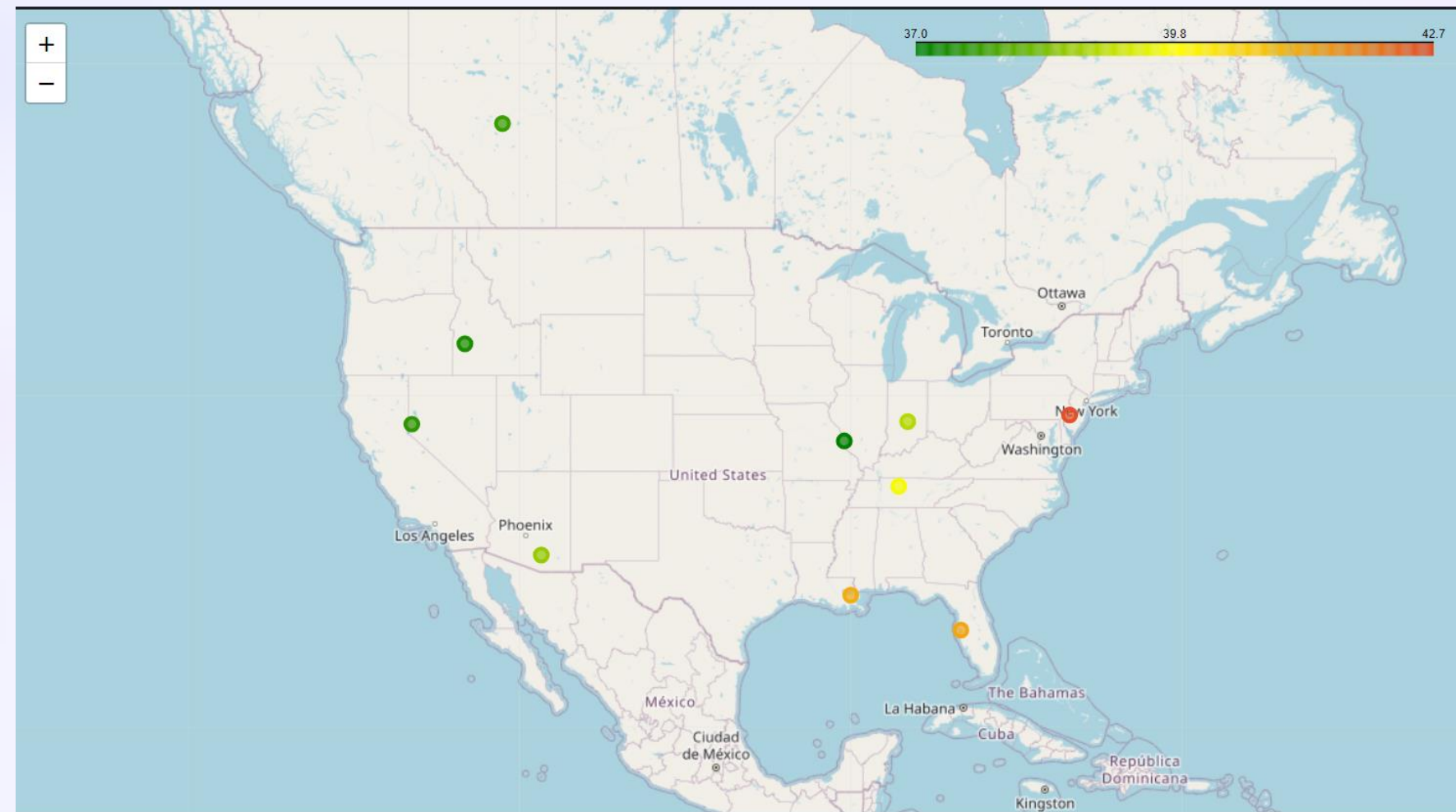
Are there any patterns in the user engagement over time for successful businesses compared to less successful ones?

- Successful business, particular those with higher ratings (above 3.5), exhibit consistent and possibly increasing user engagement over time.
- High rated restaurants maintain a steady of growing level of user engagement over time, reflecting ongoing customer interest and satisfaction.
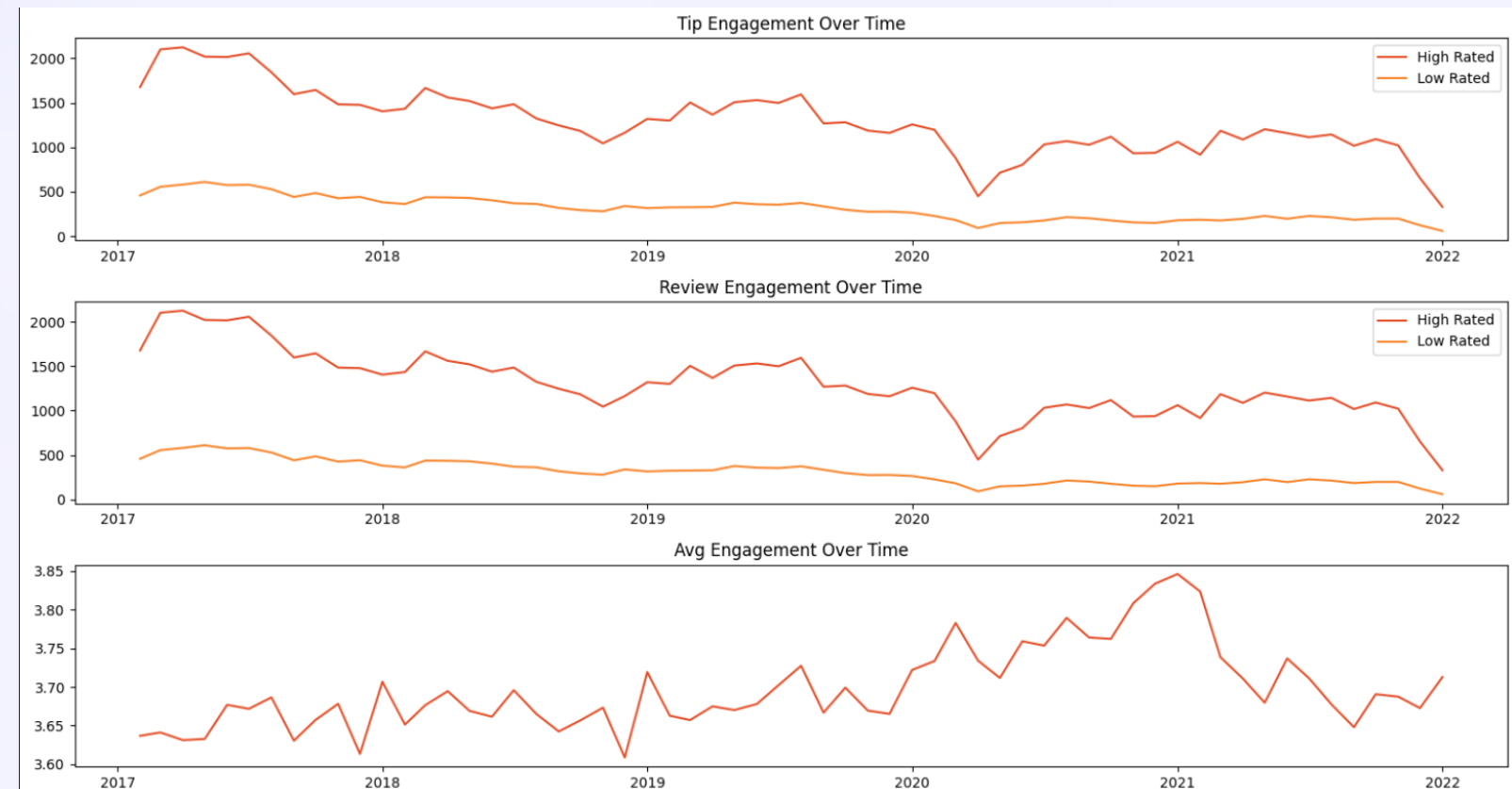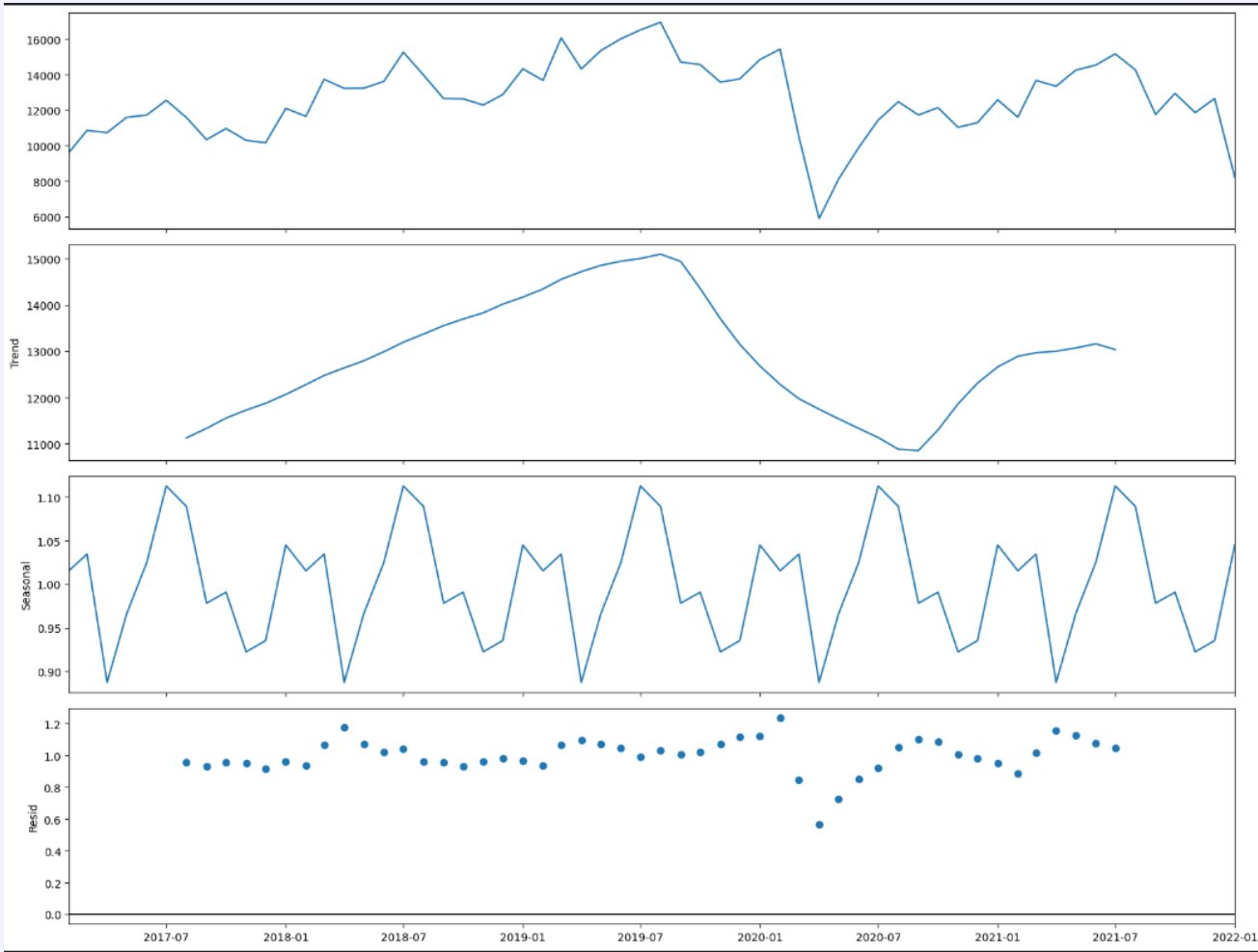
# Trend and Seasonality Analysis

Tip Count
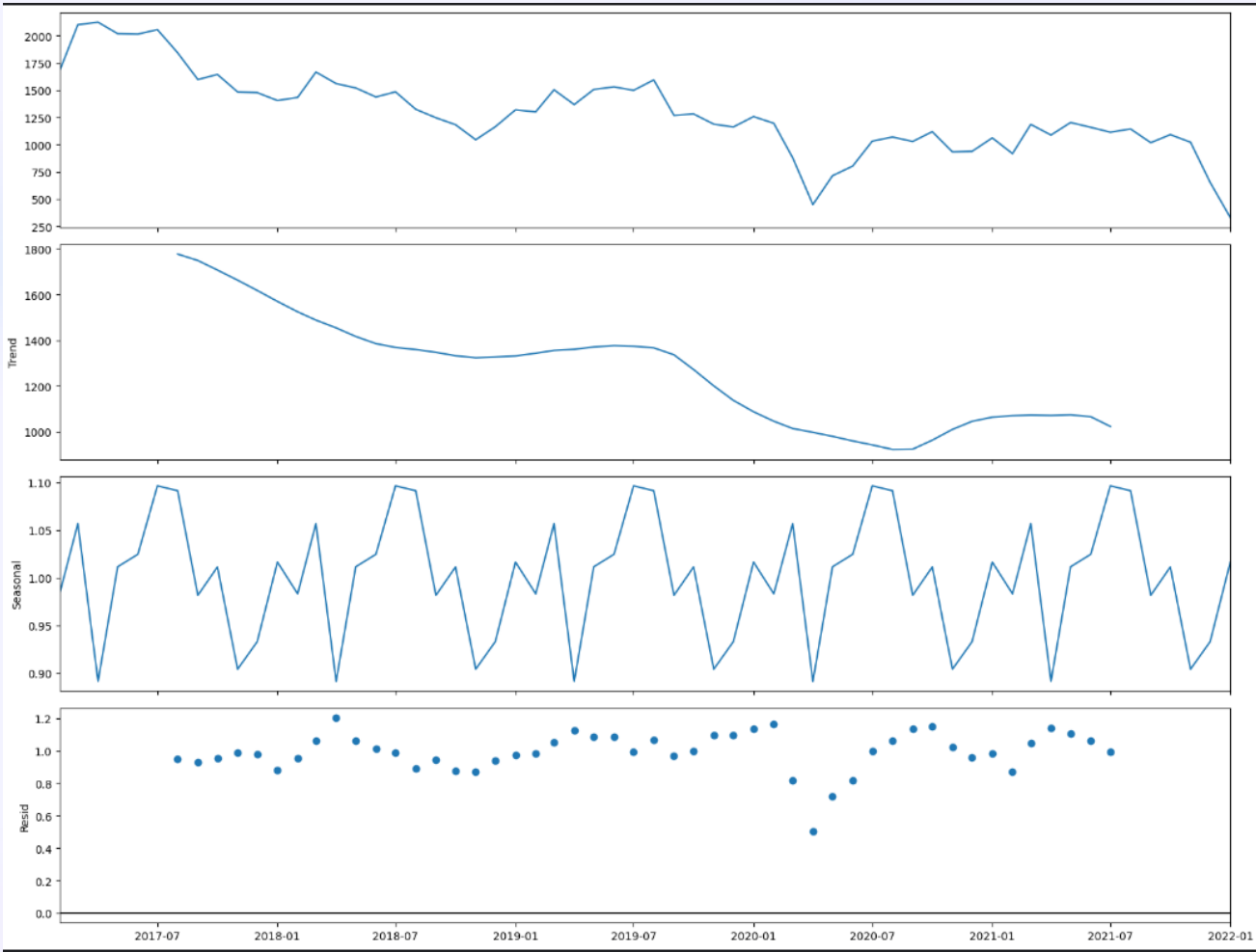
Review Count

How does the sentiment of reviews and tips (useful, funny, cool) correlate with the success Metrics of restaurants?

- "useful", "funny", and "cool" are attributes associated with user reviews. They represent the feedback provided by users about the usefulness, humour or coolness of a particular review.
- Higher counts of useful, funny and cool reviews suggest greater user engagement and satisfaction, which are key factors contributing to a restaurant's success.
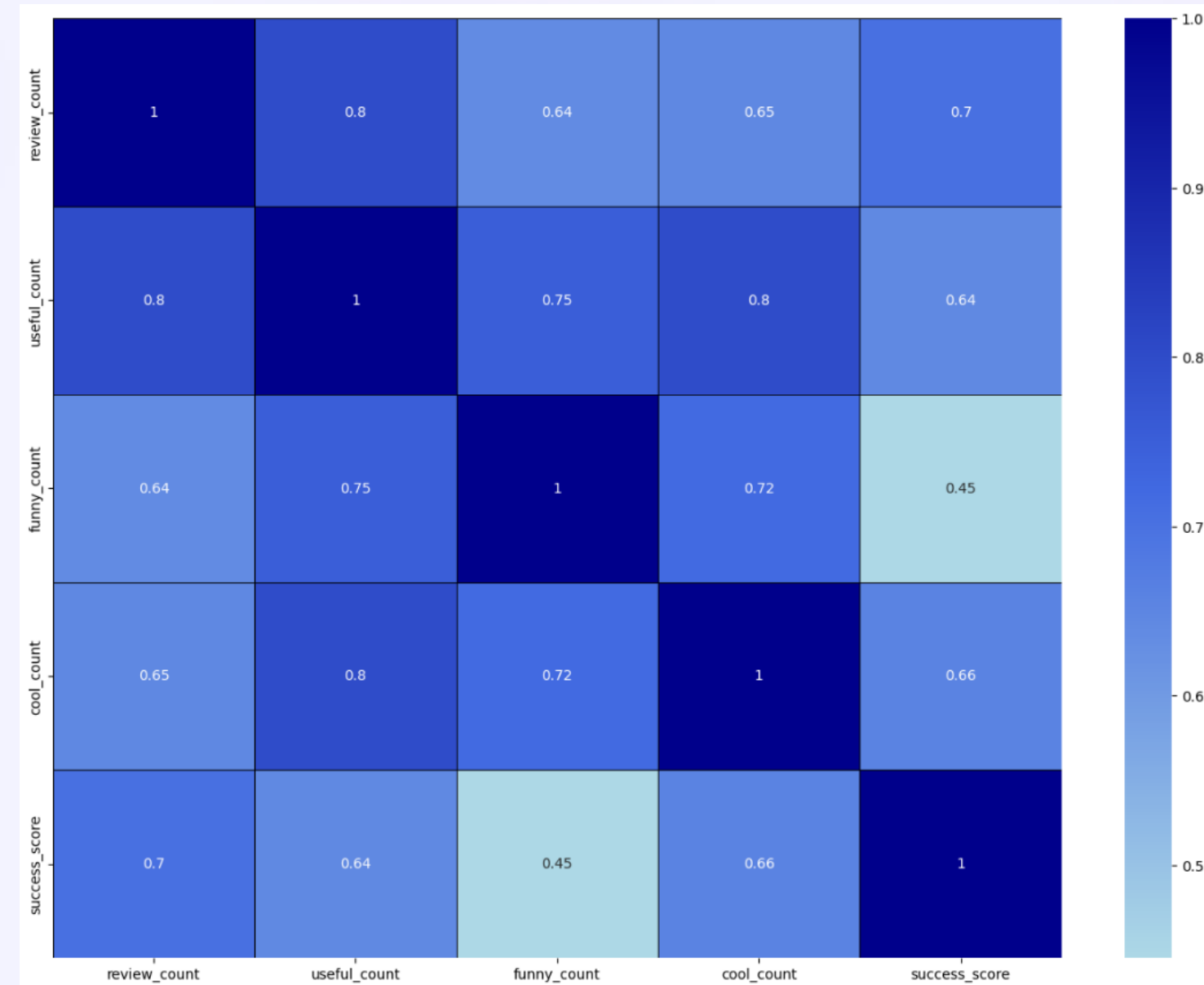
Is there any difference in engagement of elite Users and non-elite users?

- Elite users are individuals who have been recognized and awarded the "Elite" status by Yelp for their active and high-quality contributions.
- Elite users, despite being significantly fewer in number, contribute a substantial proportion of the total review count compared to non-elite users.
- Establishing a positive relationship with elite users can lead to repeat visits and loyalty, as they are more likely to continue supporting businesses they have had good experiences with.

User Distribution

Review Distribution

Not Elite 95.41

Elite 4.59

Elite 44.05

Not Elite 55.95
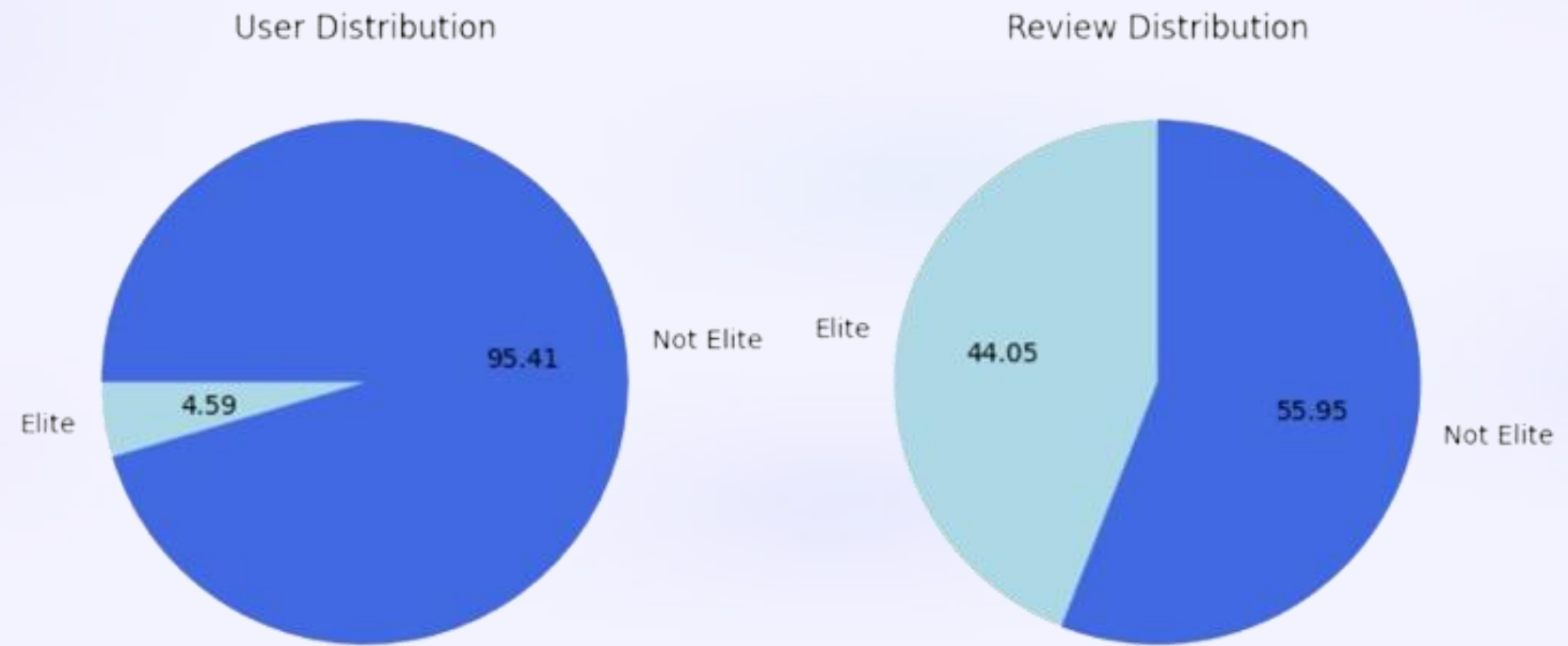
Busiest Hours

- The busiest hours for restaurants, based on user engagement, span from 4 pm to 1 am.
- Knowing the peak hours allows businesses to optimize their staffing levels and resource allocation during these times to ensure efficient operations and quality service delivery.
- The concentration of user engagement during the evening and night hors suggests a higher demand for dining these times, potentially driven by factors such as work schedules social gatherings and leisure activities.
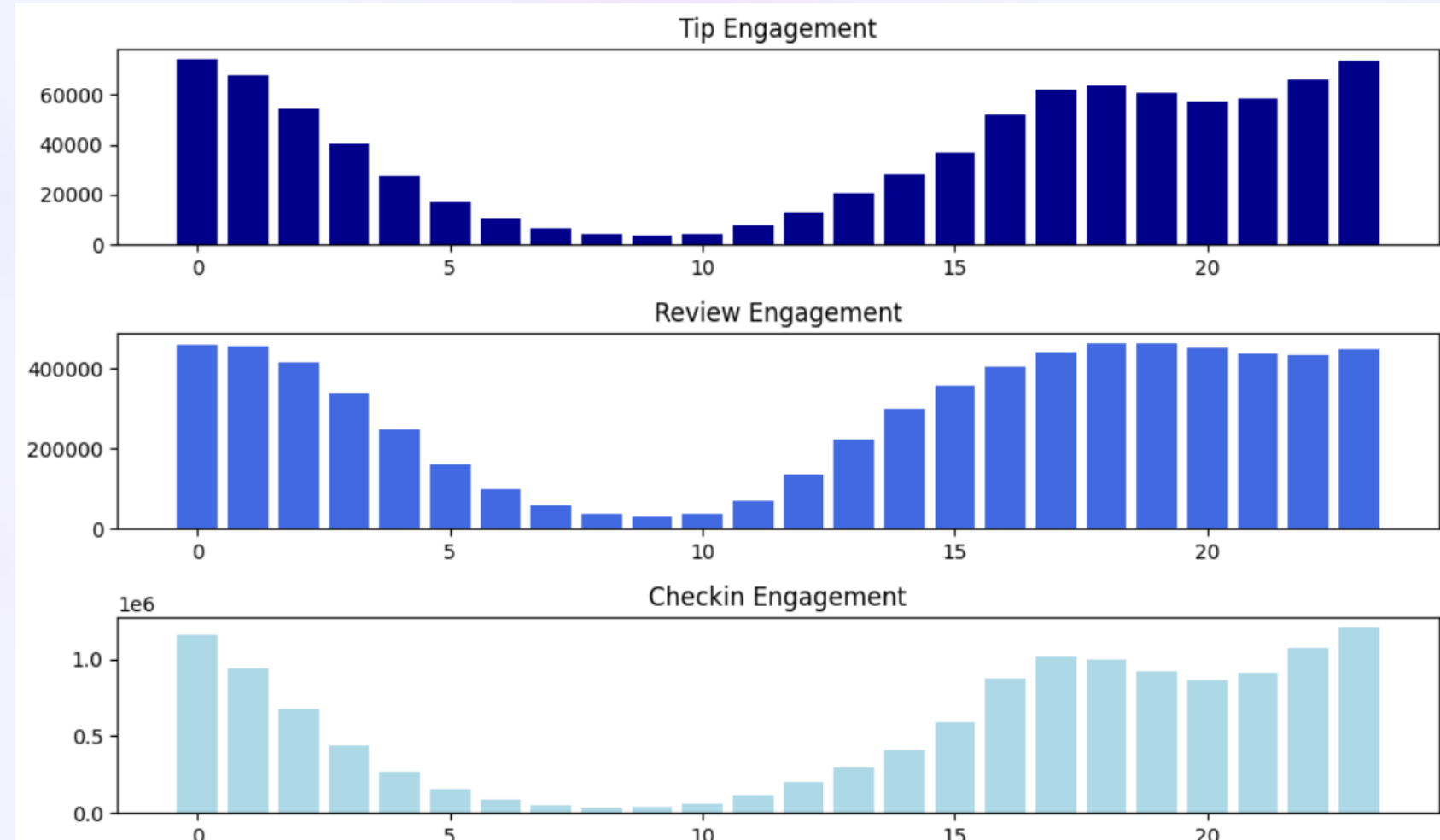
Insights and Recommendations

① **Enhance User Engagement**

The analysis reveals a strong correlation between user engagement metrics (reviews, tips, check-ins) and business success metrics (review count, ratings). Restaurants should focus on strategies to increase user engagement, such as offering incentives for reviews, hosting events, and providing exceptional customer service.

② **Optimize Business Operations**

Factors like business hours, location, and menu offerings also play a significant role in a restaurant's success. Businesses should analyze their operations data and customer feedback to identify areas for improvement, such as adjusting hours, improving accessibility, or diversifying the menu.

③ **Leverage Data-Driven Insights**

The insights gained from this data analysis can empower restaurant owners and managers to make more informed decisions about their business strategy, marketing, and resource allocation. Continuous monitoring and analysis of key performance indicators can help restaurants stay ahead of the competition.

Conclusion

In conclusion, this SQL data analytics project on the Yelp dataset has provided valuable insights into the relationship between user engagement and business success metrics in the restaurant industry. By exploring user reviews, tips, and check-ins, we've gained a deeper understanding of the factors that contribute to a restaurant's success.