# Yelp Dataset SQL Portfolio Project Report

## Introduction

The Yelp Dataset SQL Portfolio Project was developed to analyze the relationship between user engagement and business performance in the restaurant industry. With data from Yelp's public dataset, the project explores how factors such as reviews, tips, and check-ins impact a restaurant's success metrics, including ratings and review counts. The objective is to provide meaningful insights for restaurants to optimize their engagement strategies and improve overall performance.

## Research Objectives

The primary objectives of this project include:

* Analyzing the correlation between user engagement (e.g., reviews, check-ins, tips) and restaurant success metrics.
* Understanding how various factors like user sentiment, peak engagement hours, and geographic location contribute to business success.
* Developing actionable insights that restaurants can apply to enhance their customer engagement and increase their overall ratings.

## Hypothesis

The hypothesis behind this project is that higher levels of user engagement, such as frequent reviews and check-ins, positively affect a restaurant's ratings and overall business success. Additionally, elite users—those with a high level of activity on Yelp—are likely to have a stronger impact on business performance.

## Data Overview

The dataset used for this analysis is the **Yelp Open Dataset**, which contains data about businesses, users, reviews, and check-ins. The key components of the dataset include:

* **Business Data**: Information about restaurants, including their name, location, and categories.
* **Review Data**: User reviews, including star ratings and text content.
* **User Data**: User profiles, which provide details on elite status and engagement levels.
* **Check-in Data**: Records of customer check-ins, indicating peak engagement hours.
* **Tip Data**: User-generated tips, which offer concise insights into customer experiences.

The data was stored in a structured SQL database, and Python was used for further data exploration and visualization.

# Analysis and Findings

## Correlation Between Engagement and Ratings

The analysis revealed a strong correlation between higher user engagement and better restaurant ratings. Restaurants with frequent reviews and check-ins tend to have higher average star ratings and more review counts, indicating that sustained customer interaction helps boost visibility and credibility.

### Sentiment Analysis

Sentiment analysis on the review text highlighted that positive sentiment (i.e., more favorable language) directly influences a restaurant's ratings. Businesses that received more positive reviews generally had higher ratings.

### Time Trends

The analysis also looked at how engagement levels vary over time. Interestingly, many restaurants saw an increase in user engagement during peak hours from **4 PM to 1 AM**, particularly on weekends. Seasonal trends, such as holidays and special events, also had a noticeable effect on user engagement.

### Geographic Trends

There were clear differences in restaurant success based on geographic location. Businesses located in urban areas or popular dining destinations generally received more reviews and higher engagement than those in rural or suburban locations. The analysis further revealed that restaurants in states like **California** and **New York** tended to perform better in terms of user engagement and ratings.

### Elite vs Non-Elite Users

The study found that elite Yelp users contributed significantly to a restaurant's engagement. Restaurants that attracted reviews from elite users had higher credibility and more in-depth reviews, leading to improved performance metrics.

### Peak Engagement Hours

Restaurants saw a noticeable increase in check-ins and reviews between **4 PM and 1 AM**, especially during weekends. This insight suggests that businesses can optimize their operations and promotional activities around these peak times.

## Recommendations

Based on the findings, the following recommendations are proposed for restaurants to improve user engagement and business performance:

* **Engage Elite Users**: Encourage elite users to review the business by offering special deals or events targeted at this demographic.
* **Optimize Business Hours**: Focus on offering promotions or events during peak hours of engagement, particularly between 4 PM and 1 AM.
* **Leverage Positive Reviews**: Showcase positive reviews prominently on websites and social media to build credibility and attract more customers.
* **Focus on Seasonal Trends**: Use data from holidays and special events to plan promotions and increase engagement during high-traffic periods.

## Conclusion

The Yelp Dataset SQL Portfolio Project successfully demonstrated how data analysis can provide valuable insights into the restaurant industry's success metrics. By analyzing user engagement through reviews, tips, and check-ins, we have identified key factors that influence business performance, such as positive sentiment, elite user engagement, and peak activity times. These insights can be applied by restaurants to optimize their strategies and improve their overall performance.

## Appendix

* **Dataset**: Yelp Open Dataset
* **Tools**: SQL, Python, Power BI, Excel
* **GitHub Link**: [Yelp Dataset SQL Portfolio Project](https://github.com/mishradiya/Yelp-Dataset-SQL-Portfolio-Project-.git)