

Project Title: Yelp Data Analysis

Name: Ehiane Oigiagbe

Date: 7/6/2024

Table of Contents:

1. Introduction
2. Data Description
3. Analysis and Results
 - Business Performance Analysis
 - Top 10 Businesses by Review Count
 - Average Star Rating by City
 - Businesses with the Most Check-ins
 - User Engagement Analysis
 - User with the Most Reviews
 - Top 10 Most Active Users
 - Review Analysis
 - Top 10 Most Reviewed Businesses
 - Average Rating by Business
4. Visualizations
5. Insights and Conclusions
6. Appendices
 - SQL Scripts
 - Python Scripts
 - Data Sources

Introduction

This report presents the analysis of Yelp data to uncover insights about business performance, user engagement, and review patterns. The primary objectives of this analysis are to identify top-performing businesses, active users, and review trends to provide actionable insights for stakeholders.

Project Title: Yelp Data Analysis

Name: Ehiane Oigiagbe

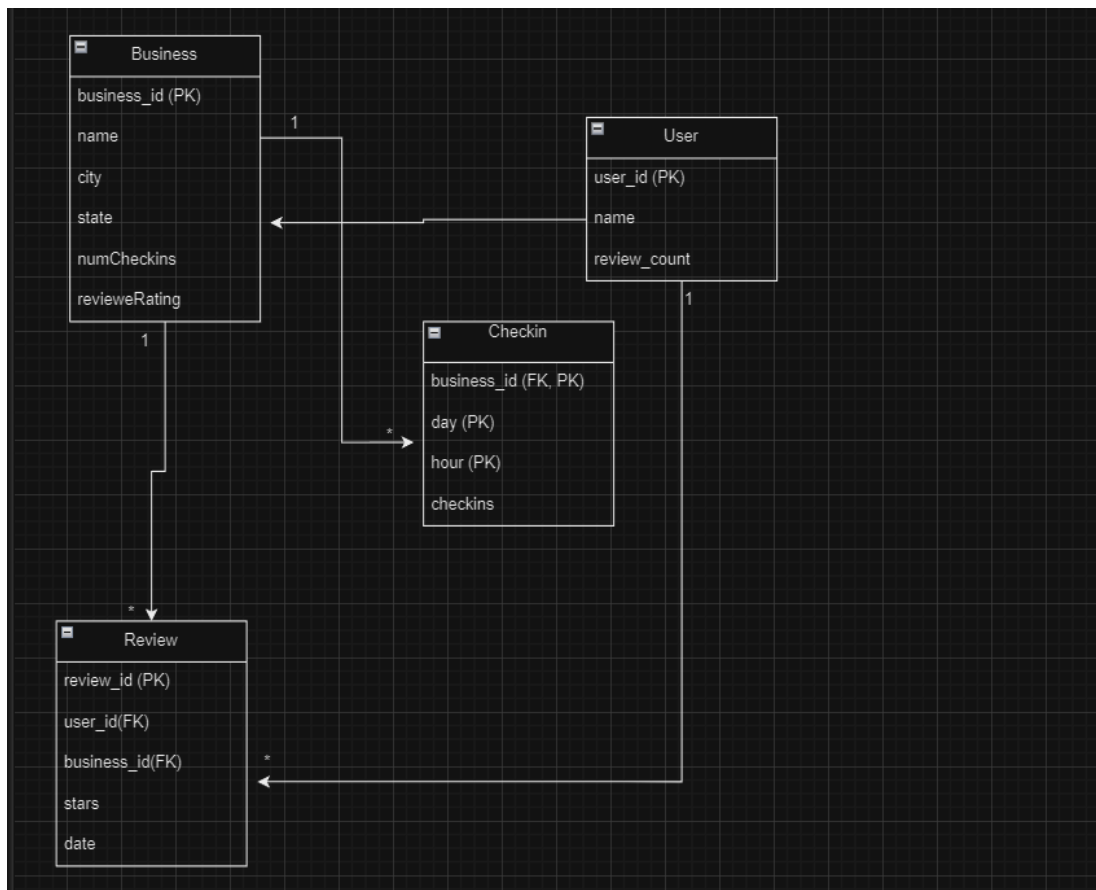
Date: 7/6/2024

Data Description:

The data used for this analysis includes information about businesses, users, reviews, and check-ins. The dataset was processed and stored in a PostgreSQL database with the following schema:

- ❑ **Business:** Contains information about businesses including their name, location, review statistics, and new attributes for number of check-ins (numCheckins) and average review rating (reviewRating).
- ❑ **Users:** Contains information about users including their name, review count, and other engagement metrics.
- ❑ **Review:** Contains review data including review text, rating, and associated business and user IDs.
- ❑ **Checkin:** Contains check-in data indicating the number of check-ins at businesses over time.

Updated ER Diagram



Project Title: Yelp Data Analysis

Name: Ehiane Oigiagbe

Date: 7/6/2024

Analysis and Results:

Business Performance Analysis:

Top 10 Businesses by Review Count:

SQL Query:

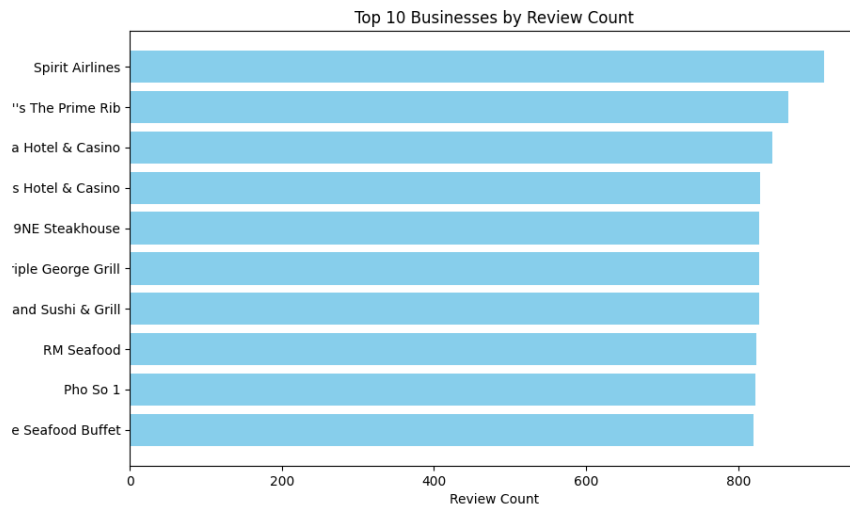
```
SELECT name_, city, state, review_count FROM Business ORDER BY review_count DESC  
LIMIT 10;
```

Explanation: This query retrieves the top 10 businesses based on the number of reviews. It includes the business name, city, state, and review count.

Project Title: Yelp Data Analysis

Name: Ehiane Oigiagbe

Date: 7/6/2024



Insights:

- Spirit Airlines has the highest review count among businesses.
- The list includes a mix of restaurants and service providers

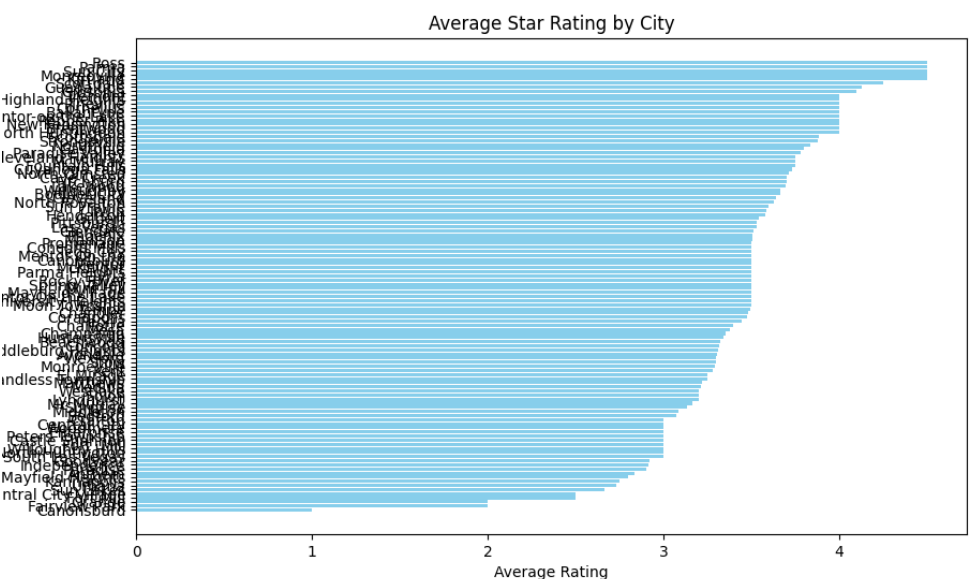
Average Star Rating by City:

SQL Query:

```
SELECT city, AVG(stars) as avg_rating
FROM Business
GROUP BY city
ORDER BY avg_rating DESC;
```

Explanation: This query calculates the average star rating for businesses grouped by city.

Visualization:



Summarized table:

Average Star Rating by City:		
	city	avg_rating
0	Ross	4.5
1	Parma	4.5
2	Sun City	4.5
3	Monreoville	4.5
4	Kirtland	4.5
..
105	Central City Village	2.5
106	Fort Mill	2.5
107	Orange	2.0
108	Fairview Park	2.0
109	Canonsburd	1.0
[110 rows x 2 columns]		

Insights:

- Cities with the highest average star ratings are identified.
- Indicates regions with higher customer satisfaction.

Project Title: Yelp Data Analysis

Name: Ehiane Oigiagbe

Date: 7/6/2024

User Engagement Analysis:

User with the Most Reviews:

SQL Query:

```
SELECT user_id, name_, review_count
```

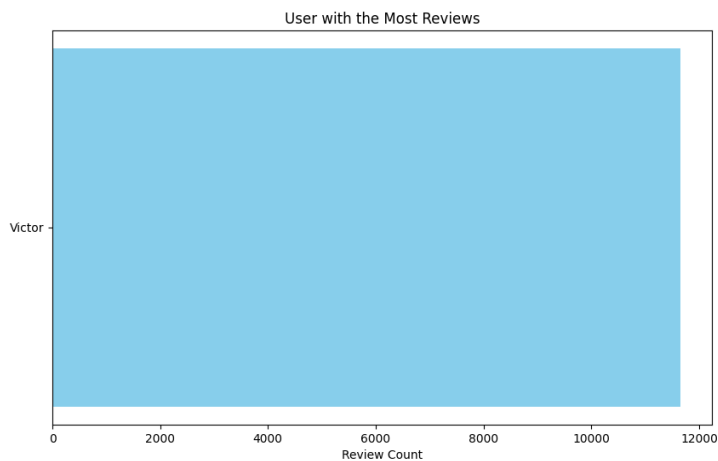
```
FROM Users
```

```
ORDER BY review_count DESC
```

```
LIMIT 1;
```

Explanation: This query identifies the user with the highest number of reviews.

Visualization:



Insights:

- Victor is the user with the most reviews, showcasing significant engagement.

Top 10 Most Active Users:

SQL Query:

```
SELECT user_id, name_, review_count
```

```
FROM Users
```

```
ORDER BY review_count DESC
```

Project Title: Yelp Data Analysis

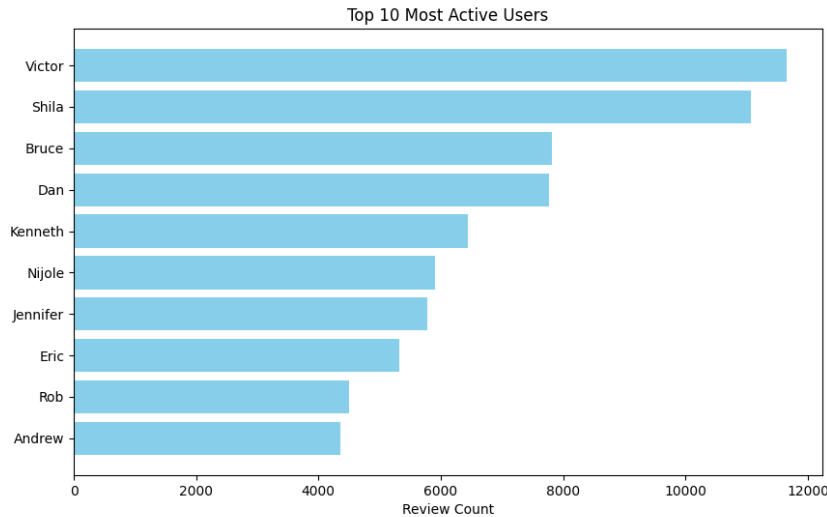
Name: Ehiane Oigiagbe

Date: 7/6/2024

LIMIT 10;

Explanation: This query lists the top 10 users based on their review count.

Visualization:



Insights:

- Lists users with the highest review counts, indicating high activity levels.

Review Analysis:

Top 10 Most Reviewed Businesses:

SQL Query:

```
SELECT B.name_, B.city, B.state, COUNT(R.review_id) as review_count
FROM Business B
JOIN Review R ON B.business_id = R.business_id
GROUP BY B.name_, B.city, B.state
ORDER BY review_count DESC
```

Project Title: Yelp Data Analysis

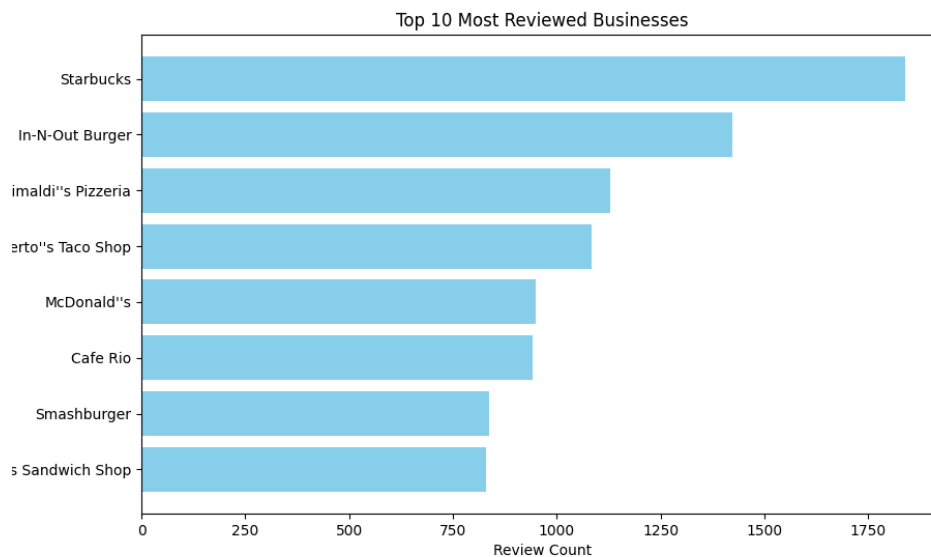
Name: Ehiane Oigiagbe

Date: 7/6/2024

LIMIT 10;

Explanation: This query identifies the top 10 businesses with the highest number of reviews.

Visualization:



Insights:

- Starbucks is the most reviewed business, followed by popular food chains.

Average Rating by Business:

SQL Query:

```
SELECT B.name_, B.city, B.state, AVG(R.stars) as avg_rating
```

```
FROM Business B
```

```
JOIN Review R ON B.business_id = R.business_id
```

```
GROUP BY B.name_, B.city, B.state
```

```
ORDER BY avg_rating DESC
```

```
LIMIT 10;
```

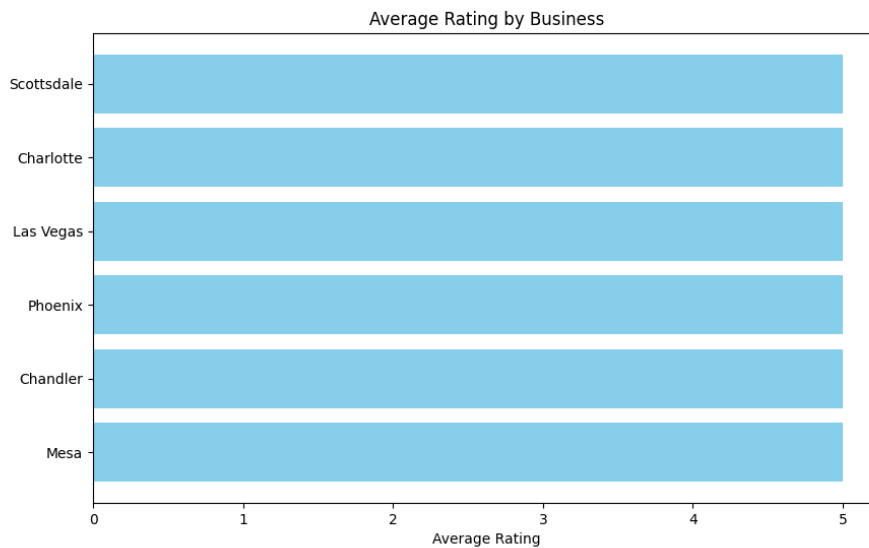

Project Title: Yelp Data Analysis

Name: Ehiane Oigiagbe

Date: 7/6/2024

Explanation: This query calculates the average star rating for businesses based on reviews.

Visualization:



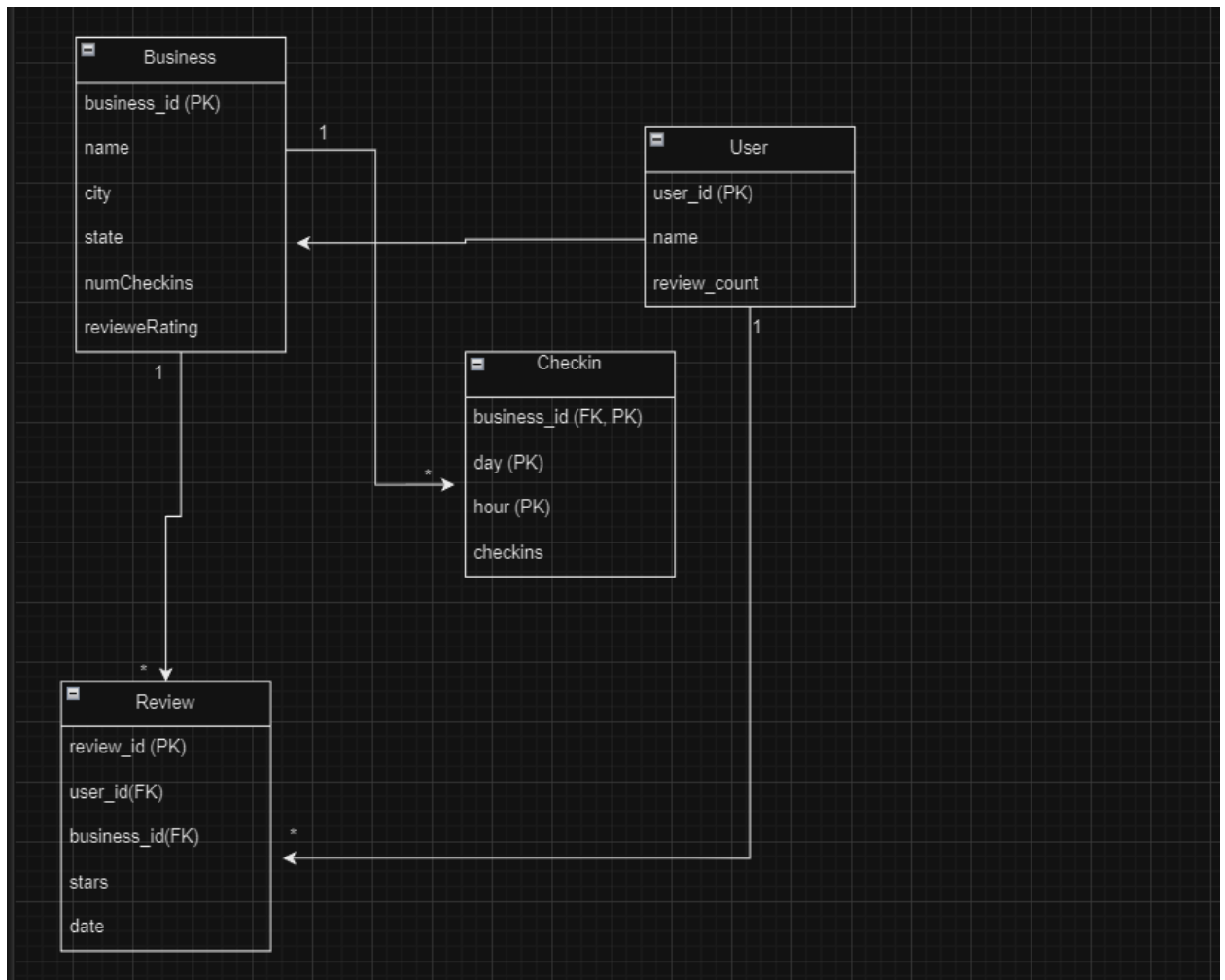
Insights:

- Highlights businesses with the highest average ratings, indicating high customer satisfaction.

Project Title: Yelp Data Analysis

Name: Ehiane Oigiagbe

Date: 7/6/2024



Insights and Conclusions:

- **Top Performers:** Businesses like Starbucks and Spirit Airlines are top performers in terms of reviews and check-ins.
- **User Engagement:** Highly active users contribute significantly to the review ecosystem, with some users posting thousands of reviews.
- **Regional Satisfaction:** Certain cities have higher average ratings, indicating regions with higher customer satisfaction.