# Project Milestone 2: User Level Analysis

## Objective

This milestone seeks to analyze user-level data to understand behavior, contributions, and influence on the Yelp community. By studying reviews, tips, and other activities created by users, we aim to understand their contributions to Arizona-based businesses.

## Dataset Details:

- A refined dataset focusing on active users within the Yelp community interacting with Arizona businesses.

## Tools Used:

- Apache Spark for efficient dataset querying and filtering.
- Python for data visualization.

## Queries and Findings

### Query 1: Most Active Users

Objective: Identify users with the most reviews.
Finding: Top users, such as Fox (17,473 reviews) and Victor (16,978 reviews), are significant contributors to the dataset.

### Query 2: Elite Users Count

Objective: Determine the total number of elite users.
Finding: The dataset includes 91,198 elite users.

### Query 3: Average Stars by User

Objective: Identify users with the highest average ratings.
Finding: Several users, including David, Shea, and Patrick, have an average rating of 5.0 stars.

### Query 4: Users Receiving Most Compliments

Objective: Determine which users have received the most compliments.
Finding: Richard leads with 201,031 compliments, followed by Brian with 84,959.

### Query 5: Top Contributors by Business Category

Objective: Find users contributing the most reviews to specific business categories.
Finding: Isabella leads with 399 reviews in the restaurant category.

### Query 6: User Influence by Average Business Rating

Objective: Analyze the effect of user ratings on business ratings.
Finding: Influential users like Rich and Jonathan maintain 5.0 user stars, closely aligned with high business stars.

### Query 7: Review Trends by Elite Status

Objective: Compare the average ratings of elite and non-elite users.
Finding: Elite users consistently give higher ratings, averaging 5.0 stars.

**Query 8: Tip Analysis**
Objective: Highlight users contributing the most tips.
Finding: Michael leads with 4,071 tips, followed by Marcia with 1,385.

**Query 9: User Influence on Business Ratings**
Objective: Analyze the relationship between user average ratings and the average ratings of businesses they review, identifying users with significant influence on business ratings.
Finding: Users like Pete, Jonathan, and Manny exhibit high average personal ratings (5.0 stars), while the average business ratings they review range from 2.5 to 4.5 stars. This suggests that while these users rate businesses positively, their ratings may not always align with the overall business ratings.

**Query 10: Review Trends Over Time by User Activity**
Objective: To analyze review trends over time for highly active Yelp users (100+ reviews) and identify patterns in their yearly contributions.
Finding: The analysis highlights that in 2005, Zachary emerged as the most active user with an impressive 171 reviews, far surpassing other contributors. Andy followed with 60 reviews, and Patrick with 51 reviews. Other notable contributors included Gourmet with 37 reviews and Keith with 24 reviews. This indicates that a relatively small but highly active group of users played a pivotal role in driving review activity during this period, showcasing their sustained engagement and influence within the Yelp community.


## Conclusion
User contributions are critical in shaping business visibility and reputation. Elite users and active contributors play a pivotal role in maintaining the quality of Yelp data.

## Recommendations

**For Businesses:**

- Focus on engaging elite and active users to enhance visibility and credibility.
- Encourage tips and detailed reviews to attract more customers.

**For Researchers:**

- Study temporal trends in user activity and their impact on ratings.
- Explore the influence of user characteristics (e.g., elite status) on review patterns.