Yelp User Data Report

**Dataset Overview:**

* **Total Entries**: 552,339 rows
* **Total Columns**: 11 columns

**Correlation Analysis – Votes Received:**

**A graph with a line of blue dots

Description automatically generated**

Examining correlations between cool votes, funny votes, and useful votes, there was a strong correlation between the three:

Cool/Funny- 0.9764113

Cool/Useful- 0.9832708

Funny/Useful- 0.9546541

This suggests that users who write reviews in one area (funny, cool, or useful) are likely to receive recognition in other categories. Highlighted by the figure, more votes in one category indicates an upward trend of votes in another category.

**Predicting Fans:**

Review Count – Weak Relationship

Writing more reviews may increase visibility, but it is not the primary reason for fans.

Average Stars – Stronger Relationship

Users with higher average ratings tend to have more fans, indicating that consistently good reviews can help increase following.

**Column Descriptions:**

1. **user\_id**: Unique ID for each Yelp user
2. **name**: User's name (optional)
3. **review\_count**: Total number of reviews written by the user
4. **average\_stars**: Average star rating given by the user
5. **cool\_votes**: Number of "cool" votes received
6. **funny\_votes**: Number of "funny" votes received
7. **useful\_votes**: Number of "useful" votes received
8. **friends**: List of friends (as strings)
9. **elite**: Years the user was elite
10. **yelping\_since**: User's registration date on Yelp
11. **fans**: Number of fans the user has

**Predicting Fans Figures:**

A graph showing a green line and a green line

Description automatically generated

A graph with red dots

Description automatically generated

A graph showing a number of numbers

Description automatically generated

**Review Count Clusters:**

Highlights differences in activity levels, with higher activity garnering more fans.

Findings: Users with more review counts generally have more fans, but it is not the driving factor behind the number of fans.

A graph with green and blue dots

Description automatically generated

**Average Stars Clusters:**

Users with high ratings belong to the “most popular” group.

Findings: Users with higher average stars are more likely to have a higher number of fans across the board.

Three clusters were chosen to represent the activity levels of all users. One category represents the least active (or completely inactive) users, the next representing more active users, who show a greater following likely due to being more active. The third category represents the most active users, which shows a high correlation between account activity and followers, which makes sense as users would not be likely to be fans of accounts with little content or consistency.

**Popularity:**

Popularity is measured in this report by the number of fans a user has. A high number of fans reflects ability to attract and maintain a consistent following.

**Impact:**

Impact is inferred from engagement (cool/useful/funny votes). These metrics show the influence and value a user provides, as their reviews create reactions and feedback from other users.

**To increase Yelp popularity, users should focus on three strategies:**

* **Write engaging reviews**
  + Engagement metrics (votes) correlate strongly with each other. Positive accolades from one category usually bring others.
* **Maintain High Ratings**
  + Users with >4.0 average rating consistently belong to cluster 3 (most impactful/popular).
* **Quality > Quantity**
  + Accounts with the most reviews do not always have the most fans.