Yelp Data Analysis

Through the use of R-Studio a dataset containing 552,339 entries was analyzed to determine various pieces of information from the data. The columns of the dataset were “user\_id”, ”name”, “review\_count”, “average\_stars”, “cool\_votes”, “funny\_votes”, “friends”, “elite”, “yelpin\_since”, and “fans”.

A graph with a line and dots

Description automatically generated

With the use of a correlation matrix, it could be determined the relationship between cool votes, funny votes, and useful votes.

| **cool\_votes** | | **funny\_votes** | | **useful\_votes** | |
| --- | --- | --- | --- | --- | --- |
|  |  | |  | |  |
| **cool\_votes** | 1.0000000 | | 0.9764113 | | 0.9832708 |
| **funny\_votes** | 0.9764113 | | 1.0000000 | | 0.9546541 |
| **useful\_votes** | 0.9832708 | | 0.9546541 | | 1.0000000 |

A graph with a line and a dotted line

Description automatically generated

From the linear regression for the graph Review Count vs. Fans, it indicates that that there is a slight correlation between the number of reviews made and the number of fans a yelper has. The linear regression coefficient is 0.071, further showing there isn’t a significant change in fans due to number of reviews.

A graph with black dots and a blue line

Description automatically generated

The Fans vs. Useful votes graph is similar to the Review Count vs. Fans graph in that there is a slight correlation that the more useful votes that a yelper gets results in a large number of fans. The linear regression coefficient is 0.026 in this case also showing no significant change in fans from useful votes.

Conclusion:

A yelper can enhance popularity and impact through consistent engagement with the community with reviews that are useful, funny, and cool. Also letting time take its course, older users tend to have more fans as time progresses.