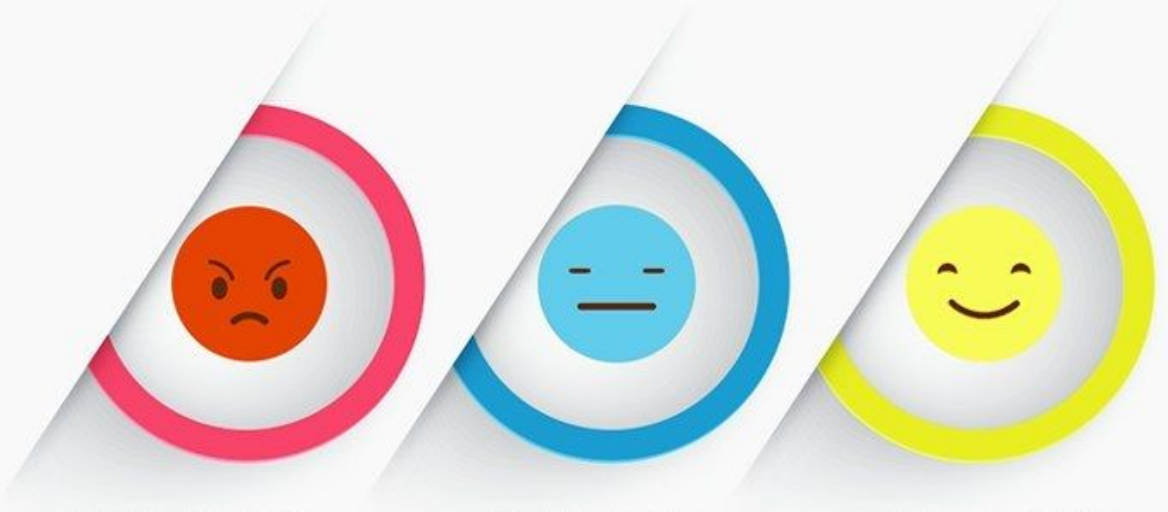
Sentiment Analysis on Text Reviews Compared To Their Respective Star Ratings

By Len Huang

Abstract

- TextBlob NLP library is widely-used for sentiment analysis
- Draw data from the Yelp Academic dataset,
  - Text reviews and star ratings
  - Analyzing them by the number of words each review had
  - Use TextBlob to get sentiment values of these reviews
  - Compared generated sentiment values with their respective user-given star rating
- Matched pairs t-test was not ideal
- Mean differences were promising

SENTIMENT ANALYSIS



NEGATIVE

Totally dissatisfied with the service. Worst customer care ever.

NEUTRAL

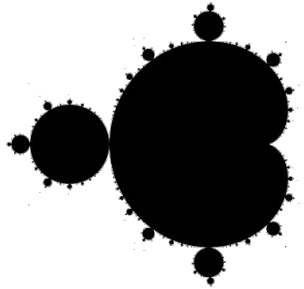
Good Job but I will expect a lot more in future.

POSITIVE

Brilliant effort guys! Loved Your Work.

Libraries

- Premade tools
- Different approaches

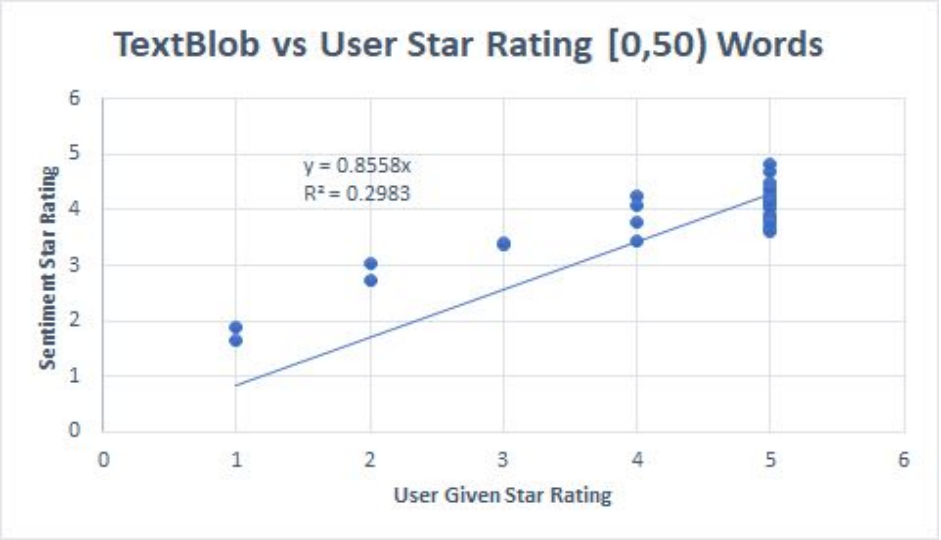


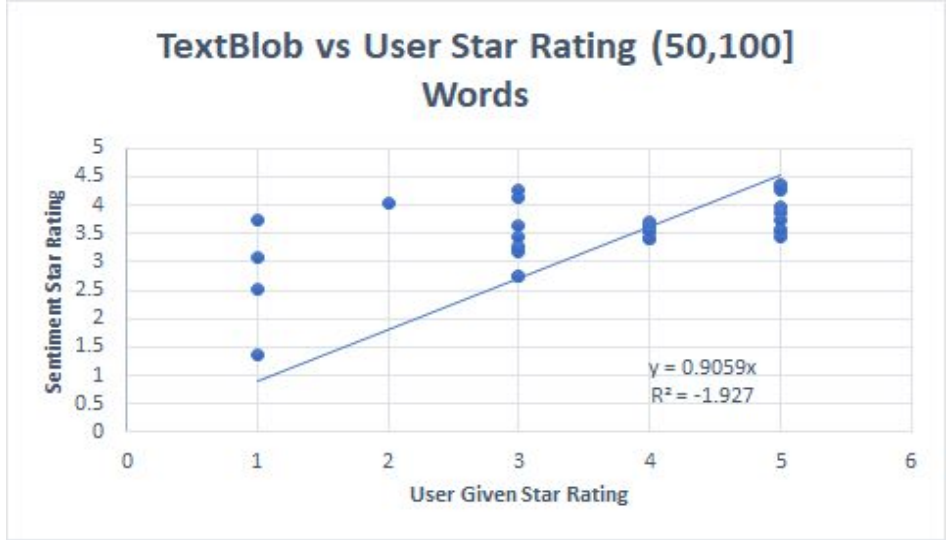
TextBlob

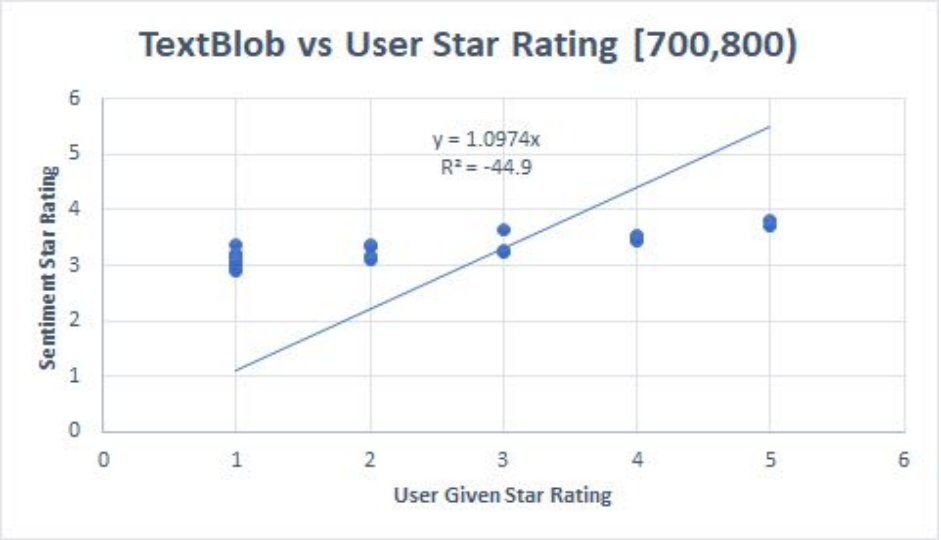
Natural Language Analysis
with Python NLTK

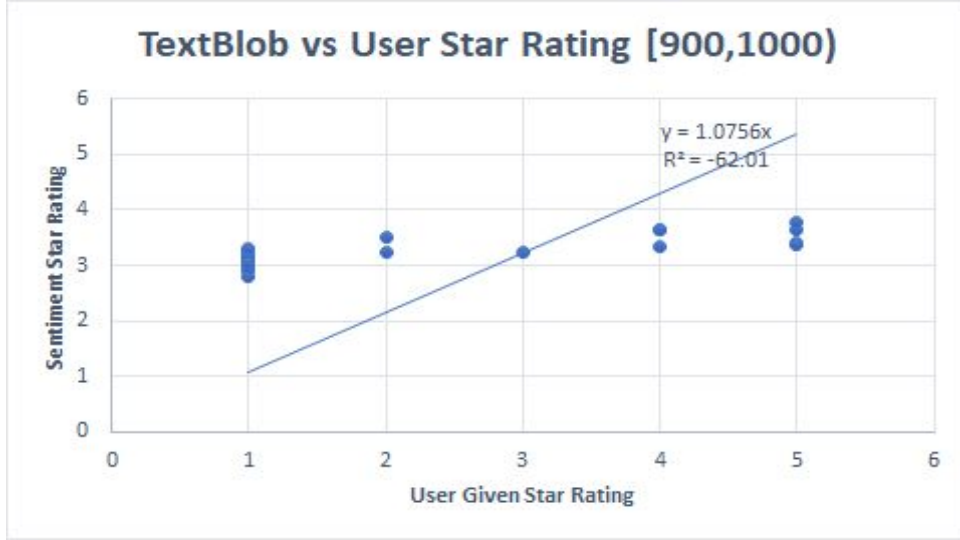
Method

- Gather JSON data from dataset
- Split all text reviews into word size categories
- Employ TextBlob Sentiment analyzer on text review
- Export scaled and user-given star rating to CSV file
- Use MS Excel to conduct a matched pairs t-test

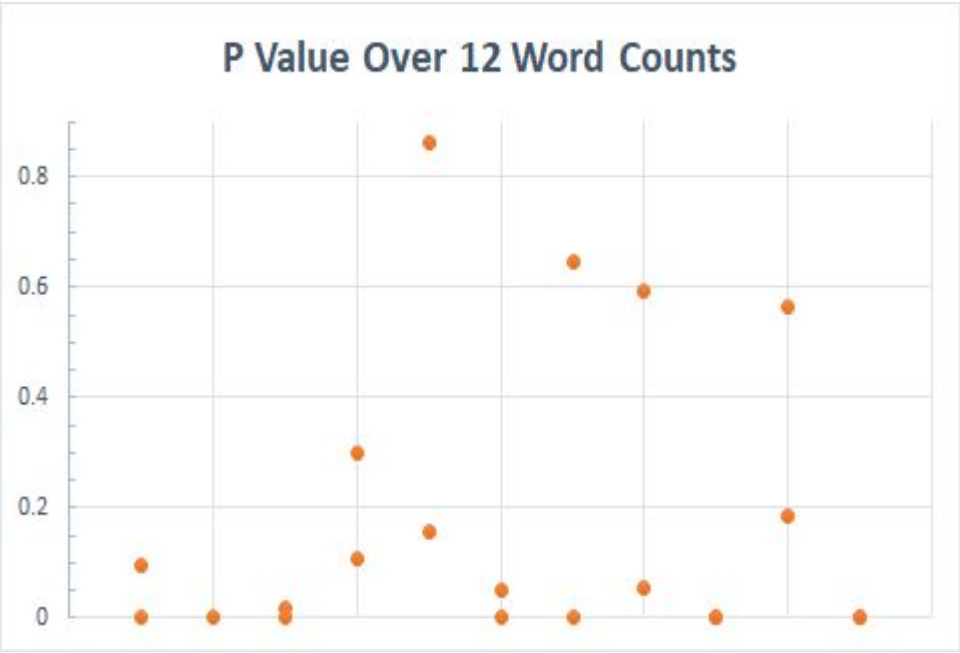

TextBlob vs User Star Rating [0,50) Words
y = 0.8558x
R² = 0.2983
Sentiment Star Rating
User Given Star Rating


TextBlob vs User Star Rating (50,100] Words
y = 0.9059x
R² = -1.927
Sentiment Star Rating
User Given Star Rating


TextBlob vs User Star Rating [700,800)
y = 1.0974x
R² = -44.9
Sentiment Star Rating
User Given Star Rating


TextBlob vs User Star Rating [900,1000)
y = 1.0756x
R² = -62.01
Sentiment Star Rating
User Given Star Rating

Concluding Diagrams



Analysis

- 10/22 of samples had p values less than alpha = .05,
  - 10 times where we could reject the null hypothesis that there is no difference
  - 12 times that we failed to reject that null hypothesis
- Reviews with fewer words trended towards the ideal line of y = x
- Reviews with more words gravitated towards y = 3
- For word counts less than 900, average residual was less than one star

Conclusions

- Reviews with more words may have more complexity
  - Various pros and cons balance out to neutral statement
  - Brief review may just be “I love this restaurant”
- Overall, the p-values did not give us strong enough evidence to assume that there is a significant mean difference
- In the long run, TextBlob appears to work better with less text than it does more text.
- While it is not perfect, it is still an effective tool.