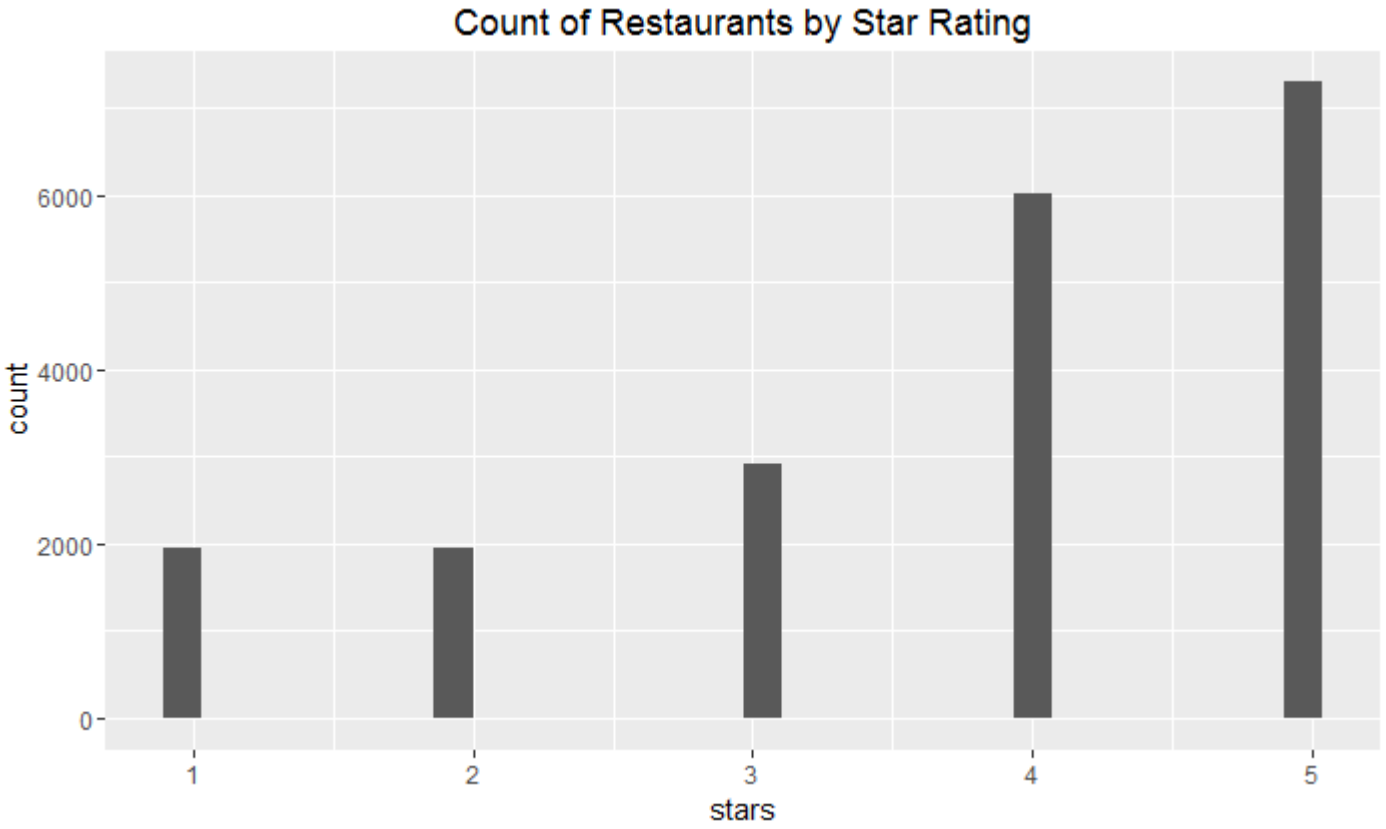
Worked with: David Wang, Terry Wu, Embert Lin

Summary Statistics



- Average Reviews per Restaurant: 2.44 reviews
- Average Reviews per User: 1.16 reviews
- Average reviews for GoodForLunch < Average reviews for not GoodForLunch
  - 2.26 reviews vs. 2.61 reviews
- Average ratings for GoodForLunch < Average ratings for not GoodForLunch
  - 3.72 stars vs. 3.74 stars

Exploratory Text Analysis

food	good	place	great	like	service	just	one
15636	13761	12298	9678	8254	8067	7929	6853
time	get	really	back	will	ordered	restaurant	
6517	6208	5907	5881	4952	4797	4610	

15 most frequent words among reviews

Wordcloud of 100 most frequent words

Text Analytics and Prediction

Unique Terms

- ~45,091 terms
- This was found by removing 99% of sparse terms and assuming that these terms were unique because they appeared in less than 1% of documents
- The difference was 46,016 terms – 925 terms = 45,091 terms

Correlation Wordcloud

Coefficient Wordcloud

At a threshold of 0.5:

Training Classification Accuracy: 64.96%

Test Classification Accuracy: 64.99%

Conclusion:

Training and test accuracy was nearly identical but not great in terms of accuracy

Positive Correlation / Coefficient

Negative Correlation / Coefficient