

A Study Concerning Employee Reviews for Google, Amazon, Apple, Facebook, Microsoft, and Netflix

Table of contents

Business Understanding

- Business objectives
- Project Plan

Data Understanding and Preparation

- Collect Initial Data
- Describe Data
- Clean Data

Modeling

- Early Exploration
- Select Analysis Technique
- Conduct Sentimental and Predictive Analysis
- Trend Analysis

Summary

- The Conclusion from The Analysis
- Business Recommendations

Business Understanding

Business Objectives:

It is an important question to ask when you try to find a new position on your career path: which company within the industry best aligns with your professional goals? In other words, where to send your resume and apply for a job? To facilitate the answer to this question, many candidates reach for the opinions of employees of these companies. Employee reviews give useful insights into the firms and provide valuable suggestions about whether to work for and grow together with the companies. Through employee reviews, we can compare the similarities between these companies, have a better understanding on which company is worth working for, and conduct the common attributes for each company based on their employee reviews.

Project Plan:

There are five steps we need to conduct: 1. collect initial data and have a good understanding of variables and strings; 2. clean the dataset and inspect all missing values; 3. explore the dataset to analyze the correlation between variables and companies; 4. Process more-in-depth analysis using text mining and trend analysis; 5. run linear regression and decision tree models to predict future opportunities for employees.

Data Understanding and Preparation

Collect Initial Data:

We collected the initial dataset from Kaggle.com, an online community to do data science projects. The original employee reviews' dataset was scraped from Glassdoor, a website where current and former employees anonymously review companies and their management. We downloaded the dataset from Kaggle and prepared to analyze it deeply.

Describe Data:

The dataset contains over 67k employee reviews for Google, Amazon, Facebook, Apple, Microsoft and Netflix, and there are 16 variables: *index*, *company name*, *company location*, *date posted*, *job-title* (this string also includes the job status of the employee, such as current and former employees), *summary*, *pros*, *cons*, *overall rating*, *work/life balance rating*, *culture and values rating*, *compensation & benefits rating*, *senior management rating*, *helpful review count*, and *link to review*. With all six firms, Amazon had the most significant portion of employee reviews with 26430 reviews in the dataset; Microsoft kept the second place with 17930 reviews; Apple contained 12950 reviews, Google had 7819 reviews, Facebook had 1590 reviews, and Netflix has 810 reviews.

Clean Data:

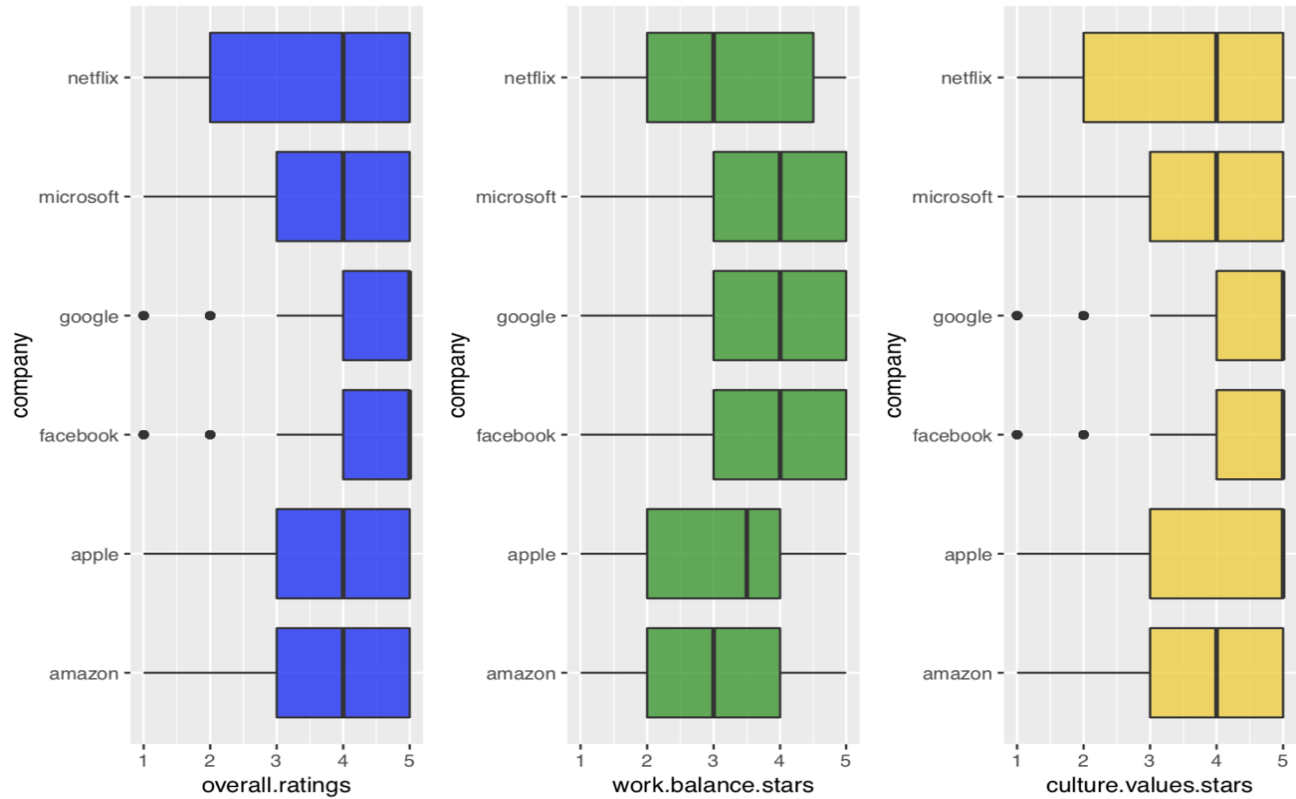
The first step of data cleaning was to transform rating variables into numeric variables and replace NA value by the phrase “none”, as there are six rating categories such as *overall rating*, *work/life balance rating*, and *senior management rating*. Then we noticed that in the variable *job-title*, it contained separate information: position and employment status, we split the *job-title* column into status and position and dropped *job-title*. Also, we transformed dates column into date-type data and company name into factor type. Lastly, we exported the dataset as the cleaned dataset for future analysis.

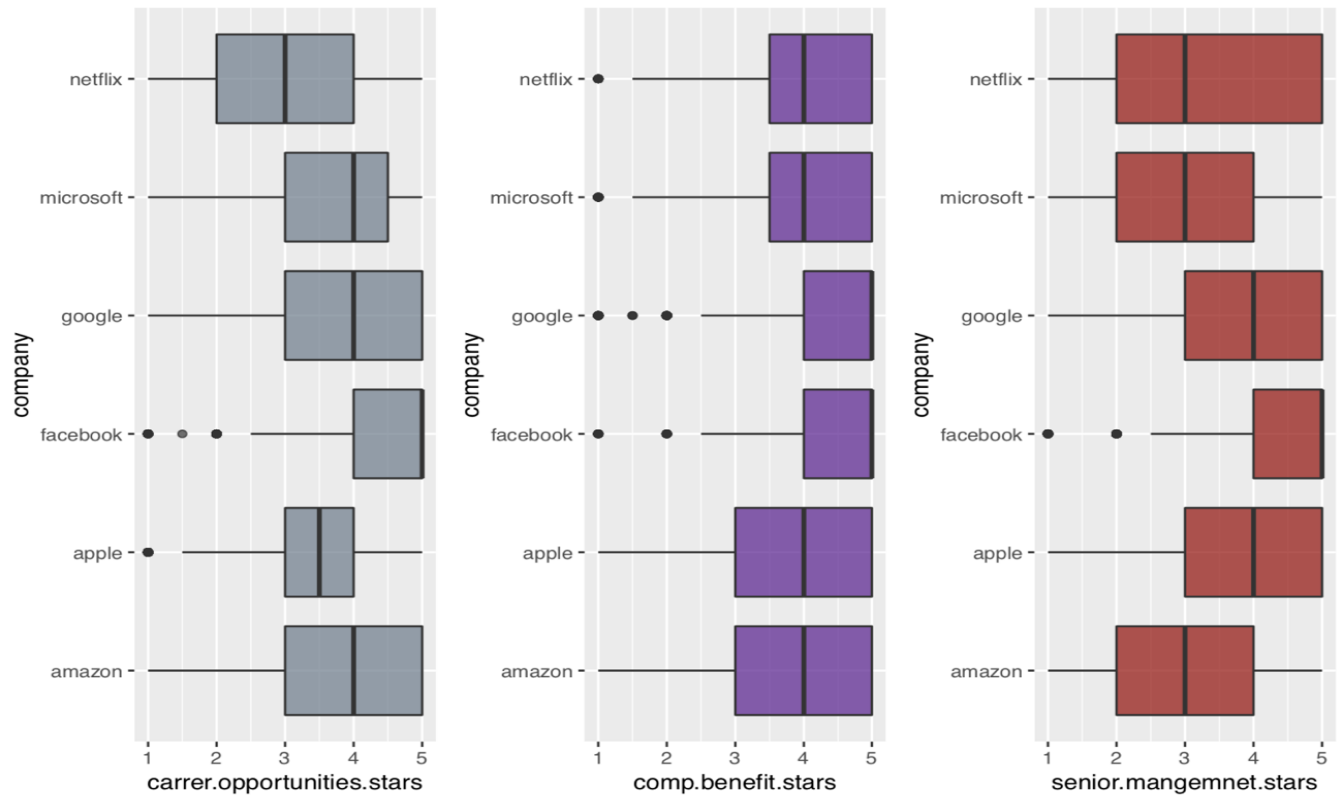
Modeling

Early Exploration:

We began our exploratory analysis looking at the comparisons between these six firms according to the numeric variables from the dataset to see the similarities and differences. An analysis of the mean score for each numeric category in each firm revealed that Facebook had the highest average rating scores in *overall rating*, *culture values*, *career opportunity*, *company benefits* and *senior management* at 4.51, 4.51, 4.36, 4.56 and 4.26 accordingly. Google held the highest score in *work balance* at 4, where Facebook kept the second highest score at 3.93. On the other hand, Netflix had the lowest average rating scores in overall rating, culture values, career opportunity, and company benefits at 3.41, 3.01, 3.52, and

3.06 correspondingly. We utilized ggplot2 and boxplot to compare the rating scores in different categories according to the firms' employee reviews and visualized the result as follows:





Furthermore, we conducted the correlation plot between these rating variables and it revealed that the correlation between *overall rating* and *culture value rating* was the highest at 0.76, as was the correlation between *senior management rating* and *overall rating* (0.73). The correlation between *overall rating* and *career opportunities rating* was also higher than others (0.69). It revealed that there was a positive correlation between *overall rating* and *culture values rating*, *senior management rating*, and *career opportunity rating*. Our analysis of the correlation between overall rating and work balance rating and comp & benefit rating implied that there was a low correlation between these variables.

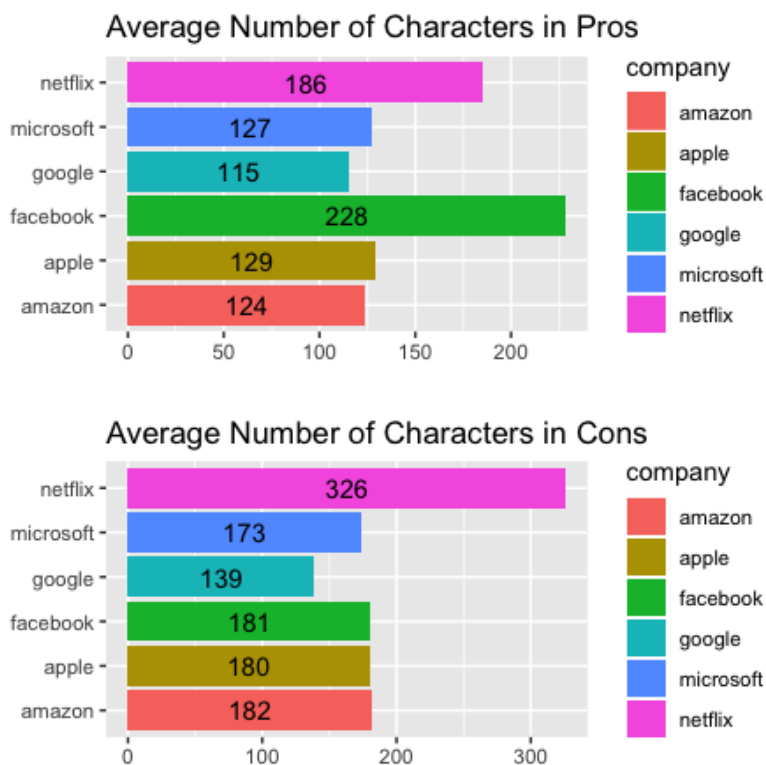
Select Analysis Technique:

We selected text mining and trend analysis to build the dataset models. The main purpose of text mining analysis is to compare the differences and similarities of employee reviews from different aspects among six companies. In addition, we can

Conduct Sentimental and Predictive Analysis:

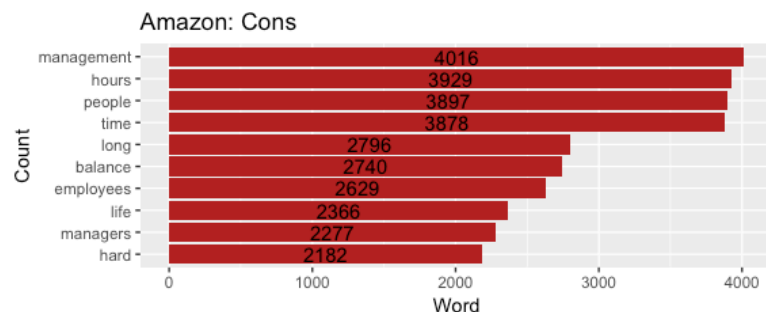
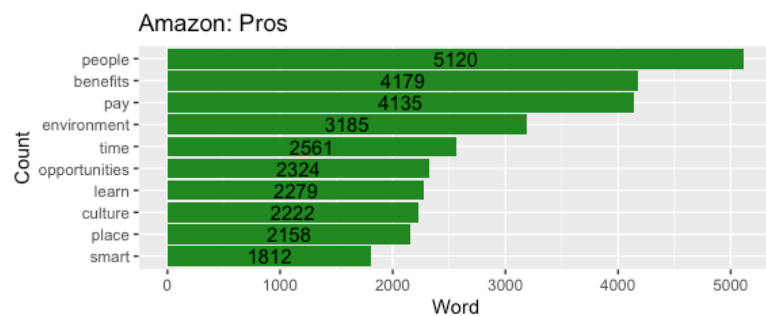
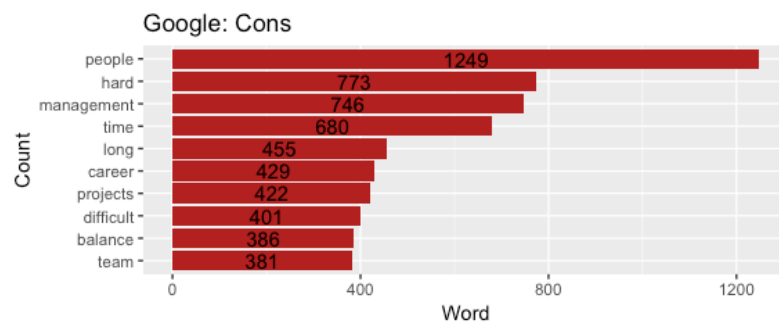
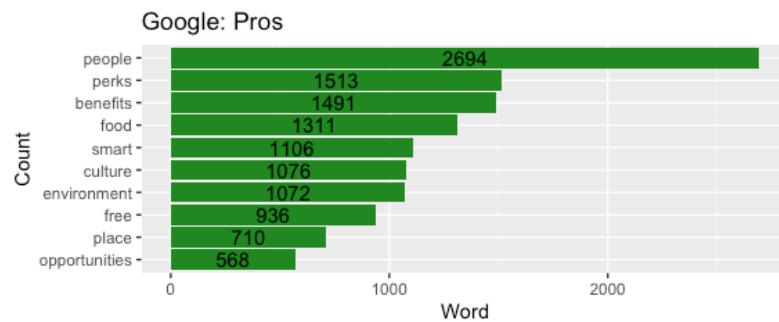
Once our dataset was sufficiently clean for analysis, we began text mining analysis for descriptive variables including *pros*, *cons*, and *summary*.

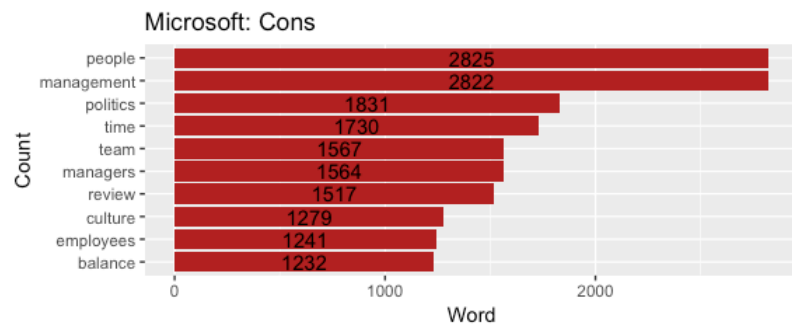
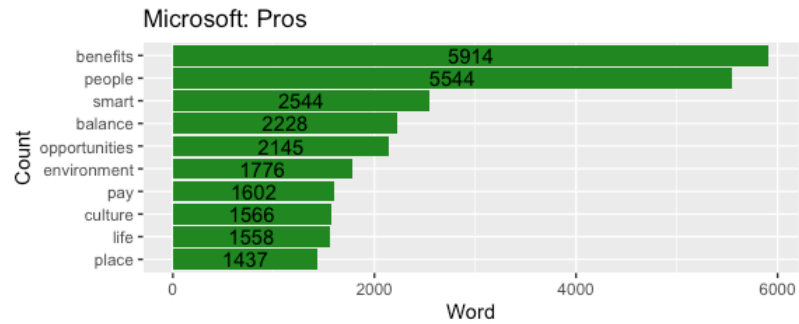
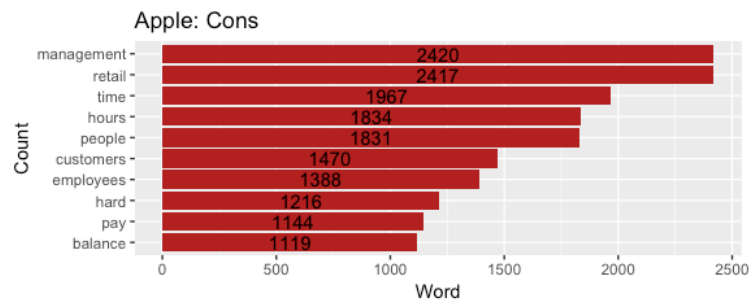
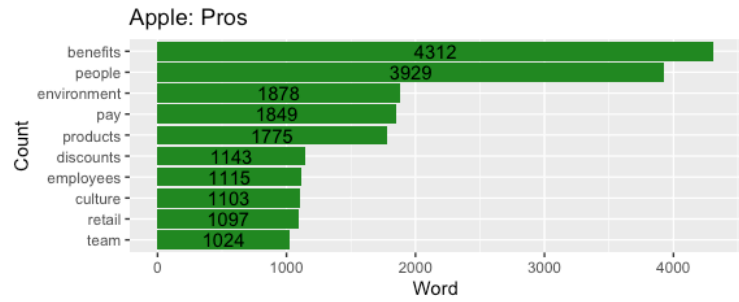
We want to take a look at the number of characters in descriptive variables such as *pros* and *cons* among six companies and we conducted the following graph. It visualized that Facebook has the highest number of words in the *pros* category whereas Netflix has wordiest employee comments on firm's negative aspect. It revealed that employees tend to describe more specific and provide more personal opinions on the firms' advantages and disadvantages. Also, we found that Facebook and Netflix contain the highest number of comments in *Pros*, indicating that there is a possibility that they have better performance on employee satisfaction than other firms. While Netflix has the highest number of characters in *Cons*, it revealed that employees tend to complain about the firm more than other companies.

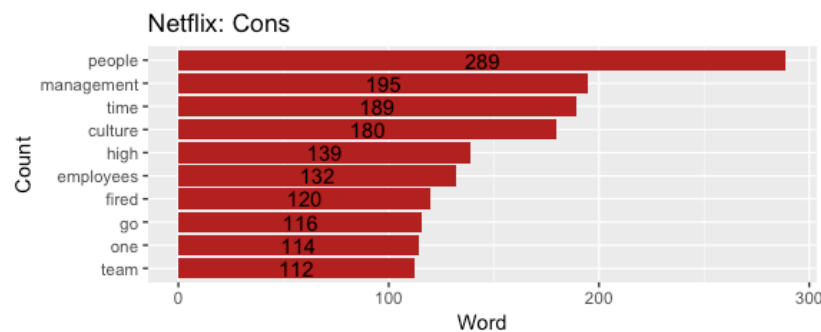
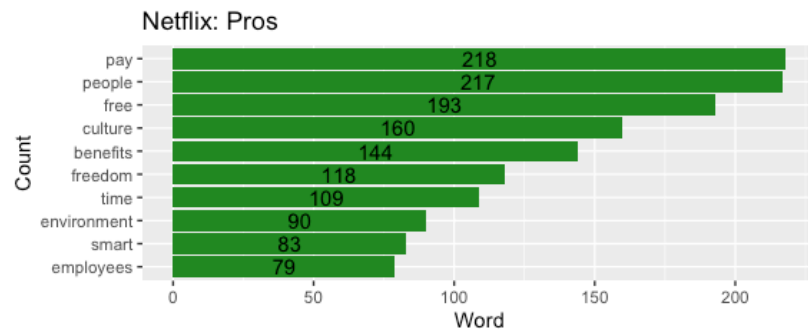
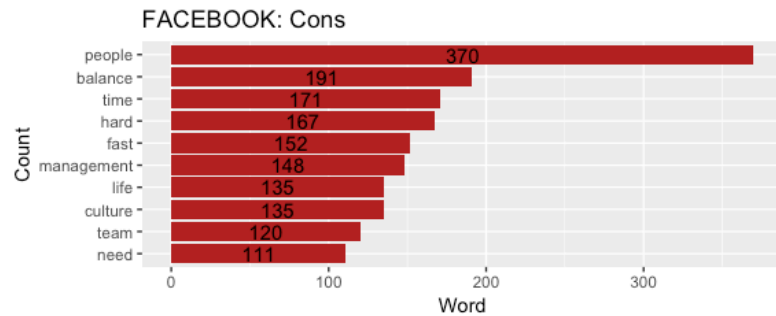
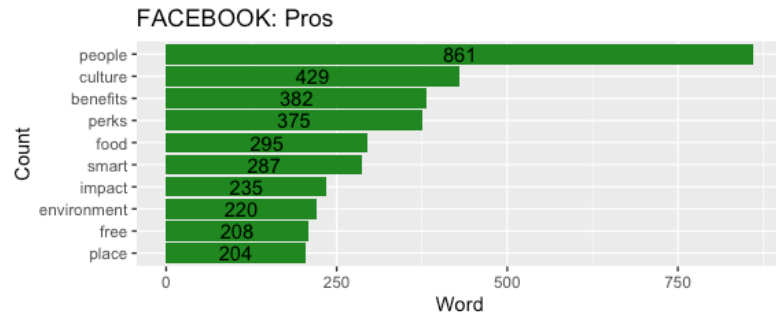


These connections reinforced us to calculate the correlation between the length of the reviews and overall ratings. Not surprisingly, there is a positive correlation between the number of characters in *pros* and overall rating at 0.0436 and a negative correlation between the number of characters in *cons* and overall rating at -0.294. (as shown in R markdown)

Next, we analyzed the descriptions in *pros* and *cons* variables using sentimental analysis to determine what keywords were used most frequently to describe these six companies. The top ten words in each category for each firm are shown as follows:







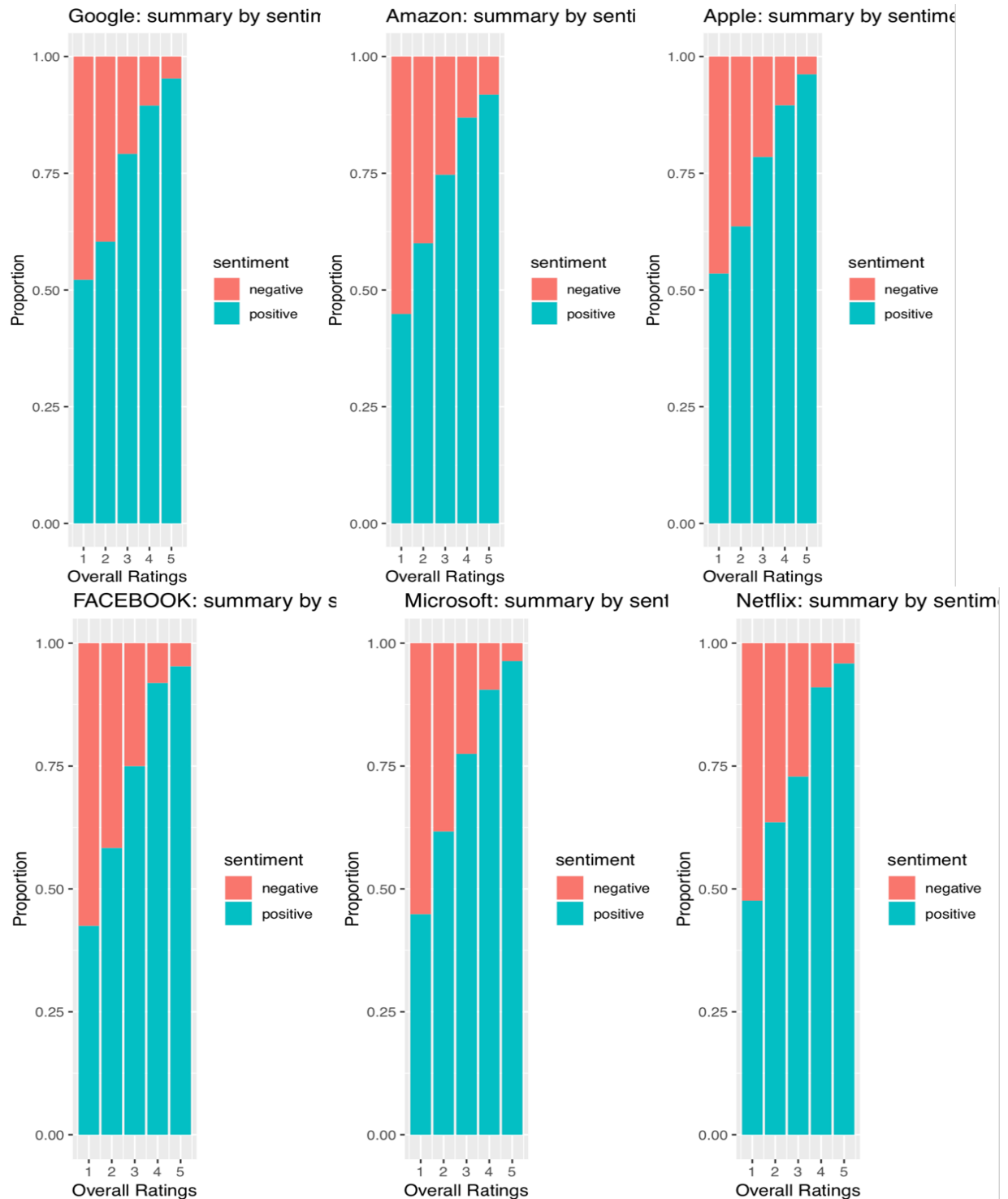
From our analysis, we observed that the term “environment” ranked higher than its ranking in *Pros* in other firms, indicating that Apple and Amazon have a better working environment; In addition, Netflix had the potential highest pay among six companies. The term “smart” was mentioned in Microsoft and Google referring that employees valued their coworkers’ intelligence. Amazon and Microsoft provided more “opportunities” for their

employees which might bring possible career opportunities along the way to success.

In terms of *Cons*, employees at every company complained about the management, while the rank of “management” for Facebook was much lower than others. So we could presumably assert that though poor management is a common problem for all six companies, Facebook adopted a more advanced management system with higher employee satisfaction. The term “fired” only appears in the top 10-word list of Netflix, which showed that dismissal was a serious internal problem of Netflix, and it could also explain why Netflix was such a controversial company with both longest descriptions in *Pros and Cons*. “Politics” was ranked the 3rd in *Cons* of Microsoft, which indicated that the employees complained a lot about the workplace politics within the company involving unbalanced power and authority that prevents a fair working environment.

In addition, we designed a Wordcloud to visualize the character’s frequency in each firm for *pros and cons* variables, and we noticed that the phrase “people” showed up everywhere as one of the most frequently used words in all six firms. It is a sign from the employee perspective that people matter the most when they think about the firms’ advantages and disadvantages.

The second part of text mining is a sentimental analysis. We were curious about the relationship between positive/negative words in the *summary* and *overall ratings*. We divided the description words in *summary* variable into two fractions: positive and negative, and we analyzed the correlation between the proportion of these two sentiments and *overall ratings* variable among six firms. The graph shows as follows:

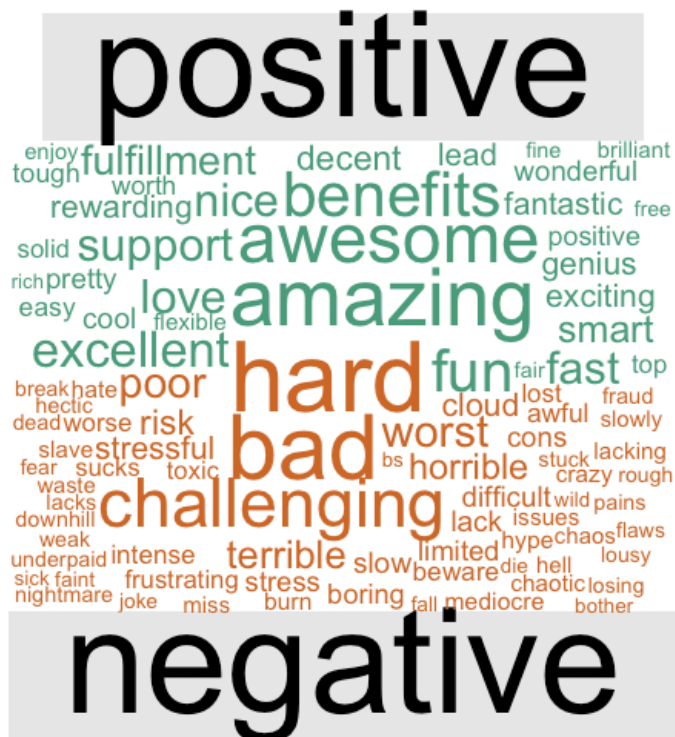


From the graph, we could tell the proportion of positive words in *summary* is positively correlated to the overall ratings, and the proportion of negative words in *summary* is negatively correlated to the overall ratings. It revealed that employees' sentiment matters the most when they evaluate the firm in general. We generated deeper thinking that in *pros* and *cons* variables, the

statistic showed the term **people** was one of the most frequently used words, and we connected the common feature both in the analysis of sentiment and character of frequency, there was a possibility that **people** were the key factor of employees' sentiment. The positiveness of people around employees may increase the fraction of positive sentiment.

Furthermore, we measured the list of words in *summary* variable and their associations with eight basic emotions (trust, anticipation, joy, surprise, fear, anger, sadness, and disgust) and two sentiments (positive and negative) using NRC lexicon. The majority of *summary* was largely positive according to our graph which plotted the positive and negative sentiments for summary of all six firms.

We then analyzed the descriptions in *summary* variable to determine what terms were used most frequently to describe the firm, and the WordCloud nicely illustrates the specific words employees used both in positive and negative aspects. The top five descriptive words in the positive side were amazing, awesome, benefits, fun, and excellent, whereas the top five descriptive words in the negative side were bad, hard, challenging, worst and terrible. The analysis revealed that descriptive words kept consistent with two sentimental categories.



Trend Analysis:

After the text mining analysis, our next step was to use trend analysis to estimate the change of *overall ratings* over years for each company. We separated the employees' current job status from the *job-title* variable into two categories: current employees and former employees, and we analyzed the correlation between the *overall ratings* and employees' job status. The results are as follows:

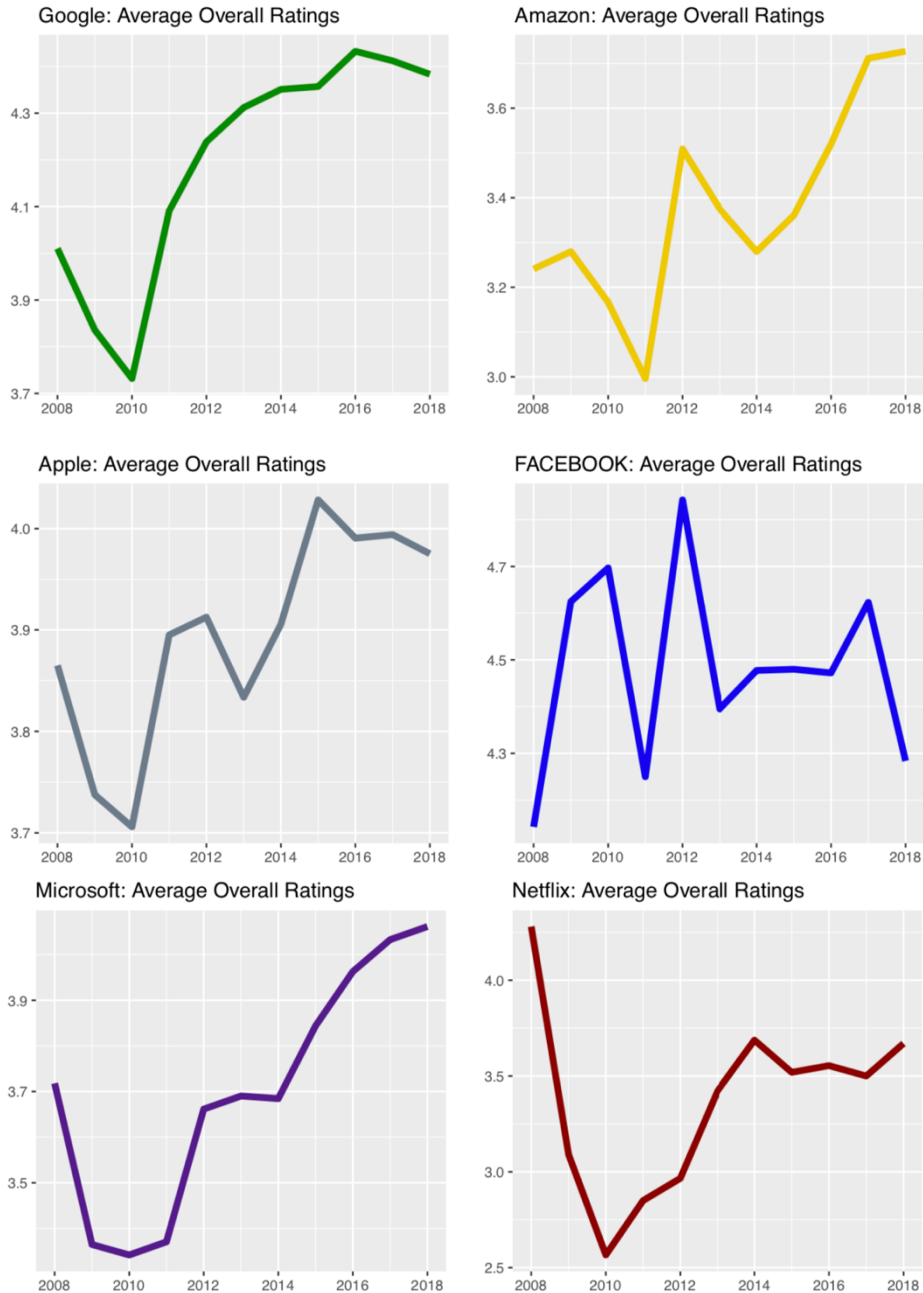
	Google	Amazon	Apple	Facebook	Microsoft	Netflix
Current employee	4.43	3.80	4.06	4.67	3.91	3.93
Former employee	4.21	3.17	3.83	3.87	3.65	2.9

There is a clear indication that current employees rated their firms higher than former employee which is not surprising to us since former employees had their reasons to leave the firms. Google, Amazon, Facebook, Apple, Microsoft, and Netflix are the most well-known companies worldwide and yet their overall ratings from former employees' perspective still have some rooms to improve.

While interesting, we started to analyze the *overall ratings work balance stars, culture values stars, career opportunities stars, company benefit stars, and senior management stars* trend over years for each firm. The package Lubridate was used to track the change and visualize the fluctuation over time. In Google, the trends of all variables were on the rise from 2008 to 2018 except for *culture values stars*. Additionally, the fluctuations of curves were bigger in the *work balance stars* and the *culture values* than other variables. It revealed that Google employees' satisfaction on work balance and culture values were decreasing in the last ten years. All of the *ratings* variables in Amazon has decreased to bottom and recovered to its peak over time, and the trends in general is on the increase except for *work balance stars*. Microsoft has similar trends with Amazon, yet the *company benefit stars* have decreased over the decade with great fluctuation. On the other hand, Apple and Facebook have some noticeable decreases in *work balance stars and culture values stars*, and other categories have wide fluctuation from 2008 to 2018. Comparing with the other five firms, Netflix performs the poorest especially in *overall ratings and senior management stars*. Netflix reached its perk in 2008 in almost all

categories and flopped dramatically throughout these years, which indicate that Netflix employees' dissatisfaction was getting stronger and stronger. It brought our interests that all rating scores had reached its bottom at the year 2010, the potential cause might be the political and economic environment which need further investigation.

Below is a graph visualization illustrating the change of *average overall ratings* from 2008 to 2018 among six firms. The fluctuation of the curve for each firm is relatively big, and the *overall rating* for the majority of firms are on the rise except for Netflix. It indicated that these firms are making an improvement to raise the satisfaction of employees. Specifically, Google, Amazon, Apple, and Microsoft are making stronger effort comparing to Facebook and Netflix.



Summary

The conclusion from the analysis:

Based on our analysis on the dataset, we had some interesting findings on employees' reviews on their company.

1. Facebook and Netflix have the longest descriptions in Pros while Netflix has the longest descriptions in Cons. It shows Facebook

would have more potential benefits for the employees, while Netflix is a very controversial company with both good benefits and unsatisfied employee experience.

2. After text mining the Pros and Cons of the six companies, we discovered that Apple and Amazon have a better working environment, Netflix has a potential higher payment, Microsoft and Google tend to hire more intelligent employees, and Amazon and Microsoft provide better career opportunities for the employees. Besides, management is an overall problem for all six companies, and Facebook tends to have a better management system. For Netflix, the dismissal problem was severe and has led to more complaints from the employees. Microsoft has a severe problem on workplace politics, and this could potentially lead to employee disappointment and generate obstacles for the company's long-term goals.

Business Recommendations:

For the six companies:

1. Facebook, Netflix, Microsoft, and Google should improve their working environments for competitive advantages.
2. All companies should adopt more advanced management system in order to optimize employees activeness and potentials.
3. For Netflix, the company should better leverage employment management and make people less intense by reducing the working pressure and improve their work-life balance.

For job seekers:

1. The Job seekers can refer to our results or use our developed models to gain a clear understanding of the culture, value, benefits, environment, and other related factors of each company. It is essential to find the most appropriate company to fulfill personal goals.
2. If you prefer environment-oriented companies, Apple and Amazon are your best options; for compensation-oriented companies, Netflix is the best place to work; Amazon and Microsoft provide more career opportunities than other firms, and these are the better place to enlighten your future career path; Google and Microsoft can provide a educational environment for job seekers who would like to learn and grow quickly from colleagues; Since Microsoft has potential office

politics problems, job seekers should think twice if equality and transparency are their top concerns for choosing employers.