

Project Deliverable 2: Employee Reviews

Team Chicken Dinner: Xinying Xu, Qian Zhao, Jiayi Jenny Lu, Junwei Xu

I. Problem Statement

Many people in the world face questions such as where to send the CV? which company is a better choice for me? etc. To help answer these questions, opinions from both former and existing employees of those companies are gathered.

The dataset we used contains over 67k detailed employee reviews of Google, Amazon, Facebook, Apple, Microsoft, and Netflix, which was web scraped from Glassdoor. We found this dataset on Kaggle.

By analyzing the ratings and review contents, we wish to reach the answer for some major problems, including:

1. What is the rating trend of different companies from 2009 to 2018?
2. What company performs best given the ratings from employees? What is the correlation between overall ratings and detailed ratings such as work balance and culture values? And what is the rating that values most for employees?
3. The driving reasons that lead employees to resign and how could the company improve itself to control the turnover rate?
4. What are the important words in the pros and cons that companies should pay attention to? And are there overlaps between the advice to management and negative comments from former employees? Therefore, what companies can do to improve those factors?
5. Do the negative reviews tend to be longer than positive ones and is it a trend that people usually complain more?
6. What are people's sentiments in the summary of their reviews?

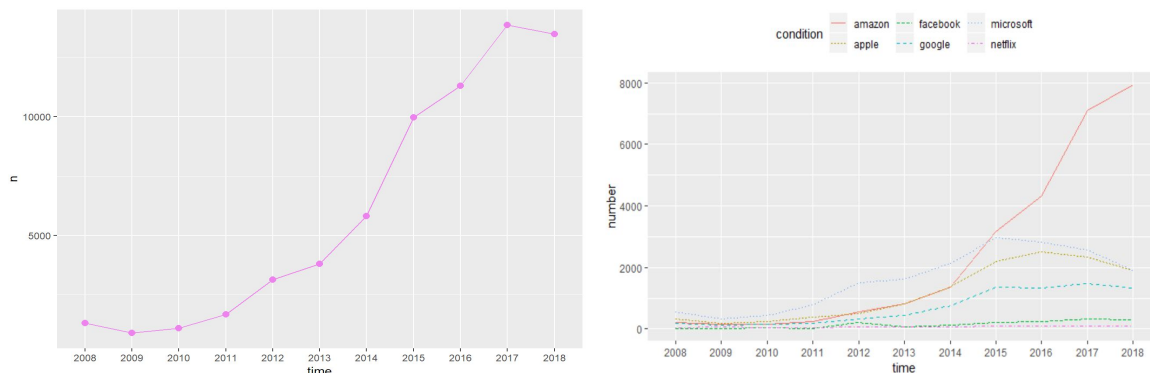
II. Analytical technique

To start with, we explore the dataset before deciding how to analyze it. Since the dataset contains both numeric and text data format, we hope to use as much information as possible, so that we could have a better understanding of the employees' opinion towards their companies. We use the Tidyverse, which gives us many advantages to explore the data. Moreover, we mainly use four analytical techniques including data visualization, feature selection, text mining and sentiment analysis. We used `na.omit`, and `filter` to clean the data, and use `summarize`, `select`, and various verbs to manipulate data. Feature selection is essential in the beginning of exploratory

analysis, because it gives us hints of which variables we should mainly select when we analyze the overall ratings. Data visualization is a straightforward way for us to visualize complex data and see the changes of past 10 years reviews. Besides, with the graphs and plots of ratings rank from different companies, we can summarize the information, generate the conclusion, and make corresponding adjustments. Text mining is definitely the key technique of our dataset because employee's reviews from six companies contains a large number of text data in different variables such as pros, cons, and summary. Text mining not only helps us obtain employees' information from words and better understand which company attracts employees more, but also gives us the reasons of employees' leaving. Sentiment analysis is also vital when we analyze the summary. It gives us employees' overall feelings of the company by comparing positive words and negative words as well as the emotions employees put in text variables.

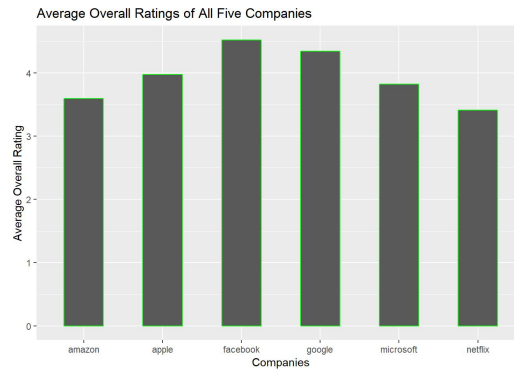
III. Analyze & Explore

Q1. The trends of numbers of reviews for companies in Glassdoor and why?



It is well-known that getting employees to provide feedback on their companies could help build an employer brand and image efficiently, and help increase employee engagement. Thus, we first explore the trends of the reviews' quantity in Glassdoor. The number of employees involved in the feedback is steadily increasing in the past years, which may be largely due to the effort of companies. The comments on Glassdoor for companies include both positive and negative contents. The cons do not necessarily mean a bad thing for companies. Because the transparent nature could help outsiders, mainly the potential candidates to better understand the company, and help the management team improve the organization to better retain the talents.

Q2. What is the correlation between overall ratings and detailed ratings such as work balance and culture values? And what is the rating that values most for employees?

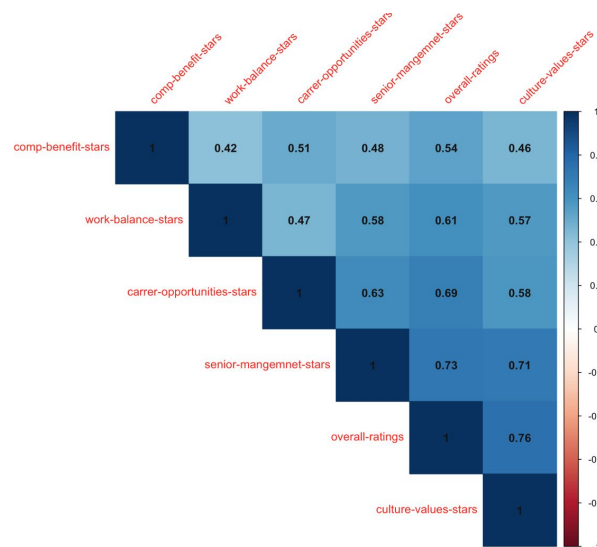


Overall rating is the most general assessment of the company from its employees, taking all aspects into account.

Facebook is rated overall the best company, followed by Google and Apple, with a score higher than or around 4.0. Netflix gets the lowest score from its employees.

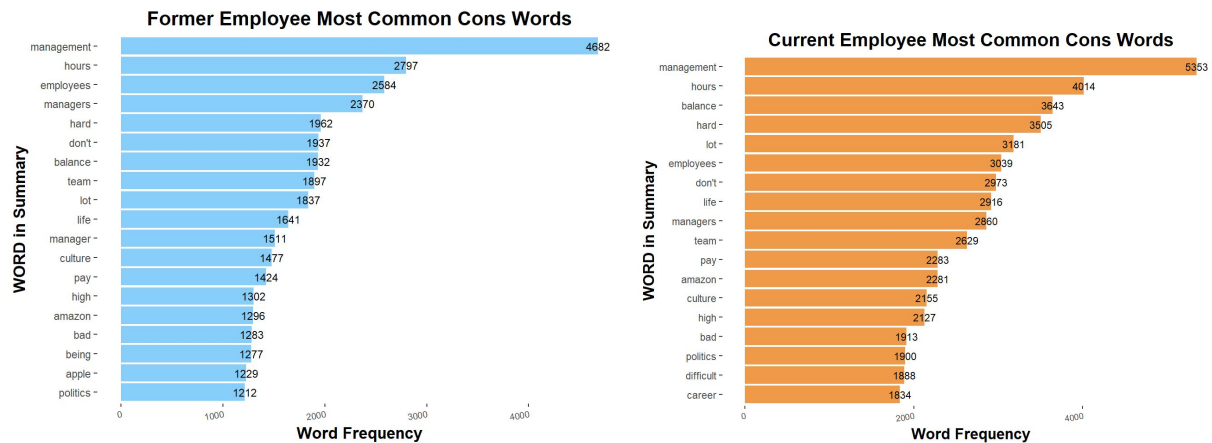
From the boxplot, both Facebook and Google get 5/5 in the median, others are just 1 point behind. The rating data of Netflix has a higher discrete degree than others, which may imply that employees attitude towards Netflix varies a lot.

In general, overall ratings are really good for these companies.



From the plot above, the overall rating is most closely connected with the culture-values (0.76), following with senior-management-rating (0.73). Together with the cons words which we analyzed before, we may speculate that employees pay more attention to a company's culture as well as their managers, which may indicate a company's organizational structure.

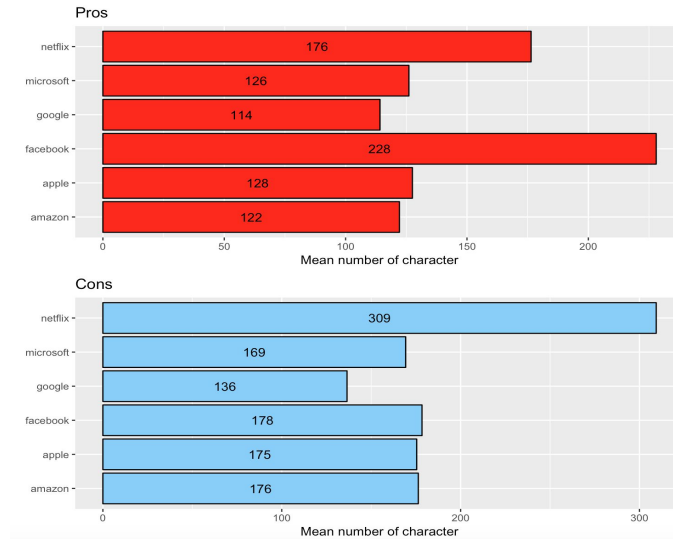
Q3. The driving reasons that lead employees to resign and how could the company improve itself to control the turnover rate?



By analyzing the cons in reviews from both the former and current employees' perspectives, 'management,' 'hours' are most frequently mentioned words. The third place's "manager" is replaced by "balance," which may indicate an improvement in the organizational structure, mainly the relationship between managers and employees, but more stress from work. "Life," "team," "pay," "culture," "politics" are still top words. Another interesting point is that employees used to mention Apple a lot in their reviews, which is replaced by Amazon from current employees' answer.

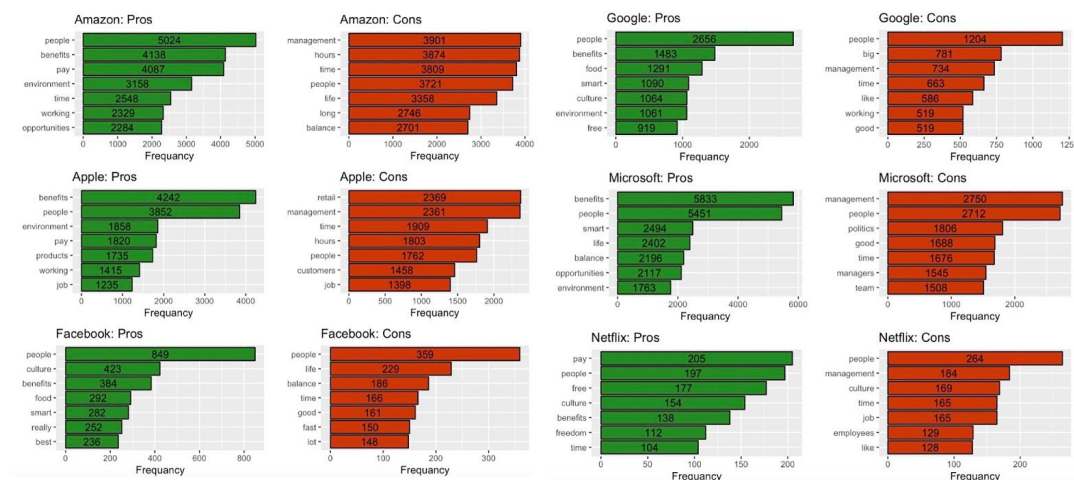
By analyzing the cons in reviews from both the former and current employees' perspectives, 'management,' 'hours' are most frequently mentioned words. The third place's "manager" is replaced by "balance," which may indicate an improvement in the organizational structure, mainly the relationship between managers and employees, but more stress from work. "Life," "team," "pay," "culture," "politics" are still top words. Another interesting point is that employees used to mention Apple a lot in their reviews, which is replaced by Amazon from current employees' answer.

Q4. What are the important words in the pros and cons that companies should pay attention to? And are there overlaps between former employees' advice to management and current employees' advice to management? Therefore, what companies can do to improve those factors?



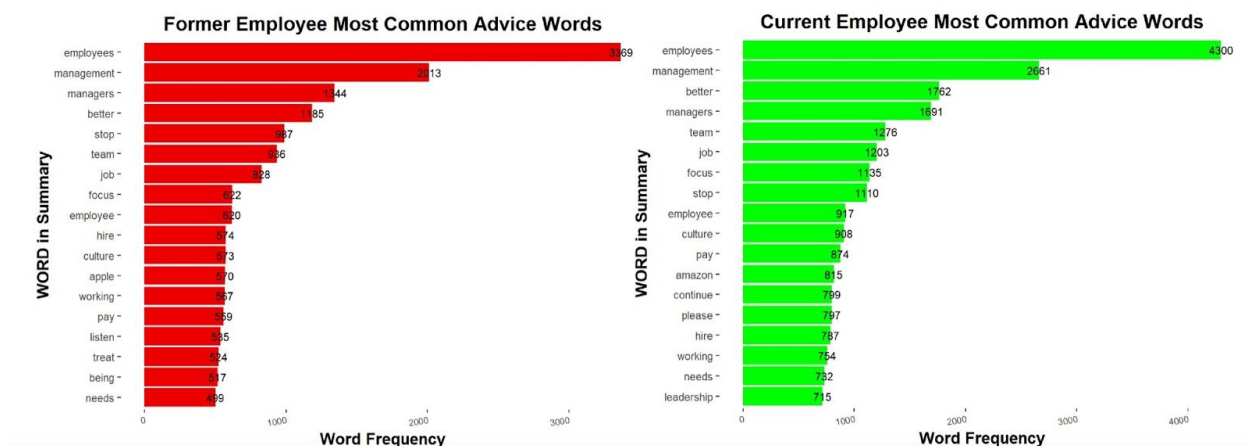
In order to answer the question, let's first explore the lengths of pros and cons. From the plot, we can depict that Facebook has the longest number of characters in pros, while Netflix has the longest number of characters in cons. Since Facebook is best rated, it's no surprise that employees can indicate many advantages. Netflix gets a lot of complaints from its employees and it's almost two times the number of characters in pros for Netflix. From the previous analysis, Netflix does not get a satisfying rating, which explains the reason that Netflix employees complain more. Another interesting result is that Google employees' reviews of pros and cons tend to be shorter than other companies', which may be relevant to the culture of the company.

After exploring the lengths of pros and cons for different companies, we come to the part of finding the most popular words for each company. We ruled out words such as company name, "work", "can", etc., because these words don't give us valuable information. Below are the plots of most frequent words in pros and cons for each company.



These plots can help companies understand what factors employees value the most and what area needs to be improved. For example, Amazon's benefits and pay are satisfying and the people there are also friendly. But Amazon should pay attention to its management and working hours, since employees have a negative attitude towards these aspects. The words "people", "benefits" and "environment" frequently appear in pros for most companies, which means these companies do a great job regarding these factors. However, a common concern of "management" and "time" from employees indicate that these companies still have space for improvement. With the help of the plots, companies can understand their employees better and reduce the employee turnover rate.

Next, we look at the advice to management from current and former employees and find out whether there are overlaps between them.



From the plots, we understand that employees are looking for better management, and they value teamwork and company culture. Since former employees advised management to be improved further and it's possible that it's one of the reasons they left, companies should pay much attention to those areas and make sure the same mistakes not happen again.

Q5. Do the negative reviews tend to be longer than positive ones and is it a trend that people usually complain more?

Following Question 4, we can further explore whether negative reviews tend to be longer than positive ones. The previous analysis gives us a general idea that negative reviews are longer than positive ones, and to ensure the finding, we run the following

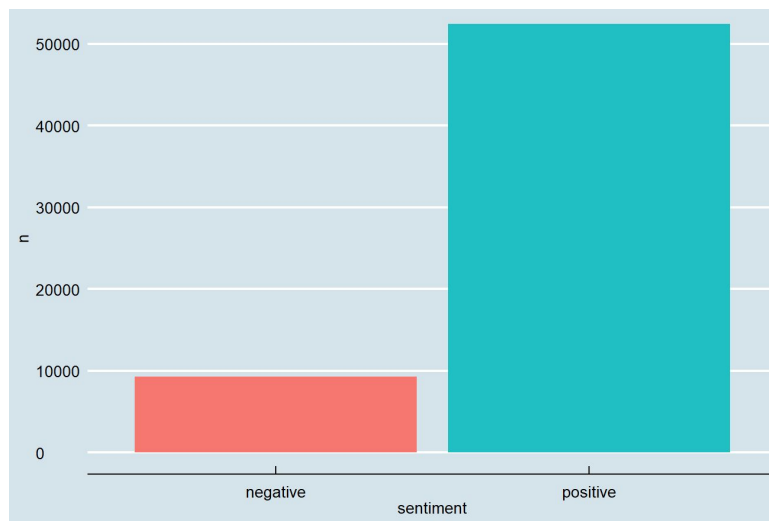
codes: `mean(nchar(mydata$pros))`, `mean(nchar(mydata$cons))`. From the result, we are certain that people usually complain more when it comes to certain problems.

Besides, are longer reviews usually associated with lower ratings? To answer this question, we need to find the correlation between overall ratings and the number of characters of cons. The code is:

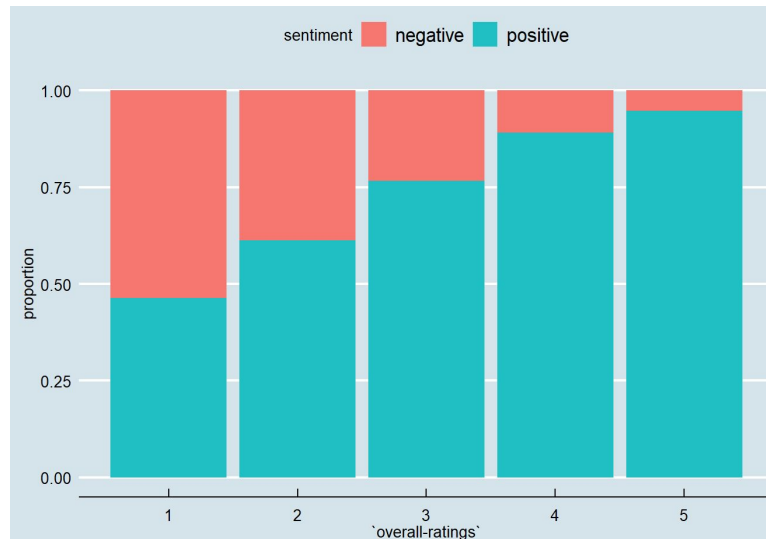
`cor(nchar(mydata$cons),mydata_clean$`overall-ratings`)` and result is -0.30. Therefore, there is a negative correlation between overall ratings and the number of characters of cons and it proves that longer reviews are usually associated with lower ratings. From the findings, we understand that people usually complain more regarding certain problems, which is normal because it's a common habit of people. This means companies should pay more attention to negative reviews as they show dissatisfactions from employees.

Q6. What are people's sentiments in the summary of their reviews?

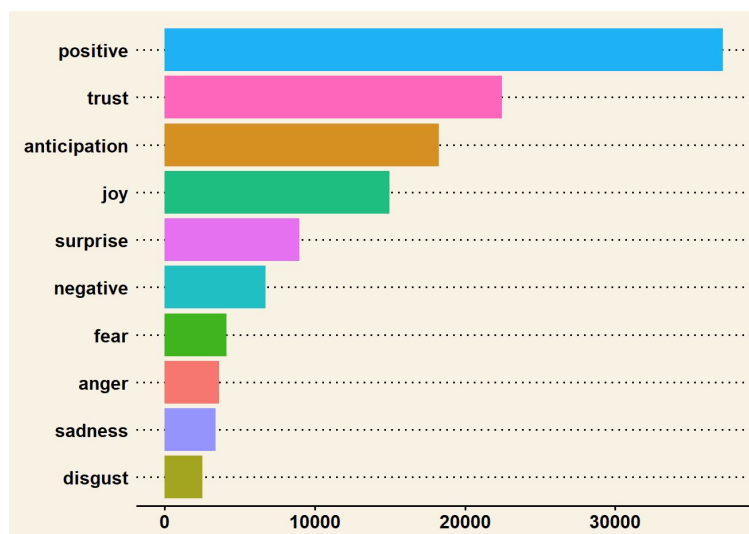
To answer this question, let's first take a look at the number of positive and negative words in the summary for all companies. From the chart below, employees are satisfied with their companies, with more than 50000 positive words in company reviews.



Furthermore, since there are a large number of positive words in the summary, it's not hard to think about the helpfulness of those positive words in rating the companies. From the chart below, we can depict that positive words take account of more than 50% of words in the summary when ratings are more than 2 stars. The fact that there are more positive words in the summary for a rating of 2 makes us doubt the helpfulness of positive words in rating the companies. But there is a possibility that employees add negative words such as "no" and "not" in front of the positive words, and that's the reason that these positive words show up many times in the rating of 2.



After exploring the positive and negative words in the summary, we come to the next step of using “NRC” lexicon to detect the emotions of different words in the summary. The plot below gives us information about employees’ emotions regarding their companies. In general, employees expressed good emotions in the company reviews, which means they are satisfied with their companies. The top five words in the plot are “positive”, “trust”, “anticipation”, “joy” and “surprise” and it shows that these companies build a relatively good relationship with their employees.



For the next part of sentiment analysis, we explored the correlation between the emotions expressed and overall ratings. When the ratings are high, there are strong correlations between good emotions such as “positive” and overall ratings. On the other hand, bad emotions do not have an obvious correlation with overall ratings.

IV. Conclusion & Recommendation

Employer branding is not just a concept used during the recruiting procedure, but also a key component to engage and retain the right talents.

From the rating perspective, Facebook is the top popular company among employees in Glassdoor. It does take a scientific approach to manage the work experience from every aspect so that it could ensure the employees not only enjoy their work but also perform the best of their abilities. Successfully to deliver a consistently positive and distinctive employment experience is an important tool to keep the talents and thus, help these tech companies achieve great innovation.

By using text mining to analyze employees' reviews for the company Google, Microsoft, Netflix, Amazon, Apple and Facebook, we can conclude that employees value benefits, work environment, company culture and management performance. Therefore, companies should pay more attention to these factors and improve employee satisfaction. When employees are satisfied with their companies, they will be more devoted to their work and improve work efficiency, which will lead to the success of companies.

Furthermore, with the help of sentiment analysis, we have a general idea that employees are relatively satisfied with their companies, with over 80% of positives words in employee reviews. And there is a strong correlation between good emotions and overall ratings.

In conclusion, we used data visualization, feature selection, text mining and sentiment analysis to analyze employee reviews for Google, Microsoft, Netflix, Amazon, Apple and Facebook. It is insightful for companies to understand their employees and amend their weaknesses. Employees will be more motivated and inspired to work for the company, therefore, company performance will be much better.