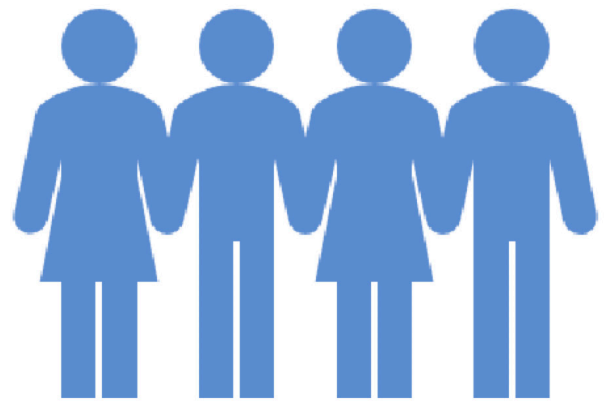# HAPPY PEOPLE

Anna Brown, Josh Chung,
Kim McKeever, Alex Snyder

Part 1: How do various rating categories impact an employee's overall opinion of their company?

Glassdoor allows its reviewers to rate their companies from 1 to 5 stars on six categories. These are company culture, work-life balance, career opportunities, senior management, benefits, and an overall rating. The overall rating is determined independently from the other five categories. We were curious as to how much each subcategory factored into a reviewer's overall rating of the company. To do this we employed a linear regression model using the scikit-learn library.

To clean the data and prepare it for the scikit-learn library we put each of the six variables of interest into a data frame. We then converted the "none" ratings into NaN placeholders. We then dropped these from the data frame. This removed approximately 1,500 data points leaving approximately 5,300. We then converted the data points from strings to floats. We put the clean data into a new data frame. For the analysis we imported:

- from sklearn.preprocessing import PolynomialFeatures
- from sklearn.linear_model import LinearRegression
- from sklearn.pipeline import Pipeline

We began by conducting a linear regression with a single variable, culture. The result was a constant of 1.2662, and a coefficient of 0.68. Thus, based on this data, starting at a base rating of 1.2662 as the culture rating increases by one star, the overall rating can be expected to increase by .68 stars on average. This model has a decent R-squared value of 0.575. This means that the model explains about 57% of the data's variability around the mean. The results are statistically significant at the 99.9% level.

We then used a multivariable linear regression model to control for work-life balance, career opportunities, senior management, and benefits. When controlling these other variables, the constant decreases to 0.6037 and the culture coefficient decreases to 0.327. For reference, the coefficients for the other four variables were work-life balance (0.1422), career opportunities (0.2558), senior management (0.0882), and benefits (0.0882). These can be interpreted as: based on the model, staring with a base rating of 0.6037 an increase in each variable by one will increase the overall rating of a review by the value of its coefficient on average. For example, as culture increased by one star, the overall rating would be expected to increase by 0.327 points on average. This model has an R-squared value of 0.707 which means that the model explains almost 71% of the data's variability around the mean. The results are statistically significant at the 99.9% level.

The results of our analysis indicate that the rating with the most impact on was culture (0.327). This was followed by career opportunities (0.2558), work-life balance (0.1422), senior management (0.0882), and benefits (0.0882). This analysis could be expanded to analyze other variables that affect a person's overall rating of a company. This information could help a company maintain high moral levels.

There are several limitations to this analysis. First, the population of this data were self-selected, as anyone can leave a review on Glassdoor. The individuals who feel compelled to write reviews are likely to feel strongly about the company they are reviewing (either positively or negatively). Second, the ratings are completely subjective. There is no way to accurately determine what each rating means to each individual reviewer. Third, we only had data for five independent variables. Other variables may impact the overall rating of a company such as salary or location. Fourth and finally, we were not able to determine if these variables were colinear.

Dep. Variable:	Overall	R-squared:	0.707
Model:	OLS	Adj. R-squared:	0.707
Method:	Least Squares	F-statistic:	3.205e+04
Date:	Thu, 21 Mar 2019	Prob (F-statistic):	0.00
Time:	21:42:51	Log-Likelihood:	-50856.
No. Observations:	53222	AIC:	1.017e+05
Df Residuals:	53217	BIC:	1.018e+05
Df Model:	4		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	0.6037	0.010	62.375	0.000	0.585	0.623
Culture	0.3270	0.003	101.918	0.000	0.321	0.333
Work_Balance	0.1422	0.003	53.448	0.000	0.137	0.147
Oppertunities	0.2558	0.003	86.253	0.000	0.250	0.262
Management	0.0882	0.002	53.527	0.000	0.085	0.091
Benefits	0.0882	0.002	53.527	0.000	0.085	0.091

Part 2: What do employees like and dislike about the Big Five?

After examining multiple regressions between review categories, we wanted to investigate the specific reasons why employees liked or disliked their jobs, specifically at the "Big Five", Amazon, Apple, Facebook, Google, and Microsoft. Fortunately, the dataset from Glassdoor contained review categories for pros and cons which allowed for free responses. We were able to analyze the written reviews to identify common themes for pros and cons at each of the Big Five.

Since each company had reviews numbering in the thousands, we reasoned that any commonly used descriptors across the reviews would would be an indicator of common trends. We utilized a frequency distribution to generate word-clouds as a visual reprentation of common themes and generated a plot of the 30 most commonly used words across pros and cons for each company. What we found that each company had different things (but also similar things) to say about what they liked and dislike about their company.
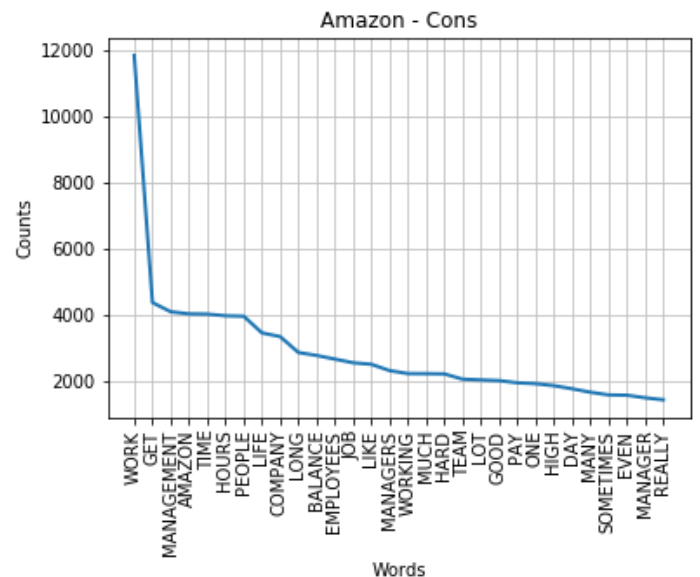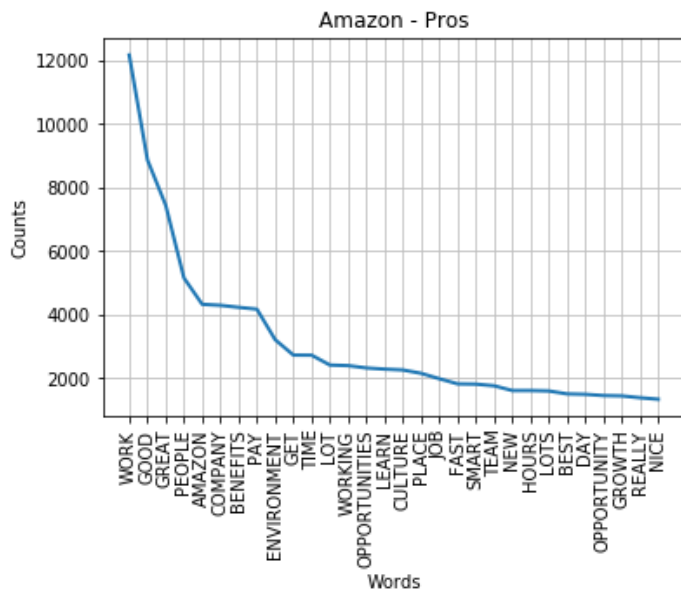
Across the board, the most commonly used word was "work" for each company and for both pros and cons. Naturally, writing about your workplace means that you probably use the word work fairly frequently. Aside from "work", each company had the words "good" or "great" within the top five words mentioned among pros. Again, since many reviews only contained a handful of words (e.g. "great place to work"), this is not a surprise. Additionally, every category of review contained the company name repeatedly.
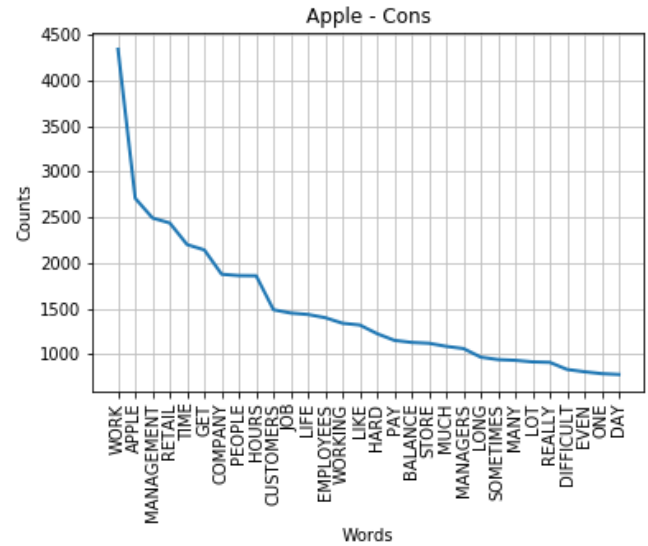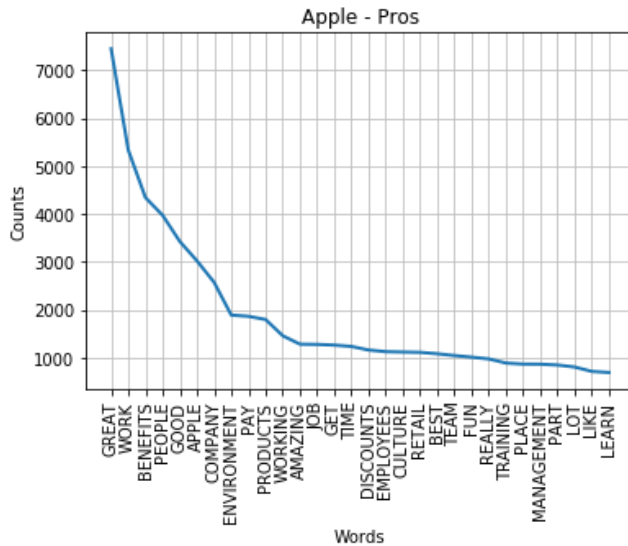
Some of the more interesting insights came from other commonalities between the Big Five. Almost of the pros also contained "benefits", "people", and "culture". Addionally many of the reviews talk about "smart people" which is no surprise considering the reputation of these companies.

In terms of cons, many of the reviews specifically mentioned difficult hours or lack of work-life balance. Since these companies are known for good benefits and compensation, it is not surprising that they can also be demanding. Similarly, several of the companies mentioned management or managers in their reviews, most notably Amazon. While difficulties with management could be related to high expectations and lack of work-life balance, it seems counter-intuitive that both good culture and bad management are both within the same company. Below is a breakdown of each company by frequency distribution.
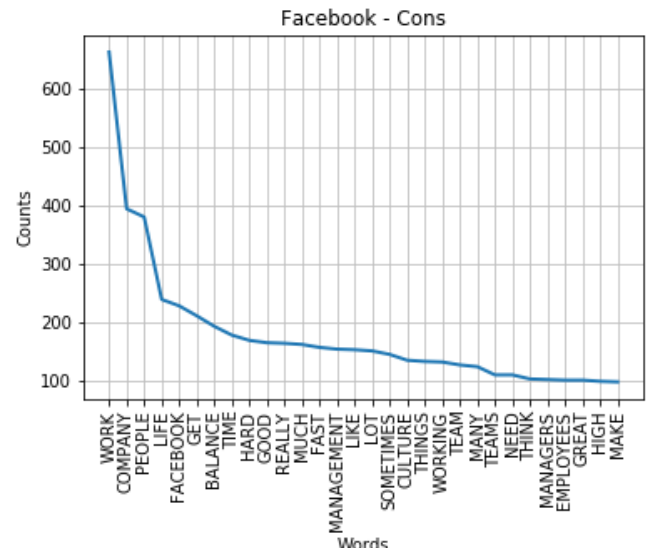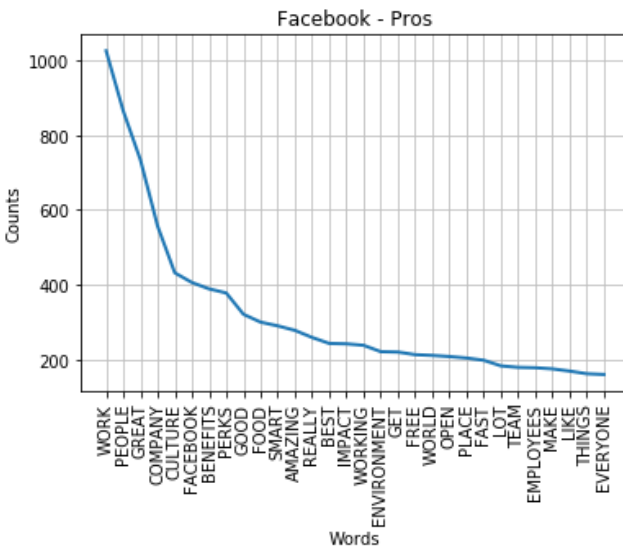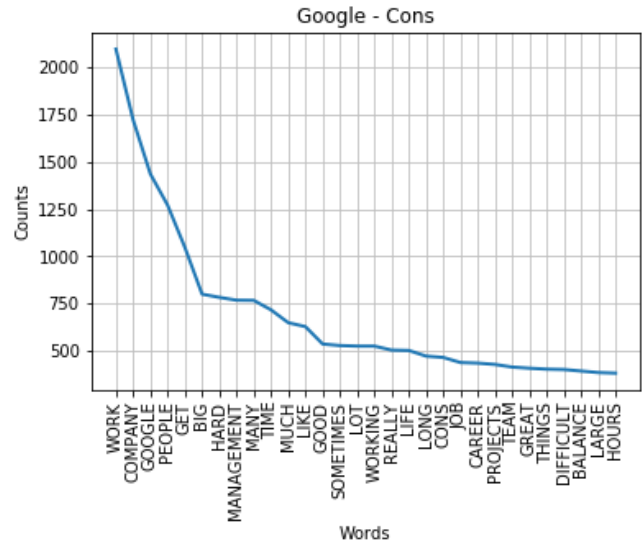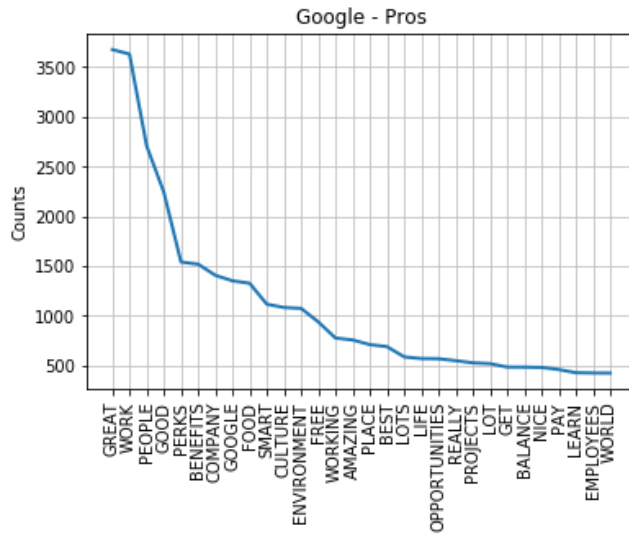
Amazon

Apple

Apple - Pros

Apple - Cons

Facebook

Facebook - Pros

Facebook - Cons

Google

Google - Pros
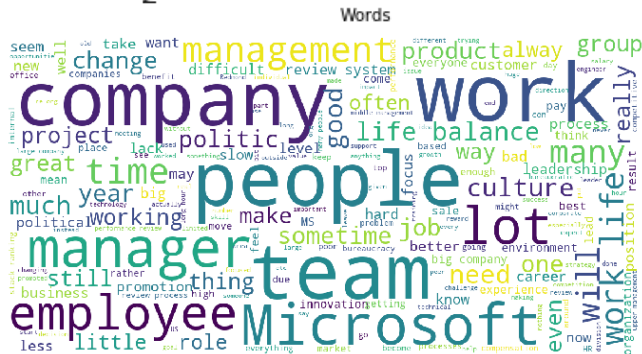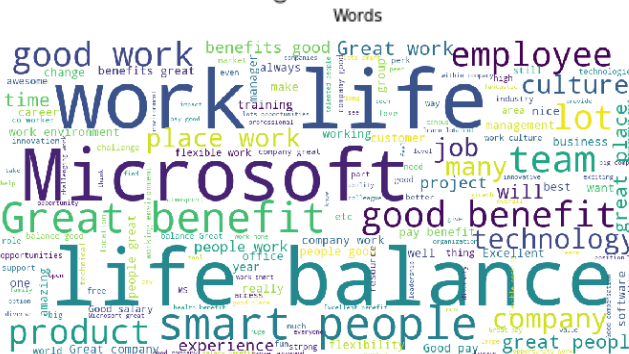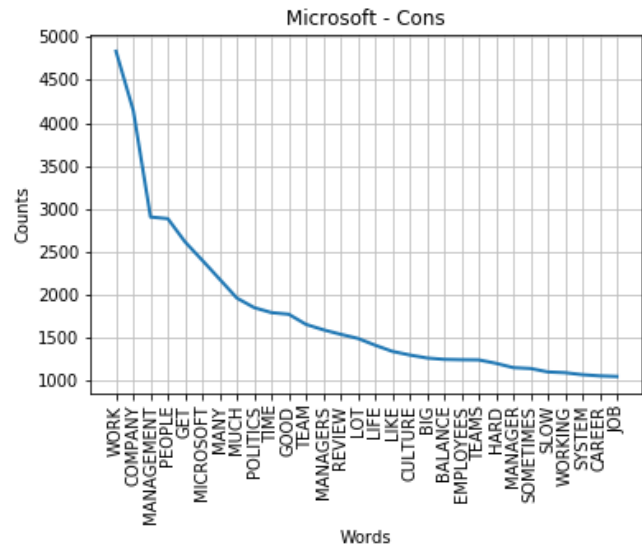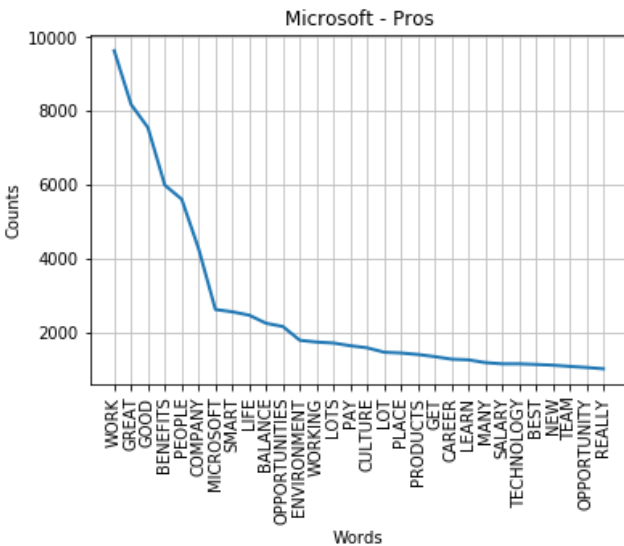

Google - Cons
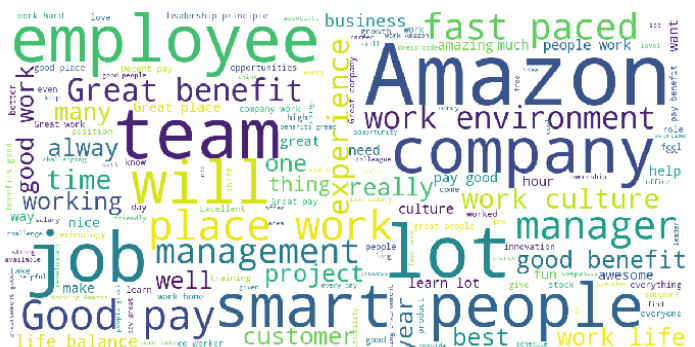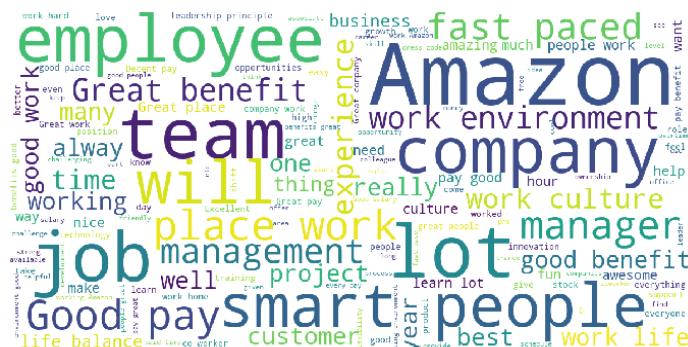
Microsoft

Microsoft - Pros

Microsoft - Cons

# Part 3: Does the unemployment rate affect job satisfaction?

We believe that unemployment rate does have an effect on job satisfaction. We analyzed a dataset of Glassdoor ratings and found that job satisfaction, as measured by overall ratings of the Big Five companies, appears to dip during times when unemployment rates are high (see Fig 1).

We started by obtaining unemployment data from the Bureau of Labor Statistics and compared data for the overall national unemployment rate to the unemployment rate for workers over the age of 25 with bachelor's degrees. Although it is likely that many of the employees at these companies fall into this subset, there are most likely a significant number of employees at these companies that do not fall into the aforementioned category, which prompted us to use the overall unemployment rate in our analysis. The difference between overall unemploymen and unemployment for 4-year graduates over the age of 25 is depicted in Figure 2.

We then analyzed the mean overall rating for each company by year and compared them to each other (Figure 3). Almost every company experienced a dip in their overall ratings around 2009-2010 which corresponds with unemployment rate and the 2008 recession. The one notable exception to this trend is for Facebook, which did not experience a dip in ratings until 2011. All of the companies then experienced a positive surge in ratings towards the end of the recession in 2012 which shows a strong relationship to the steady decline of unemployment since 2010.

There are several limitations that we wanted to acknowledge with this particular analysis of the dataset. The most significant limitatino of the dataset is in the volume of reviews, especially in the early years of Glassdoor. As you can see in Figure 4, 2007 to 2012 have relatively few reviews in comparison to more recent years. Glassdoor has grown immensely in the last five years as transparency within companies has become a more common topic of discussion in recent years.

Additionally, we must consider a shift in internet culture over the last years. It has become increasingly common practice to review just about everything online, from restaurants and other businesses on Yelp, to retail items on sites like Amazon. Online reviews are largely polarizing, with people tending to take the time to write reviews when they have a notably positive or negative experience. Perhaps people who were unhappy with their employers were more likely to seek out a service like Glassdoor before it was well known, contributing to a heavier skew towards negative ratings in the early days of Glassdoor.

Figure 1

Figure 2

Figure 3

Figure 4
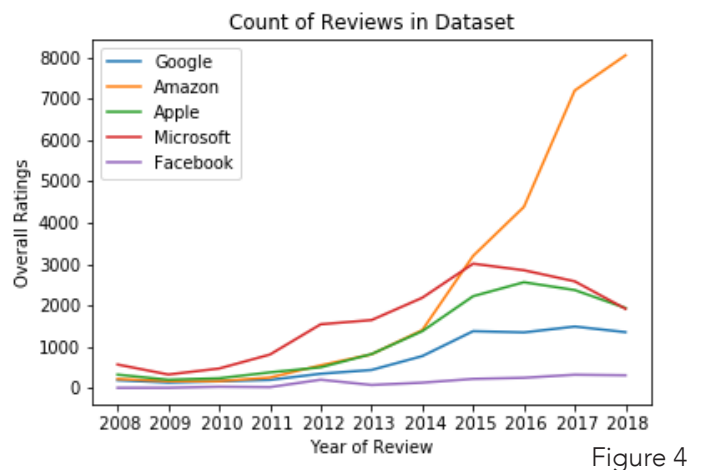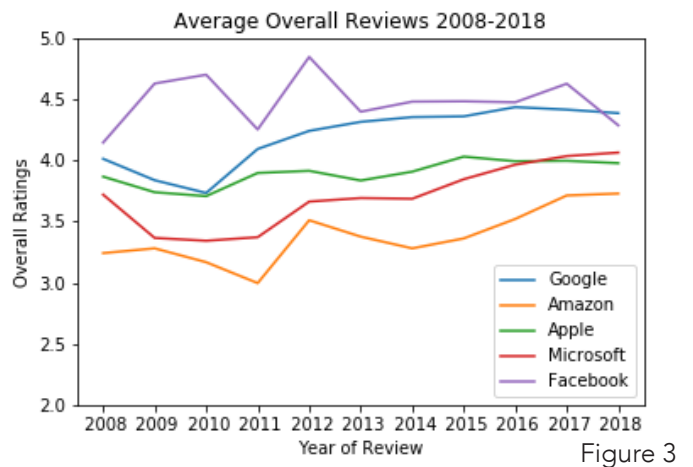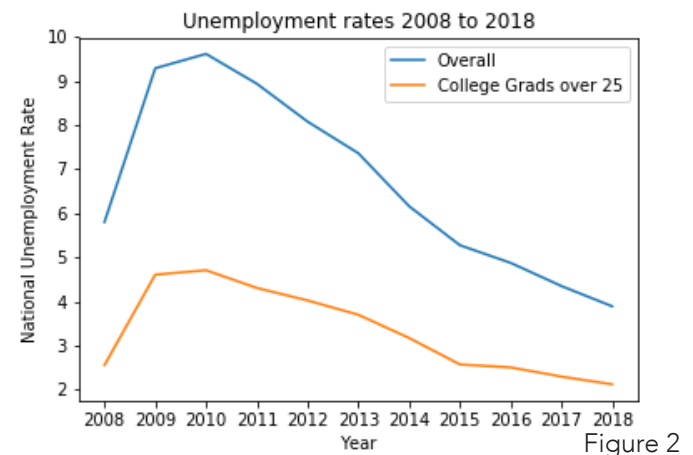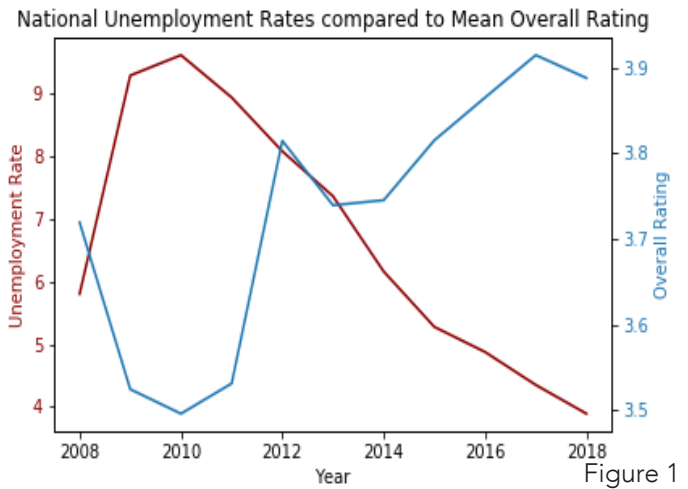
Part 4: Does the success of the company affect job satisfaction?

No, there does not appear to be any correlation between company success for the company and employee satisfaction. We examined the "Big Five" tech companies: Apple, Amazon, Google, Microsoft, and Facebook. Stock values were used as an indicator for success of a company. Historical values for stock data were obtained from www.worldtradingdata.com.

All of the data from Glassdoor was first grouped by company, then employee satisfaction was plotted over time. Employee satisfaction was represented by the overall star rating a reviewer gave the company (on a scale of one to five stars). Each date that a Glassdoor review was posted was then paired with the closing company stock value for that date. This was done to see if the trends in stock value would correlate with patterns of employee satisfaction. The daily results were then grouped by month and averaged so that the average star-rating for a month would match with the average stock value for the same month. Taking an average of the ratings reduced some of the noise that was seen when comparing reviewer ratings on a day to day basis. Finally, the ratings and the stock values were normalized. This was done by taking the stock values and satisfaction ratings and subtracting the minimum value for each data point, followed by dividing by their respective maximum values. The trends are more comparable now, varying from 0 to 100 over the whole range (aka lowest point is zero and highest point is 1).

Information could then be plotted on the same graph to see if there were any patterns in the ratings that responded to the patterns seen in historical stock data.

If employee satisfaction ratings could be modeled on company success, there would have been increases in satisfaction mirroring the increases in stock values. Apple (Figure 1) is a great example of how the ratings do not correlate with the obvious rises and falls in stock value. Instead, we can see that employee ratings are fairly consistent over time, even as the stock value climbs.
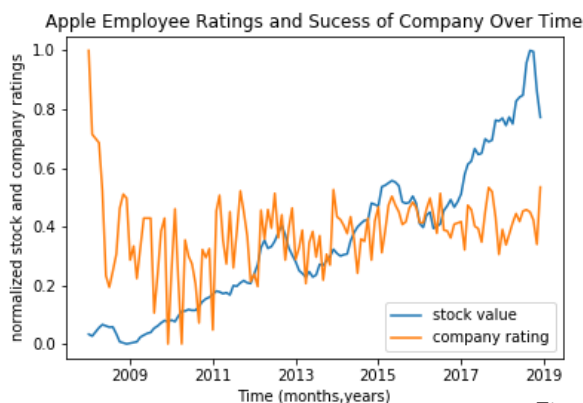

Figure 1

Microsoft(Figure 2) had steadily slow increase in reported employee satisfaction after peak unemployment rate for this period. If stock was a good indicator of happy employees, a bump in approval ratings would have been expected near the increase in stock value around 2018.
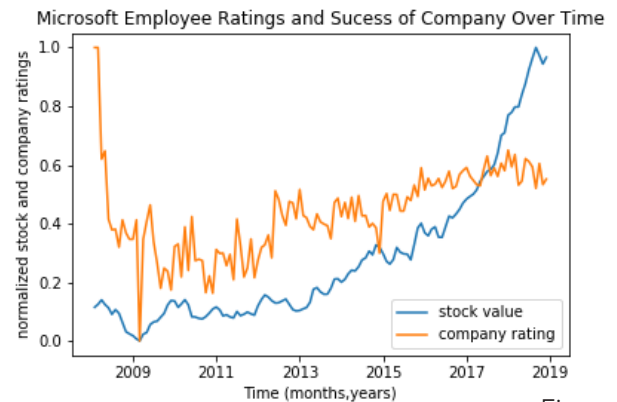

Figure 2

Amazon (Figure 3) and Google (Figure 4) employee ratings dipped around the time that the economy was in recession and became more consistent as years went on. No obvious correlation to the increasing value of stock for this company.
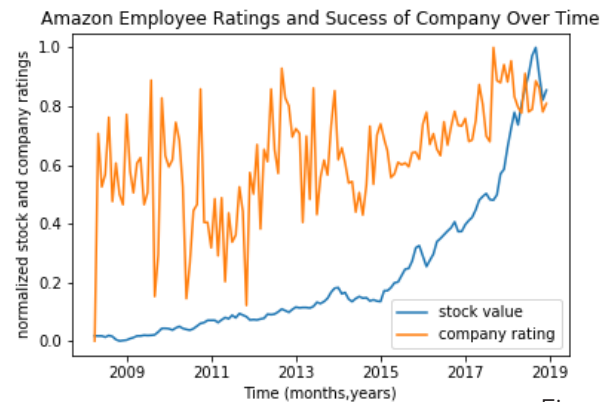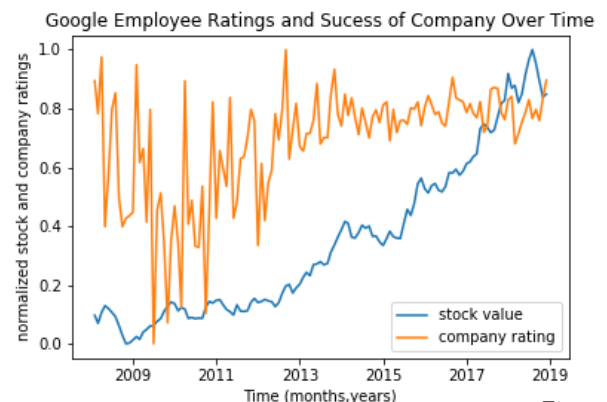

Figure 3

Figure 4

Facebook (Figure 5) had the most variable ratings for employee satisfaction towards the beginning of Glassdoor. This is in part because the sample size for ratings in the first few years was very small. There was also not a complete trend of stock information to follow, since Facebook did not become publicly traded until about 2012. It is also the only

analysis where trends in stock value seems to hint at modeling employee satisfaction in the company. Both start on a similar downwards trajectory starting around 2018.
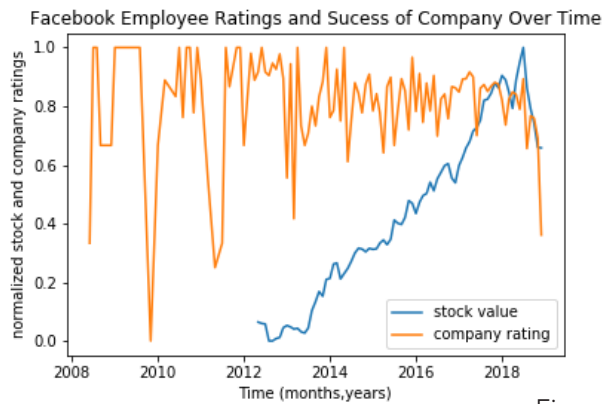

Figure 5

Potential further study

Measuring the success of a company with different methods to see if there is a better indicator of what makes a happy employee. Other ways that company success could be measured could be by customer approval, their financial report, or the rate of attracting new customers. For example, looking at the trends for Facebook, there is a dip in both stock and employee satisfaction that correlates with one of the first big scandal Facebook had, with Cambridge Analytica.