Prof. Lapets - CS 506
Final Project Report
Due: 12/11/18

Analyzing the Relationship Between Employee Sentiment & Company Stock Performance

Kai Munechika (U84907066) & Helen Zhang (U55764049)

Goal

Using the tools and methods from this class -- we want to see if we can find any relationship between how employees internally view a company and how the company's stock performs. We'll base our findings by analyzing Glassdoor employee reviews and stock data from Alpha Vantage and IEX.

Motivation

We found the following sources and reports online that suggest there exists a strong relationship between how positively employees internally view the company and the company's stock performance. We aim to measure the validity of this claim through this project.

1. https://www.eur.nl/sites/corporate/files/Sentiment_analysis_and_the_impact_of_employee_satisfaction_on_firm_earnings.pdf

2. https://www.reuters.com/article/us-investors-workers-sentiment-insight/does-a-happy-employee-make-for-a-healthy-stock-price-idUSKBN0MZ01V20150408

3. https://www.cnbc.com/2015/03/19/happy-employees-happy-shareholders-study-finds.html

Resources

Glassdoor Reviews

We will scrape employee reviews from Glassdoor and use this data to measure how positively or negatively the employees internally view their company.

Alpha Vantage API & IEX API

These resources will provide all the stock related data we'll use to analyze the trends and see if there is any correlation between the stock's performance and the employees' sentiment.

Caveats

Considering a few aspects that we were not able to control, but may have had an impact on our results.

Potential Bias of Glassdoor Reviews

It's entirely possible that the Glassdoor reviews are a flawed resource since people may be biased when writing the reviews. Generally speaking, it seems that employees who leave Glassdoor reviews tend to fall in the extremes of absolutely loving their company or completely hating it.

Industry Bias

Our data mainly focuses on companies in the technology sector -- so while our results may suggest whether a relationship exists, it may not be the same for other companies or industries. It's possible that different industries show a similar or more extreme relationship.

Differences in Employee Rankings

Most of the people who leave reviews on Glassdoor tend to be entry-level or mid-level employees -- it's possible that a higher ranked employee's views on the company may have a more significant impact on the company's stock. This is another area that seems unclear.

Project Outline

For our research we will mainly be focusing on technology companies -- with a few exceptions (like Ulta - a cosmetics brand and Lululemon - a clothing brand). We are also curious to see if the correlation between employee sentiment and company stock performance was different for larger companies vs smaller companies. We'll take this into consideration when analyzing the our results.

Specifically, we will be focusing on the following categorized companies:

Large Companies

- PayPal
- Ulta
- Facebook
- Apple
- Netflix

Small Companies

- Square
- Lululemon
- Workday
- Etsy
- Shopify
- Grubhub

Scraping the Glassdoor Reviews

Since there is no Glassdoor API, we will have to manual scrape for the company reviews using Beautiful Soup. For both the large companies list and small companies list, we will build a JSON file containing all the reviews.

For each review, we will have the following data points:

- Rating (out of 5 stars)
- Pros section (positive feedback on the company)
- Cons section (negative feedback on the company)
- Date of the review

Project Phases

We will break up our project into **3 different phases:**

- **Clustering Glassdoor Employee Reviews**
    - Take the raw text from the Glassdoor employee reviews (ignoring the pros and cons label) and see if we can cluster them into 2 distinct groups and map them back to the 'true label' groups

- **Applying Sentiment Analysis**
    - See if there is a relationship or direct correlation between the language of a Glassdoor employee review and the number of stars it has
- **Finding the Pairwise Correlation**
    - Compare the overall sentiment of Glassdoor reviews with a company's stock trends over time

Phase 1: Clustering Glassdoor Employee Reviews

Goal

To see if people reliably label their 'pros' and 'cons' text into distinct groups.

Approach

Using the KMeans and TF-IDF methods learned from class, we will blindly cluster the review 'pros' and 'cons' text into 2 groups, and then test to see how close our results were to their 'true labels'. If we find that our KMeans algorithm was able to consistently divide the unlabeled 'pros' and 'cons' text, then we can conclude that the employees leaving reviews tend to be consistent with what language they put in their 'pros' and 'cons' comments.

Steps:

1. Go through all the company reviews (in both the large & small list) and for each of those reviews:
    a. Create an empty array for all the unlabeled, raw 'pros' and 'cons' text, and another array that will hold the 'true labels'
    b. Then apply tf-idf to format the strings into vectors
    c. Scale the data down to 200 components using SVD
    d. Apply the KMeans algorithm to make predictions on which label each unlabeled, review text will get (out of 2 possible choices).
    e. Compare the predicted labels with the true labels and see whether our approach was able to consistently group the text into 2 clusters

Results

We found that around 71% of the time, pros and cons were consistently grouped into 2 distinct groups based on the language used. To determine the success of our assigned labels and the 'true' labels (from the original Glassdoor reviews), we would take the most popular matching pair to be one successful label, and then consider the opposite pair to be the other successful label.

For example, with Netflix we have the following result:

currently on company: Netflix
<**class** **'scipy.sparse.csr.csr_matrix'**> (1618, 1468)
(1618, 200)
{(0, 1), (1, 0), (0, 0), (1, 1)}
(0, 0)   793
(0, 1)   434
(1, 1)   375
(1, 0)   16

We take (0,0) to be one pair, and the other successful label would be its opposite (1,1). To calculate the percentage, we would add the total matches for (0,0) and (1,1), and then divide it by the total number of reviews. In this case, 793+375 / 1618

Below we've provided a table summarizing the percentage of reviews that were successfully grouped into a 'pros' or 'cons' label, and the following table provides a more detailed output from our program. It also seems that company size did not have much influence on the clustering.

Smaller Companies	**Larger Companies**
Square: ~79%	Netflix: ~72%
Lululemon: ~70%	Paypal: ~62%
Workday: ~69%	Ulta: ~76%
Etsy: ~80%	Facebook: ~67%
Shopify: ~62%	Apple: ~65%
Grubhub: ~78%	

Avg = 71%

currently on company: Square <**class** **'scipy.sparse.csr.csr_matrix'**> (784, 846) (784, 200) {(0, 1), (1, 0), (0, 0), (1, 1)} (1, 0) 372 (0, 1) 249 (1, 1) 143 (0, 0) 20 .79 currently on company: Lululemon <**class** **'scipy.sparse.csr.csr_matrix'**> (2762, 1300) (2762, 200) {(0, 1), (1, 0), (0, 0), (1, 1)} (1, 0) 1359	currently on company: Netflix <**class** **'scipy.sparse.csr.csr_matrix'**> (1618, 1468) (1618, 200) {(0, 1), (1, 0), (0, 0), (1, 1)} (0, 0) 793 (0, 1) 434 (1, 1) 375 (1, 0) 16 .72 currently on company: PayPal <**class** **'scipy.sparse.csr.csr_matrix'**> (5266, 2324) (5266, 200) {(0, 1), (1, 0), (0, 0), (1, 1)} (0, 0) 2524

```
(1, 1)    818
(0, 1)    563
(0, 0)     22
.70

currently on company:  Workday
<class 'scipy.sparse.csr.csr_matrix'> (1852, 1526)
(1852, 200)
{(0, 1), (1, 0), (0, 0), (1, 1)}
(1, 0)   891
(1, 1)   540
(0, 1)   386
(0, 0)    35
.69

currently on company:  Etsy
<class 'scipy.sparse.csr.csr_matrix'> (264, 257)
(264, 200)
{(0, 1), (1, 0), (0, 0), (1, 1)}
(1, 0)   127
(0, 1)    84
(1, 1)    48
(0, 0)     5
.80

currently on company:  Shopify
<class 'scipy.sparse.csr.csr_matrix'> (582, 582)
(582, 200)
{(0, 1), (1, 0), (0, 0), (1, 1)}
(1, 0)   280
(1, 1)   211
(0, 1)    80
(0, 0)    11
.62

currently on company:  Grubhub
<class 'scipy.sparse.csr.csr_matrix'> (820, 710)
(820, 200)
{(0, 1), (1, 0), (0, 0), (1, 1)}
(1, 0)   396
(0, 1)   247
(1, 1)   163
(0, 0)    14
.78
```

```
(0, 1)   1882
(1, 1)    751
(1, 0)    109
.62

currently on company:  Ulta
<class 'scipy.sparse.csr.csr_matrix'> (7854, 2581)
(7854, 200)
{(0, 1), (1, 0), (0, 0), (1, 1)}
(0, 0)   3893
(1, 1)   2119
(0, 1)   1808
(1, 0)     34
.76

currently on company:  Facebook
<class 'scipy.sparse.csr.csr_matrix'> (3156, 1788)
(3156, 200)
{(0, 1), (1, 0), (0, 0), (1, 1)}
(1, 0)   1542
(1, 1)   1002
(0, 1)    576
(0, 0)     36
.67

currently on company:  Apple
<class 'scipy.sparse.csr.csr_matrix'> (25520, 5430)
(25520, 200)
{(0, 1), (1, 0), (0, 0), (1, 1)}
(1, 0)   12612
(1, 1)    8717
(0, 1)    4043
(0, 0)     148
.65
```

Phase 2: Applying Sentiment Analysis

Goal

To see if there is any consistency with the sentiment in employee reviews and the star rating they assign to the company.

Approach

Using Textblob, a 3rd party library for sentiment analysis, we used its ‘polarity score’ function to determine how positive or negative the text from a review was.
The polarity score has a scale from -1 to 1, where a value closer to -1 meant the review was more negative, and vice versa for a value closer to 1. A polarity score of 0 meant that the text was relatively neutral.

Steps:

1. Create a dictionary to hold the review scores based on the star rating of that review
2. Go through up to 500 reviews for all the companies (in both the large & small list) and for each of those reviews:
    a. Calculate the overall polarity score by adding the polarity score of the review’s ‘pros’ text with the polarity score of the review’s ‘cons’ text
    b. Update the dictionary based on what rating the employee gave
3. Calculate the average polarity scores split based on the star ratings

Results

We found that there was a direct relationship between the polarity scores of a review and the star rating that the employee gave. The following is a dictionary of results, where the key is the star rating (1 being the lowest, 5 being the highest).

{1: {'review_count': 1765, 'polarity_score': 468.8436637284275, 'average_score': 0.26563380381214025},
2: {'review_count': 2273, 'polarity_score': 814.727823507667, 'average_score': 0.3584372298757884},
3: {'review_count': 4651, 'polarity_score': 1947.9523991615251, 'average_score': 0.4188244246745915},
4: {'review_count': 7604, 'polarity_score': 3504.1687242375597, 'average_score':0.46083228882661226},
5: {'review_count': 8946, 'polarity_score': 4654.197614079766, 'average_score': 0.5202545958059206}}

Phase 3: Finding the Pairwise Correlation

Goal

Find out if there is any correlation between employee sentiment and stock prices based on data from the last 4 years.

Approach

Using data from IEX and Alpha Vantage, we will calculate the monthly average stock price for all of our companies, and compare it with the average monthly sentiment found in the Glassdoor reviews to see if any relationship exists.

Steps:

1. Create a JSON file that stores the monthly historical stock price average dictionary for each company, such that the key is a string date and the value is the monthly average stock price. (ex: '2018-09':124.06)
2. Create a JSON file containing the average monthly sentiment based on the Glassdoor reviews. Instead of the values being the average monthly stock price, they will be the average monthly review sentiment.
3. For each company:
    a. Make a data frame for both the monthly stock prices and monthly sentiment
    b. Concat the data frames together and apply the pairwise correlation function (build into Pandas)

Results

Below are the results we found. The table categorizes the companies based on their correlation significance.

Negative Correlation	Positive Correlation	Weak Correlation
Apple -0.09990978186930763 Etsy -0.1433104768673486 Lululemon -0.05607332389134671 Square -0.01048908912523195 Grubhub -0.2045696903435296	Facebook 0.20941687636012443 Netflix 0.04660470405883612 Shopify 0.5499292645088557 Workday 0.1196661421525131	Ulta -0.008774692916692296 PayPal 0.001125894949762557

Conclusion

Overall, it seems that from our results, **there is not enough evidence to support the original claim** of a strong correlation between employee sentiment and stock performance. Out of the 11 companies, only 4 of them had a positive correlation -- while the other 7 had a negative or weak correlation between employee sentiment and stock performance.

Future Considerations

Given our previously mentioned caveats, in a future iteration it may be helpful to somehow normalize the data. Access to another resource for unbiased employee reviews or employee sentiment may also allow more meaningful results. Another way to add a dimension to this project would be to consider the role of the employees leaving each review and putting a higher weight on those with more executive positions. Testing if this correlation is any different with other industries could also add more insight to this topic.