Data Breach Corporate Apology Comparisons

DS3500 Final Project

Luke Abbatessa, Yitian Liang, Naman Razdan, Jasmine Wong, Yu Xiao, & Yuting Zheng

Northeastern University, Boston, MA, USA

Abstract

A data breach is an incident where sensitive data is accessed, stolen or used by an individual unauthorized to do so. The increasing prevalence of these breaches coupled with an increased consumer demand for data transparency makes this topic particularly relevant. Our group selected 11 massive data breaches that happened in the past decade to investigate the aftermath of data breaches. We created our own databases of the incident timeline, apologies, and stock price. We resorted to sentiment analysis to identify the patterns in apologies, followed by an analysis of stock price fluctuation over the time when apologies were published. We hypothesized that cumulative sentiment scores would increase throughout a text, and that stock price percentage drops would be mitigated by apologies with more positive sentiment scores than apologies with more negative sentiment scores. We built an interactive dashboard to present our findings. The first tab groups companies based on sensitivity of stolen data, showing cumulative apology sentiment scores and their relationships with each companies' stock fluctuations; the second tab allows users to compare cumulative sentiment scores, Heaps' Law plots, and stock fluctuations of two user chosen companies. Through our analysis, we found that 8 of the 11 company statements we analyzed ended with a higher sentiment score than they began with, and that there was no clear correlation between a company's average sentiment score and stock price percentage decrease. In addition to testing our hypotheses, one of our main goals was to build an accessible and interactive learning tool to raise awareness about the threat of data breaches through our dashboard.

Introduction

In the past decade, the number of data breaches has more than doubled. These breaches range in severity, compromising data as unsusceptible as name, email, and gender or as sensitive as complete credit card information, social security numbers, driver's license numbers, and even passport information. What they have in common is that real people are always affected. To respond to scandals and apologize to customers, companies carefully craft apologies with the hope of earning goodwill and regaining trust.

It would be interesting to observe trends in apologies' sentiments to gauge the strategy and authenticity of the companies' responses. To do this we performed VADER sentiment analysis

on each apology, plotting the cumulative sentiment score of the apology against its progression. Coupled with a Heaps' Law plot, this sentiment analysis allows us to understand how the tone, mood, and verbiage change throughout the apology. After understanding the nature of each apology, our team's primary goal is to understand the relationship between corporate apologies following a data breach at a certain firm and the stock price of that company before and after the breach and apology. This research could provide insight into the effectiveness of particular styles of apology statements in pacifying public disapproval.

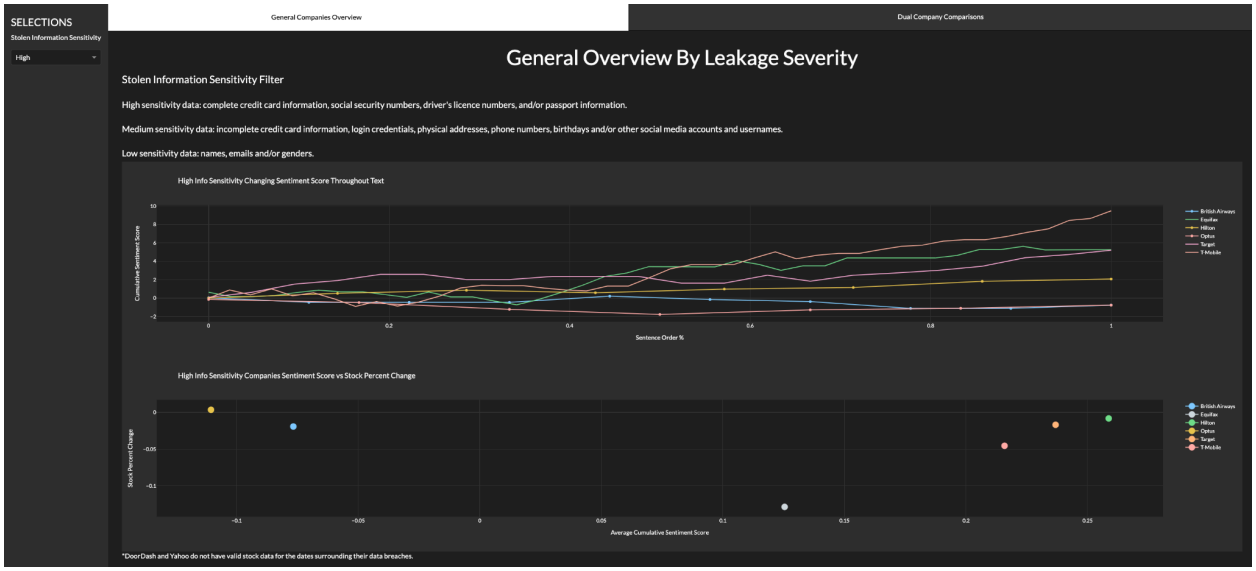
There is existing research on the relationships between corporate apologies and stock market volatility, but this research does not analyze the sentiments of apology statements, nor do they focus on data breaches. In an analysis of corporate apologies after non-financial crises, Racine finds that companies must match their formal response strategies to the severity of the crisis [1]. Apologizing when the firm is not directly responsible, or not effectively apologizing when the firm is directly responsible, directly reduces shareholder wealth. Consequently, apologizing when responsibility should be taken and refraining from apologizing when responsibility should not be taken mitigates shareholder losses. This suggests that an authentic and an effective understanding of public perception is invaluable for companies responding to crises. With this project, we hope to further explore the importance of authenticity in corporate apologies for data breaches.

Methods

Since there are no databases containing a consolidation of data breach timelines and their corresponding corporate apologies, our team researched and created our own mini-database of historic data breaches. In looking for data breaches to add to our database, we searched with three criteria in mind: 1) sufficient media coverage to be able to cross-verify timeline facts, 2) publicly available access to apology statements, 3) large companies that may have relevant stock data. After researching online for relevant data breaches, our team narrowed down on 11 companies (British Airways [1], DoorDash [17], Equifax [16], Hilton [11], LinkedIn [6], Optus [5], Pearson [8], Target [2, 13], T-Mobile [12, 14], Twitter [15], and Yahoo [3, 9, 18]) and scraped their apologies into .txt files. Once we generated the .txt files, we generated .json files associated with each .txt file that included the name of the .json file, the name of the associated company, and the actual text that comprised the company's apology as key-value pairs. Additionally, we researched the timeline of each individual breach and extracted the following dates: day before the public knows about the data breach, day the public knows about the data breach, day that the company responded, and day after company response. Using these dates to select the stock prices for each of the 11 companies using Yahoo Finance [19], we compiled a csv file that contained the desired closing prices for each of the 11 companies.

Analysis

The main outcome of our project is a dashboard allowing users to explore the sentiment scores of 11 corporate apologies and the corresponding stock fluctuations caused by the initial breach and the apology release. Our dashboard is divided into two tabs: the first tab groups companies based on severity of the breach, providing information on cumulative apology sentiment scores and their relationships with each companies' stock fluctuations; the second tab allows users to select two companies and compare their cumulative sentiment scores, Heaps' Law plots, and stock fluctuations.

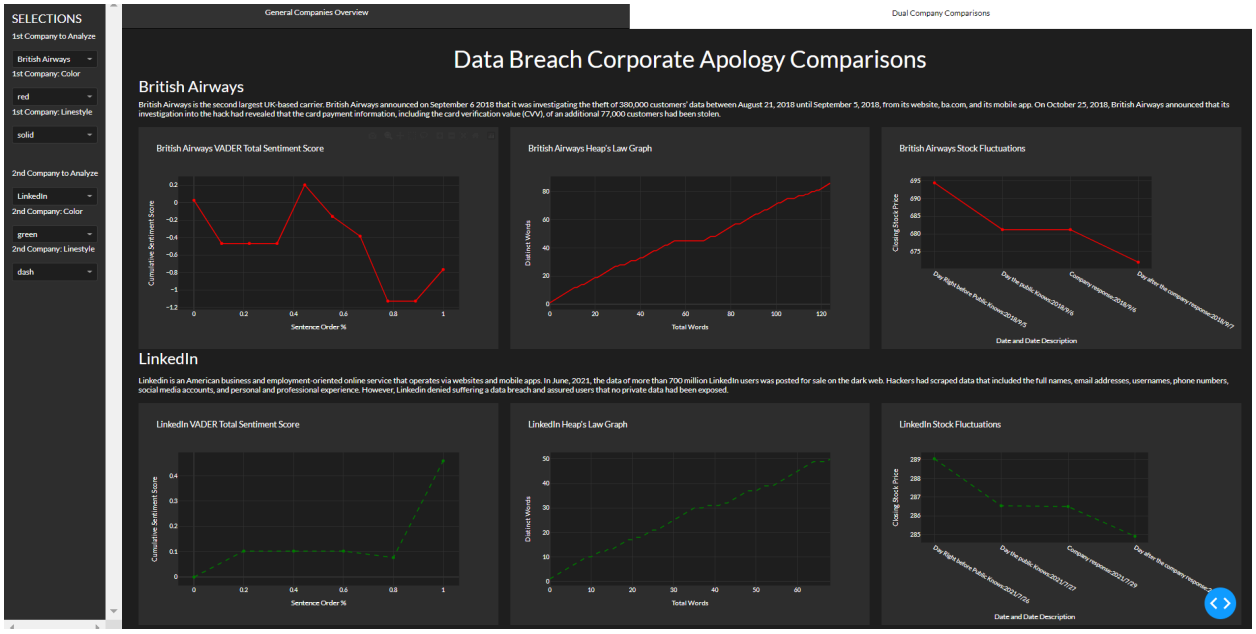


Tab 1: Cumulative Sentiment Scores and Sentiment Score vs Stock Percent Change plots (High Sensitivity Data Breaches)

The three groupings for the first tab of our dashboard are as follows: Highly Sensitive Information, including full credit card details, social security numbers, license numbers, and/or passport numbers; Moderately Sensitive Information, such as usernames and passwords for other social media accounts, incomplete credit card information, physical addresses, phone numbers, and birthdays; and Data of Limited Sensitivity, such as gender and/or name. In our chart section above, we have presented all 11 companies' VADER sentiment scores in one chart for better comparison. At the bottom, we present the VADER sentiment score for each of the 11 companies as a scatter plot against the percentage of the stock.

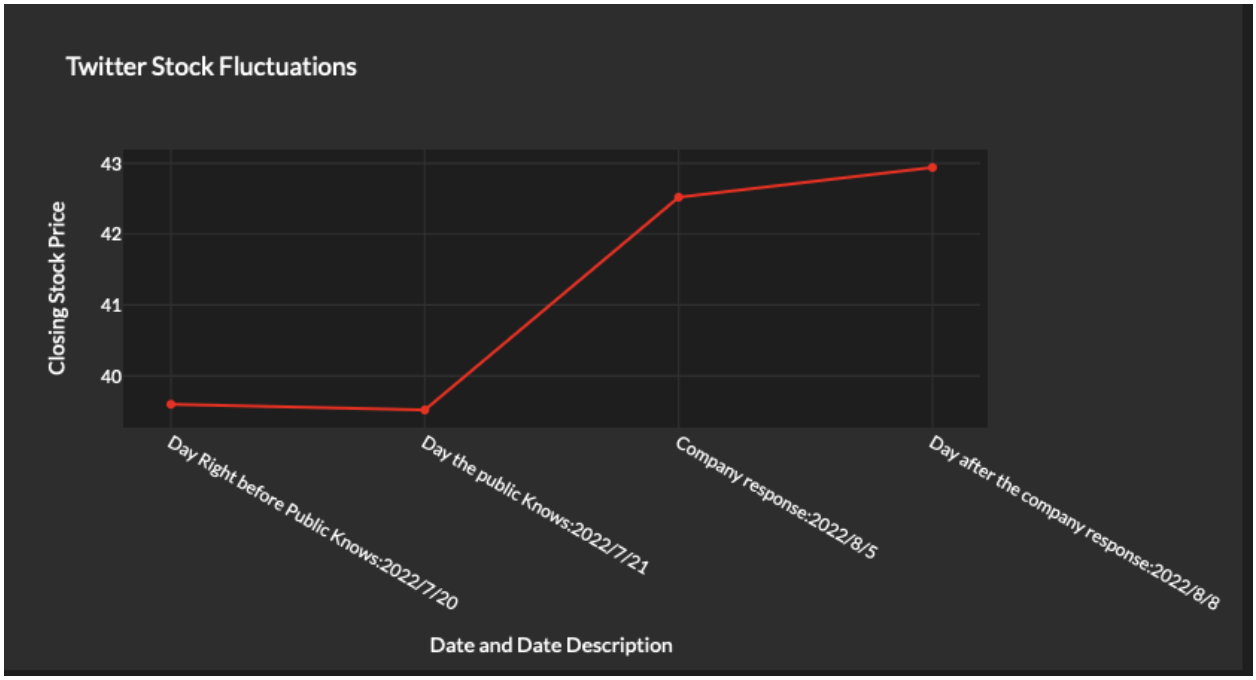
By choosing different information sensitivity levels and their scatter plot distribution, we can observe whether they have a positive or negative impact on the stock price, and the extent of their impact. For example, if we choose highly sensitive data, the scatter plot of the lower part is almost all distributed below 0, which means that leaking highly sensitive data will definitely have a negative impact on the company's stock price.

The second tab of our dashboard features a selection panel on the left, allowing the user to select two companies to compare. In order to make the dashboard more accessible to colorblind audiences, we also allow the user to customize the color and linestyle for each plot. Each company's section contains a Cumulative Sentiment Score vs Sentence Percentile chart, a Heaps' Law chart, and a Stock Price chart. Lining each company's chart with the other company's respective chart allows users to compare the two graphs with minimal visual effort and increased accuracy (when compared to alternate dashboard layouts).



Tab 2: Cumulative Sentiment Scores, Heaps' Law Graphs, and Stock Chart (British Airways and LinkedIn)

Twitter's data breach stands out from the rest due to investors' unique response to the crisis. Twitter's stock price only dropped $0.08 after knowledge of the breach was made public, and after the company response, the stock went up $3.00, and the stock continued to rise the day after. Interestingly, Twitter was the only company whose stock trended higher after the breach than before the breach. Potential reasons can be gleaned from our description of the breach: "In January, 2022, Twitter had previously confirmed the existence of the vulnerability" and the exposed data only consisted of "Users' phone numbers and email addresses." Because investors already had knowledge of the vulnerability, perhaps their perception of the stock was already adjusted. Second, the breach was only of medium severity, and although many serious breaches have occurred in recent years, perhaps investors were unperturbed by the incident. Finally, we must consider confounding factors. Twitter was recently acquired by Elon Musk after a tumultuous year of negotiations, and its stock price has fluctuated highly and often as a result. Because of this, it becomes difficult to draw conclusions from stock data a week apart even in the context of a data breach accessing the contact details of 5.4 million accounts.



Twitter Closing Stock Price Fluctuations Graph

Conclusions

Our project focused on the aftermath of data breaches. To display a dashboard that allows for varied analysis to the users, we combined sentiment analysis of the firm's apology after the data breach with changes in the stock price of the company during the precise time period of the data breach. The project was completed in three phases: brainstorming phase, data gathering phase, and development phase.

We investigated 11 of the largest data breaches in a decade (British Airways, DoorDash, Equifax, Hilton, LinkedIn, Optus, Pearson, Target, T-Mobile, Twitter, and Yahoo) and gathered the original apology statements published by each company after they suffered a data breach. Using the gathered apology statements, a VADER sentiment analysis was conducted, combined with stock data, to look at the relationship between VADER sentiment scores and stock prices. We plotted VADER Sentiment Total Score, Heaps' Law graph, and Stock Fluctuation in the Dual Company Comparisons tab and found that all companies, except Twitter, showed a downward trend in stock before and after the company issued a statement. From this, it can be concluded that after the data breach, the event had a negative impact on the company's stock, regardless of the high VADER sentiment score of the apology letter. After reaching our conclusion, we conducted a separate analysis of the special case, Twitter, and we found that Twitter was the only company whose stock trended going up after the breach. The potential reasons for this can be gleaned from our description of the vulnerability. We found that the exposed data included only "users' phone numbers and email addresses."After finalizing the company comparisons, we realized that there was one factor we had overlooked which was the sensitivity of the data being

compromised. So we created another tab on the dashboard, General Overview By Leakage Severity, to further explore whether different severity level data breach caused a positive or negative impact on the stock price.

In general, we have reached the investigation of finding the VADER sentiment score of companies' statements after data breaches, finding the change of stock price before and after data breach, and finding the ratio of breach and stock price change for different sensitivity levels of data.

In future works, our group hopes to build a more comprehensive dataset to enrich our dashboard. We hope to analyze more cases of data breaches to draw more comprehensive and accurate conclusions. We also want to find a better solution to improve the readability of our descriptions and add hoverable descriptions on graphs to help users better understand our findings. We hope that the results of this project will raise the public awareness about privacy and data security which will finally spur on companies to implement more security measures.

References

1. BBC. (2018, October 25). *Ba investigation into website Hack reveals more victims*. BBC News. Retrieved December 7, 2022, from https://www.bbc.com/news/technology-45953237
2. *A Bullseye View. behind the scenes at Target*. Target Corporate. (n.d.). Retrieved December 7, 2022, from https://corporate.target.com/press/releases/2013/12/a-message-from-ceo-gregg-steinhafel-about-targets
3. Condliffe, J. (2020, April 2). *A history of yahoo hacks*. MIT Technology Review. Retrieved December 7, 2022, from https://www.technologyreview.com/2016/12/15/106901/a-history-of-yahoo-hacks/
4. Cyberknow. (2022, September 28). *Optus Data Breach timeline*. Medium. Retrieved December 7, 2022, from https://cyberknow.medium.com/optus-data-breach-timeline-c02d8c5298c4
5. Guardian News and Media. (2022, October 3). *Optus reveals at least 2.1 million id numbers exposed in massive data breach*. The Guardian. Retrieved December 7, 2022, from https://www.theguardian.com/business/2022/oct/03/optus-commissions-independent-review-of-data-breach
6. Lovejoy, B. (2021, June 29). *LinkedIn breach reportedly exposes data of 92% of users, including inferred salaries [U]*. 9to5Mac. Retrieved December 7, 2022, from https://9to5mac.com/2021/06/29/linkedin-breach/
7. *Managing risk together*. ORX. (2022, December 6). Retrieved December 7, 2022, from https://managingrisktogether.orx.org/

8. *Pearson Customer Notification*. Pearson plc. (n.d.). Retrieved December 7, 2022, from https://plc.pearson.com/en-US/news/pearson-customer-notification
9. Perlroth, N. (2016, September 22). *Yahoo says hackers stole data on 500 million users in 2014*. The New York Times. Retrieved December 7, 2022, from https://www.nytimes.com/2016/09/23/technology/yahoo-hackers.html
10. Racine, M., Wilson, C., & Wynes, M. (2018). The value of apology: How do corporate apologies moderate the stock market reaction to non-financial corporate crises? *Journal of Business Ethics*, *163*(3), 485–505. https://doi.org/10.1007/s10551-018-4037-5
11. Schwartz, M. J., & Ross, R. (n.d.). *Hilton Hotels: We were breached*. Bank Information Security. Retrieved December 7, 2022, from https://www.bankinfosecurity.com/hilton-hotels-we-were-breached-a-8703
12. Sievert, M. (2021, August 27). *The cyberattack against t-mobile and our customers: What happened, and what we are doing about it. - t-mobile newsroom*. T. Retrieved December 7, 2022, from https://www.t-mobile.com/news/network/cyberattack-against-tmobile-and-our-customers
13. Star Tribune. (2014, March 27). *Target Data Breach timeline*. Star Tribune. Retrieved December 7, 2022, from https://www.startribune.com/target-data-breach-timeline/252562691/
14. *T-mobile investigating claims of Massive Customer Data breach*. VICE. (2021, August 15). Retrieved December 7, 2022, from https://www.vice.com/en/article/akg8wg/tmobile-investigating-customer-data-breach-100-million
15. Vincent. (2022, August 23). *Twitter data breaches: Full timeline through 2022*. Firewall Times. Retrieved December 7, 2022, from https://firewalltimes.com/twitter-data-breach-timeline/
16. Weise, E. (2017, October 3). *A timeline of events surrounding the equifax data breach*. USA Today. Retrieved December 7, 2022, from https://www.usatoday.com/story/tech/2017/09/26/timeline-events-surrounding-equifax-data-breach/703691001/
17. Whittaker, Z. (2019, September 26). *Doordash confirms data breach affected 4.9 million customers, workers and merchants*. TechCrunch. Retrieved December 7, 2022, from https://techcrunch.com/2019/09/26/doordash-data-breach/
18. Yahoo! (n.d.). *Yahoo security notice september 22, 2016 | yahoo help - SLN28092*. Yahoo! Retrieved December 7, 2022, from https://help.yahoo.com/kb/sln28092.html
19. *Yahoo is part of the Yahoo family of brands*. (n.d.) Retrieved December 7, 2022, from https://finance.yahoo.com/?fr=sycsrp_catchall