



Flying Into Turbulence: Sentiment Analysis of Southwest Airlines' Policy Changes

Amanda White, Isabella Solomon, Aaron Broering, Thomas Skaggs

May 2025

Abstract

Consumer sentiment can be defined as “How customers feel about a brand, their products and services, and their customer relations.” While it can’t be quantified into a dollar-sign figure, it can be used to explain a drop in revenue or customers for a business. Another measuring stick is the stock price - increases and decreases reflect how people (large funds most importantly) feel a business is performing now and how they feel the company will perform in the future. Specifically, we are focusing on Southwest Airlines - a domestic, low-budget airline based in the US. Recently, they have announced policy changes that removed cornerstones of their airline, chiefly a free checked bag and unassigned seating. Our solution is 2 models - one that tracks public sentiment via Reddit activity and another that tracks investor sentiment via the stock price. Taking data from different points in time, we can plot these metrics using a line graph to show how Southwest is trending. Using these 2 models for Southwest, we found that the general public reacted very negatively to these changes - which makes sense, as the airline had taken away 2 of the things that made flying with them over their competitors more attractive. However, the stock price shot up, meaning that investors still feel positively about Southwest and their future.

Introduction

Our project revolves around Southwest Airlines and their potential decline in positive sentiment from customers. Since Southwest was founded in 1967, they have built a loyal customer base through their customer-centric policies, created with the goal of being cost efficient and more affordable for those who need it. These policies included checking a bag for free, unassigned seating, and allowing customers to select their desired seat after boarding. On March 11th of 2025, Southwest Airlines announced changes to these policies amongst others that shocked many. Starting May of 2025, checked bags are no longer free, assigned seating will be implemented with nine boarding groups instead of three, and new premium seat options will be offered. In addition to their more well-known policies, there will also be implementing expiration dates for any fare credits of only a year or less, depending on the value of the original purchase.

Following the announcement of these changes, the public has been contemplating whether or not they have lost their competitive advantage against other airlines and ultimately, if they will be losing more customers. Individuals have been quick to jump on public social media platforms such as X, Reddit, and Facebook to express their opinions. With our project, we will be gauging the sentiment of the public to see if this has led to any indications of change in profitability. To properly transform this business problem into an analytics problem, we decided to apply it as a classification problem where we classify the average sentiment as either positive or negative. In order to do so, we have retrieved comments from a post regarding the policy changes on Reddit and performed a sentiment analysis. We have also collected the changes in stock price before and after the announcement. These varying sources allow us to understand the perspective of not just customers but also any financial investors. As it can be seen later in the report, we ultimately found there was an overall negative sentiment meanwhile stock prices jumped greatly; this indicates that while customers were unhappy with these changes, investors approved and saw it as an opportunity to increase profit.

Considering the business context, this problem matters greatly to Southwest Airlines and their stakeholders because a strong shift in sentiment now could indicate a much larger reputational decline further down in time. Negative sentiment may result in a loss of customers and an inability to gain new ones if there is nothing that differentiates themselves from other airlines. In order to best understand their customers and grasp how to act following the public's reaction, it is important to keep a pulse on their opinions by utilizing sentiment analysis. If Southwest Airlines were to continue leveraging the same tasks that we implemented within our project, the company would be empowered to grow an understanding of their customer base in order to balance their interests with maximizing profitability.

Data Collection and Preparation

We used two main sources to collect our data. One of them was a Reddit thread found under the r/SouthwestAirlines subreddit. The thread was titled "Southwest will now charge you for checking bags," and it sparked a lot of reactions from users. To collect this data, we first had to set up a Reddit developer account, which gave us access to a client ID and client secret. These credentials were necessary for

authenticating our connection to Reddit's API, which we used to pull the content from the thread.

Initially, we attempted to use the praw library, which is the standard tool for accessing Reddit data in Python. However, because our project was being developed in a Google Colab environment, praw generated runtime errors. After troubleshooting, we discovered that asyncpraw, an asynchronous version of the library, was required for compatibility with Colab. Using asyncpraw allowed us to bypass these runtime issues and successfully collect the data without interruption.

When scraping the data, we were careful to only extract the original posts in the thread and their associated posting dates. We intentionally did not include the reply comments underneath each post. This decision was made to maintain the integrity of our sentiment analysis. Including replies could introduce misleading signals, like if someone responded positively to a negative post, the overall tone might be misrepresented. For instance, if someone posted, "This new baggage policy is terrible," and a reply says, "Totally agree! You nailed it." If we included that reply, it could artificially boost the sentiment score toward positive when the original message was clearly negative. To avoid this kind of distortion, we stuck strictly to the top level posts.

After gathering the data, we went through a detailed cleaning process. We removed excess white space, line breaks, and any posts that had been deleted or were no longer available. We also ensured proper text encoding so the data could be interpreted correctly by analysis tools. Additionally, we reformatted the post dates to follow a standardized date-time format that could be easily sorted and analyzed later on.

Once the data was cleaned and prepped, we loaded everything into a pandas DataFrame in Python. We then removed any duplicate entries to make sure our dataset was as accurate and reliable as possible. Finally, we saved the cleaned dataset as a CSV file that could be opened in tools like Microsoft Excel. However, one important issue we ran into was that when opening the file in Excel, the text encoding often got scrambled and was in Windows-1252 format. To avoid this, we found it necessary to use the "Get Data" function in Excel and manually import the

file with UTF-8 encoding selected into a blank workbook. This step ensures the characters display correctly and the data can be analyzed without issues.

For our second data source, we collected stock price information over time by scraping publicly available financial data from Yahoo Finance. This platform provides a large amount of historical stock data that is already relatively well-structured and clean, which made the initial data preparation process much easier compared to other sources. We didn't have to do extensive cleaning, but we did have to manipulate the raw data to extract the specific metrics and visualizations we needed for our analysis.

In our first round of data collection, we focused on a short window of time surrounding the announcement of Southwest's new baggage policy. This timeframe aligned with the Reddit posts we analyzed earlier, allowing us to make direct comparisons between public sentiment and stock market performance. After querying the Yahoo Finance API, the data was returned in JSON format, which we then converted and loaded into a structured format for processing. Instead of relying solely on the high or low prices for each trading day, we calculated an average of both the high and low values to get a more balanced and stable representation of the stock price for each day.

To observe the immediate financial impact of the baggage policy announcement, we grouped stock price data into two one week periods, the week before and the week after the announcement. From there, we calculated the percentage increase in stock prices during this time. This change was then visualized using a graph, where a clear upward trend becomes apparent. The visual representation helps emphasize the market's response, which we will discuss in more detail in the analysis section of our project.

However, to understand the full scope of how policy changes have influenced Southwest's stock over time, we expanded our timeline further. We traced stock prices all the way back to July of 2024, which is when the company first announced its shift in the assigned seating policy. Using the same method of averaging daily high and low prices, we created a time series graph spanning from July 2024 to March 2025. This allowed us to see fluctuations over a longer period and capture the effects of multiple major events.

In addition to the assigned seating and baggage policy changes, Southwest also went through a round of layoffs, which likely influenced investor behavior. These layoffs are marked on the graph as well, providing further context to the trends we observed. Overall, this broader stock price analysis gives us a clearer picture of how the company's strategic decisions and operational challenges have been reflected in the financial market over time.

Data Analysis and Evaluation

For this part, we focused on exploring and modeling the public sentiment around Southwest Airlines before and after major announcements, especially the baggage policy change on March 11, 2025.

We started with an exploratory analysis of the Reddit comments and Yahoo Finance data. Everything looked clean and Reddit comments had proper timestamps and content, and Yahoo Finance only had market-open dates (so weekends and holidays were naturally excluded). That was important because it helped us align the two datasets by date.

We used two different models to run sentiment analysis on the Reddit comments: VADER and BERT. VADER was easier to apply and gave us a compound score for each comment, which helped us quickly visualize the average sentiment over time. But it didn't do great with sarcasm or more nuanced language. As shown in Figure 1, most of the sentiment scores were leaning slightly positive, even when we knew from reading the comments that they were clearly negative or sarcastic. One big issue was that VADER was not picking up on these specificities.

To improve accuracy, we ran the same dataset through a pre-trained BERT model from Hugging Face. BERT doesn't give a continuous score like VADER, it just gives a label (POSITIVE or NEGATIVE) and a confidence score. So, we had to label Positive as 1 and negative as -1 and take averages over time. This model performed better at picking up sarcastic tones and overall intent. You can see that more clearly in Figure 2, where scores dipped more negative and had a bigger swing right after the policy drop. Still, even BERT missed a few complex sarcastic posts. For example, sarcastic posts like "Like I'd fly with the current administration" and "~~We do

things a little different ~~We're just like everyone else" were labeled as positive when they should have been negative.

It is important to note that we had to make the assumption that the BERT model was more accurate than VADER. A quick check of the classifications held up to this theory. We used an uncased version of BERT, a mini version of the actual BERT model. This choice was made so that we were able to efficiently run the code without it taking up too much computational space and time. We also wanted to make sure that the model we chose was not trained on Reddit comments so that it would not overfit our data.

The baseline here would be assuming all sentiment is neutral or using a single-label basic model like VADER. Compared to that, using BERT as a statistical model gave us stronger and more nuanced insights, even though it wasn't perfect.

We also recognize that no sentiment model is going to be 100% reliable, especially when sarcasm and slang are involved. That's why we visually inspected a few of the comments at major sentiment spikes to confirm whether the model's score actually made sense.

We chose to make sure that both graphs were broken down by hour instead of day because the Reddit thread only had comments across a tight 2-day window. The points are the mean of the compound scores or labels grouped by hour.

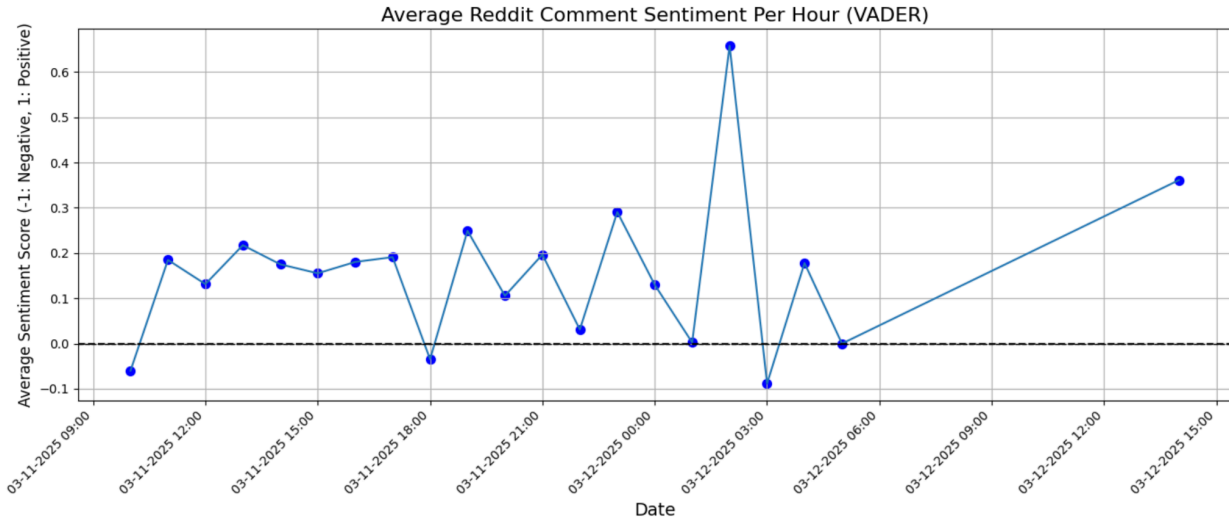


Figure 1: VADER Sentiment Analysis of Reddit Comments

As aforementioned, VADER labeled most of the data points as positive or neutral. A simple scroll through the subreddit would show that this is definitely not accurate. The negative scores that it did pick up on are not nearly as polar as they should be, which could be because of the average, but this graph showed us we needed a better model to capture true sentiment.

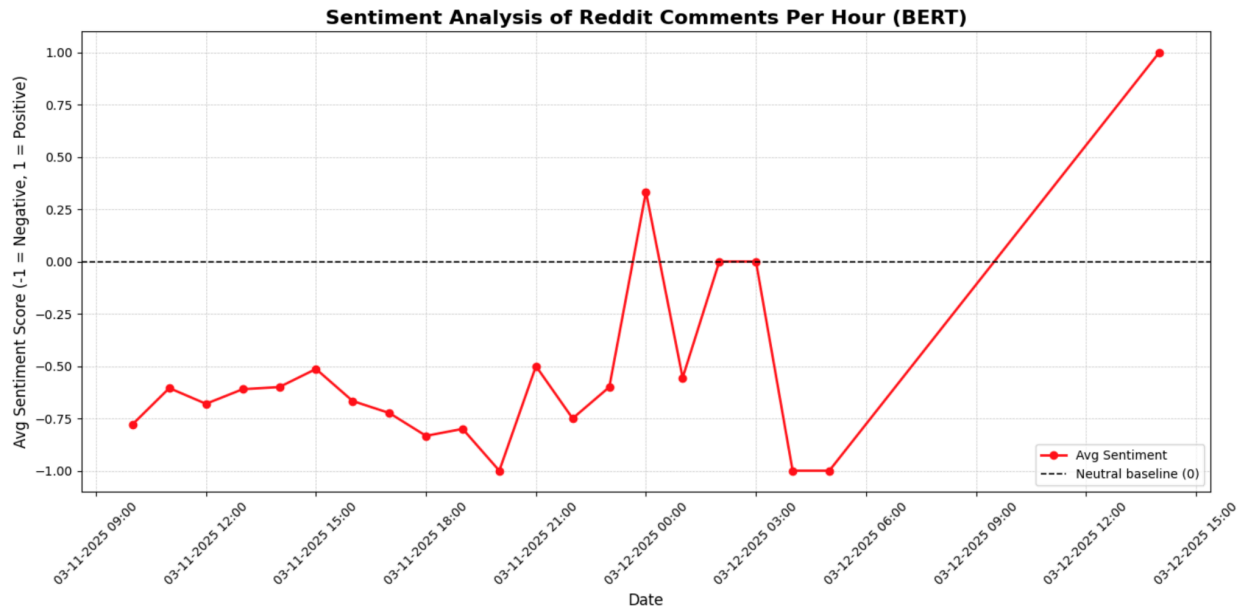


Figure 2: BERT Sentiment Analysis of Reddit Comments

BERT did a much better job of picking up on the nuanced and sarcastic language. The trend stays primarily below the x-axis at 0, and pretty close to negative one at that, so we can say with confidence that the BERT model is more accurate than VADER. The outlier on 3-12-2025 14:00 is apparent on both graphs, much more prominently in Figure 2, and this is where the sarcastic comments that BERT did not pick up on become obvious. This is likely because the number of posts started to die down at this point and these very complex sarcastic posts were the only ones in this time window, so they were not averaged out with any negatives.

To pair the sentiment data with real-world impact, we pulled Southwest's stock prices using Yahoo Finance. We looked at two timeframes: short-term (Figure 3) and long-term (Figure 4).

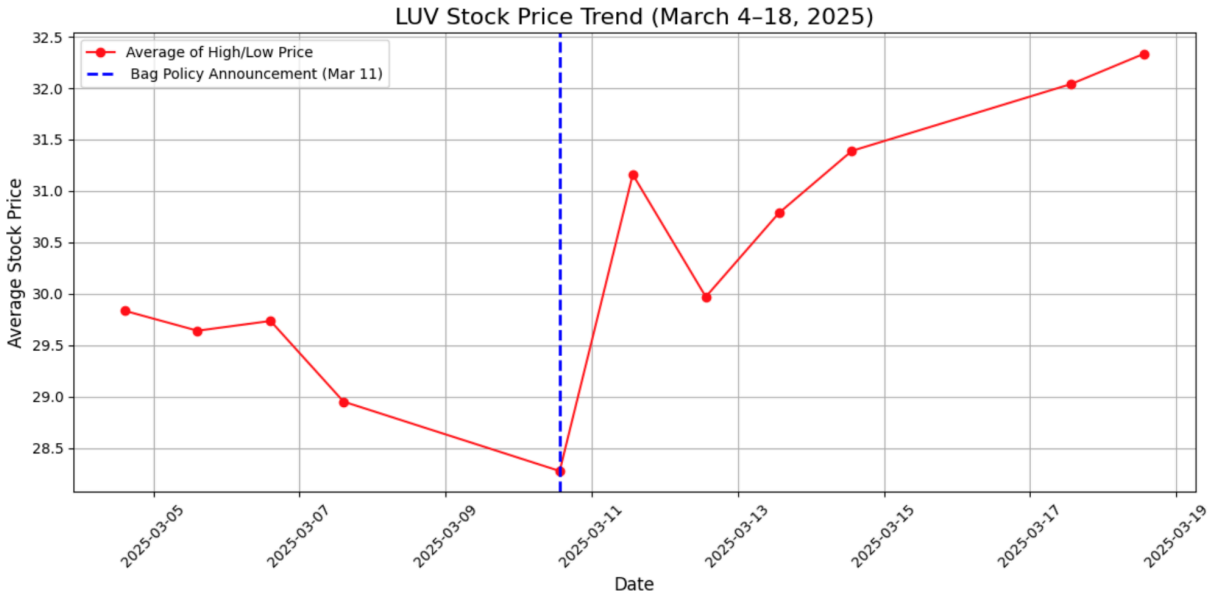


Figure 3: Baggage Policy Effect on Southwest Stock Price

In the short-term chart, we saw a pretty significant spike with a 6.81% increase in stock price right after the March 11 baggage policy announcement. It seems that Southwest was on a downward trend in their stock price before this. The stock market has many other factors that influence it other than just how the company is doing, like the economy as a whole. This spike was surprising considering the mostly negative sentiment from Reddit users. It might point to a disconnect between consumer perception and investor optimism, and investors may have seen it as a long-term efficiency move to keep up with other airline's revenue.

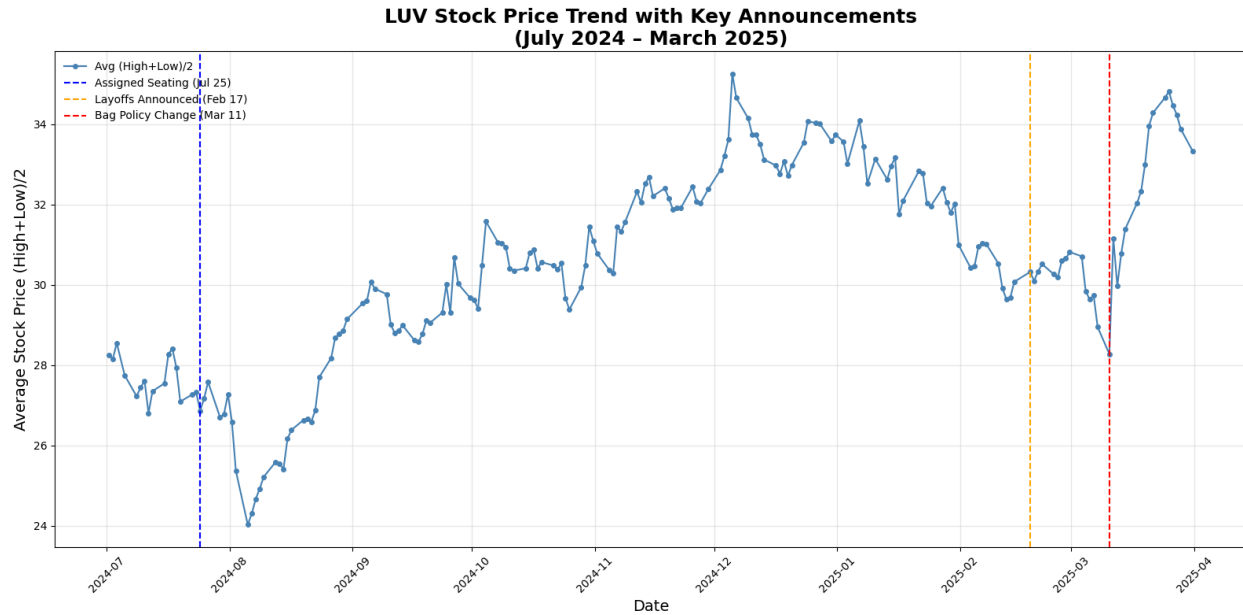


Figure 4: Extended Timeline of Southwest Stock Price

Zooming out to the long-term trend, we marked key events like the announcement of assigned seating in July 2024, layoffs in February 2025, and the bag policy change in March 2025. The trend line shows Southwest stock mostly responded positively to these announcements, especially the last one. The dip in August 2024 lines up with a general economic downturn, and the sharp decline in early March 2025 was driven by tariffs affecting the whole market, not just Southwest or the airline industry.

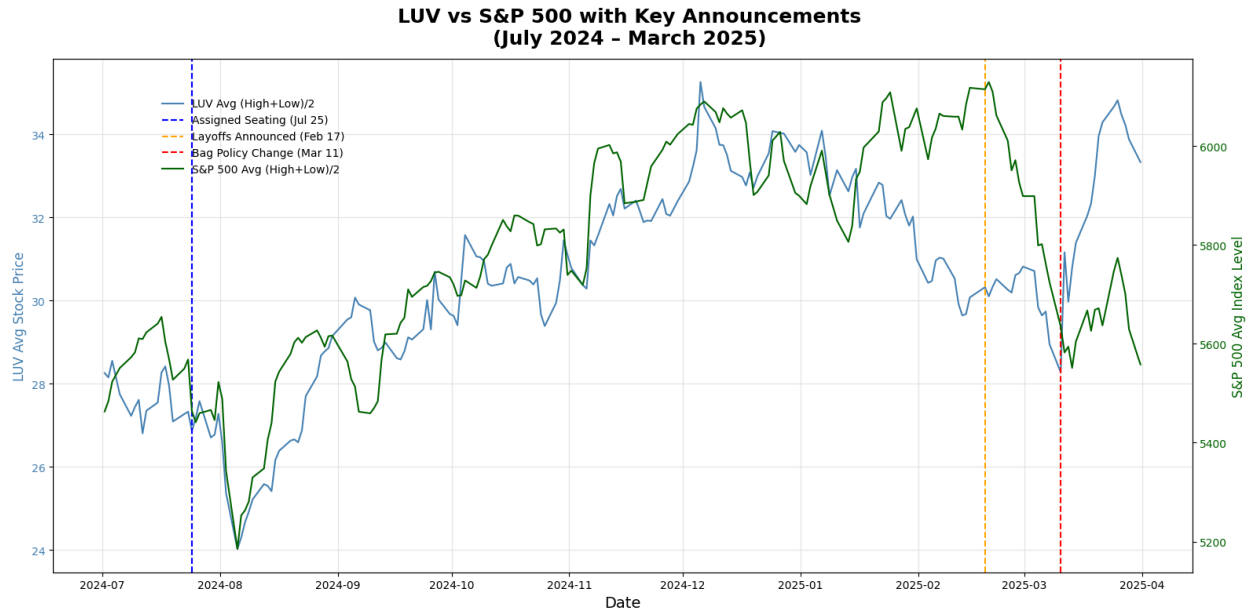


Figure 5: Extended Timeline of Southwest Stock Price vs the S&P 500

To get a better sense of how Southwest's performance compared to the broader market, we plotted LUV's average stock price against the S&P 500 index from July 2024 to March 2025. This helped us figure out if the reactions we were seeing, especially after the announcements, were specific to Southwest or part of a larger market trend. What stood out was that while both LUV and the S&P 500 dipped sharply in early March (likely due to external factors like imposing tariffs and a highly volatile economy), LUV recovered faster after the March 11 baggage policy announcement. This suggests there might have been a company specific boost tied to that news, even while the broader market stayed lower. It also adds more context to our earlier observations that LUV's post-announcement spikes were not just a fluke because it outpaced the S&P 500 in the same window, which points to a unique investor response to the announcements.

Another visualization that we wanted to make to tell more of the story was to overlay the stock price data directly on top of the sentiment analysis to see if there were any immediate, hour-by-hour market reactions to the baggage policy announcement. Our plan was to compare how sentiment changed in real time with how the stock price fluctuated during those same hours. However, we ran into a limitation with the free Yahoo Finance API, it only provides daily stock data, and not hourly. Because of this, we couldn't create a true side-by-side comparison between

hourly sentiment and stock price. Instead, we had to rely on daily averages to observe broader stock trends before and after the announcement.

Overall, these findings highlight a clear disconnect between consumer and investor viewpoints. Our analysis suggests Southwest's recent policy changes primarily reflect investor preferences rather than customer needs. Investors typically prioritize long-term financial outcomes, whereas customers focus on immediate experiences and satisfaction. This difference likely explains the wide gap in opinions. Ignoring customer dissatisfaction is risky, because if Southwest's loyal customers begin turning to competitors, the company could face significant challenges. We recommend Southwest prioritize customer sentiment in future policy decisions, as customers ultimately sustain the business by purchasing tickets.

Conclusion

While we were still able to create an applicable solution to Southwest's problem within a fairly short time frame, there are many other considerations we would have integrated had there not been a limited amount of time or free resources. Initially, we were interested in training our own model and using it alongside the pre-trained one which we later discovered could have been too time consuming considering the deadline. If we were able to dedicate more time to training our own model, this could have eliminated the issue of making an assumption that it was trained properly as we would have done it ourselves. We would have liked to compare the BERT model with our own to see not only the accuracy, but how similar they were to one another.

In addition, there are various resources that would have been more ideal for the data collection phase of our project. Given the stipulation that we were not allowed to use APIs previously utilized in class, The New York Times and News API keys were both eliminated for our use. In addition, Trustpilot data would have appealed to our group more if it contained an API key as opposed to raw data. As a result of this, it required us to web scrape, leading us to the largest issue in relation to Trustpilot—the data was thirty pages long which makes this process substantially more difficult.

Despite the varying components that could be improved upon, the benefits of this project and its solution are still quite prevalent to Southwest and its stakeholders. The greatest advantage of our solution is the duality of our evaluation through both the analysis of sentiment and stock prices. This allows us to gain perspective of not only the general customer opinion, but how investors view these policy changes in the context of the business. By being able to understand both views, Southwest is inclined to make better decisions based upon the company's priorities and values. Another benefit is that Southwest can continue to use this solution while more subreddits become available. Whether it is more articles on the now-old policy changes or new ones in the future, this provides an opportunity to keep monitoring public opinion in real time. Lastly, one great advantage to our solution is that if other companies become aware of this new strategy for Southwest, they can also easily integrate it into their company and improve their own decision making processes.

Although the benefits greatly outweigh the drawbacks to this solution, there are still a few stipulations that hinder its effectiveness. Primarily, the main disadvantage to using Reddit was that we only found one thread with enough comments that stemmed directly from the policy change. While we found more articles pertaining to Southwest, they either were not relevant to the business problem or did not contain a significant amount of comments to use, ultimately limiting our data for analysis. Additionally, we had to make the assumption that the BERT model was trained properly which could have reflected the accuracy of our solution.

REFERENCES

R/southwestairlines on reddit: Southwest will now charge you for checking bags.
(n.d.).

https://www.reddit.com/r/SouthwestAirlines/comments/1j8nu86/southwest_will_now_charge_you_for_checking_bags/

Southwest Airlines Co. (LUV) Stock Historical Prices & Data - Yahoo Finance. Yahoo! Finance. (n.d.). <https://finance.yahoo.com/quote/LUV/history/>

Yahoo! (2025, May 5). S&P 500 (^GSPC) charts, Data & News. Yahoo! Finance. <https://finance.yahoo.com/quote/%5EGSPC/>