TRACKING BOEING TWITTER SENTIMENT USING NLP AND DEEP LEARNING

By Sagar Garg

https://sagargarg.io/

https://www.linkedin.com/in/sagargarg1/

sagargarg@gmail.com

March 13th, 2022

Abstract

This paper uses Natural Language Processing (NLP) methods to analyze sentiment from Twitter data about The Boeing Company to show changes in public opinion over time. SNScrape, a Python library, was used to extract historical tweets from the last ten years related to Boeing. The Sentiment140 dataset contains 1.6 million annotated tweets representing the tweets' polarity. It is used to train and validate a Recurrent Neural Network (RNN) and assign extracted tweets a compound score on a negative to positive scale. The sentiment scores from the RNN model are aggregated and used to analyze how public opinion towards Boeing has shifted over the last 10 years to understand the significant factors and events that influenced opinion. Public sentiment towards Airbus was also analyzed to understand if the two companies faced similar issues in the commercial airliner market.

Keywords: Natural Language Processing, NLP, Deep Learning, RNN, NLTK, Python, 737 Max, The Boeing Company, Airbus, Aviation, Aerospace, Airplane, Crash, Accident, MCAS, Twitter, Sentiment Analysis, Brand Sentiment, Public Opinion, Public Relations, Reputation, Social Media

Introduction

The Boeing Company is currently facing its most turbulent period in its 105-year history due to the recent deadly crashes of two brand new Boeing 737 Max airplanes in October 2018 and March 2019 killing a total of 346 innocent lives. Both crashes were initiated by a single malfunctioning sensor that activated erroneously and forced the planes into nosedives shortly after takeoff. The 737 Max was promptly grounded by aviation authorities around the world while Boeing worked on a solution. More recently, the COVID-19 pandemic caused travel, especially international, to come to a screeching halt in March 2020. During this time, Boeing faced significant public outrage and frustration due to these recent operational failures and service interruptions. As the incidents fade from memory and the world looks to resume domestic and international travel, Boeing can benefit from public sentiment tracking and analysis to monitor public relations and marketing efforts and understand the driving factors of public sentiment towards the company in recent years to determine the best course of action to improve.

By inspecting historical events that caused an abrupt change in public opinion – both positive and negative – the PR team would discover ways to improve Boeing’s public outlook. This would occur by either focusing on positive announcements or working on damage control in the case of a disaster and taking the appropriate actions to minimize damage done to Boeing’s reputation. It is also important to track public sentiment towards Airbus to understand if they face similar issues in terms of sentiment or even inversely benefits when Boeing faces a negative public relation in the commercial airliner market. The commercial airliner market is held in a global duopoly between Boeing and Airbus, controlling 88% of the airliner industry market (Morris 2016).

Twitter was chosen to be the social media platform to scrape data from due to being a rich source of public data comprised of real-life conversations by various people and organizations. Ten years' worth of tweets mentioning Boeing were exacted from Twitter. A Recurrent Neural Network (RNN) model was developed to analyze sentiment from Twitter data after extensive hyperparameter testing. RNNs are ideal in situations involving a sequence of numbers, such as sentences where each word is represented as a token. Manually tagging tweets for training as positive or negative can be exhausting but is the most crucial part to ensure accurate training data. The Sentiment140 dataset trained the neural network model to prevent this manually tagging process. The Sentiment140 dataset contains 1,600,000 tweets extracted from Twitter that have been annotated to represent the polarity of the tweet.

In a search to find the best-performing model, tests will be conducted by changing architectural features such as the number of layers and hidden nodes, regularization factor, and directionality. The overall best performing model assigns a sentiment score for all tweets and after aggregation of scored, used to track public opinion towards Boeing and Airbus over the last 10 years and inspect historical events that caused an abrupt change in public opinion to discover the best way to improve Boeing's public outlook.

Background

In October 2018, a brand-new Boeing 737 MAX jet operated by Lion Air crashed into the Java Sea, tragically killing all 189 passengers and crew on board. A few months later, in March 2019, a second Boeing 737 MAX jet, this time operated by Ethiopia Airlines, crashed onto land only six minutes after take-off, killing all 157 passengers and crew on board. The Boeing 737 MAX was immediately grounded worldwide while the National Transportation Safety Board (NTSB) investigated the cause of both accidents. The NTSB found out that the cause of both accidents was a malfunctioning sensor in a redesigned flight control system. To quickly introduce the aircraft alongside the rival Airbus A320Neo, Boeing created the Maneuvering Characteristics Augmentation System (MCAS) so the plane would fly identically as previous models of the 737, decreasing pilot training time typically needed. However, Boeing did not mention this new system that altered the plane's trajectory since, ideally, it would perform like previous models. Still, pilots did not know how to overcome the system when this system gave warnings due to faulty sensors. Soon, the Boeing Company and later the FAA came under extreme scrutiny for the speedy certification process of the 737 Max, leading to public outrage of both organizations, especially towards Boeing. It took Boeing 20 months to develop fixes to the MCAS system and 737 Max, and was certified to safely fly by international aviation authorities in December 2020.

Literature Review

Sentiment Analysis and NLP, in general, is a relatively new area of development, but the entire field has grown tremendously in the last few years thanks to improved algorithms. With the world of unstructured data expanding, the ability to derive sentiment from text increases in popularity. One way of deriving the meaning of a text is to use sentiment analysis to classify a piece of text as containing either a positive, negative, or neutral connotation. The most common method is to apply a sentiment score to each token included in the text after tokenization. Based on the proximity to either positive or negative words, the sentiment analysis method can assign a sentiment score towards different topics found in the text. By taking the sum of the sentiment score, the text can be classified into a spectrum of varying polarity categories ranging from very positive to very negative (Appel et al., 2016).

Sentiment analysis can be used in two ways, for real-time processing or analyzing historical data, with many relevant works in both methods related to using Twitter data. In a paper for the American Society of Civil Engineers, Tang, Zhang, Dai, and Yoon identified four user clusters: construction works, construction companies, construction unions, and construction media and analyzed sentiment analysis, topic modeling, link analysis, geolocation analysis, and timeline analysis to gain a better understanding of real-world situations in the construction industry. A sentiment analysis result yielded that construction workers tended to have a more significant proportion of negative tweets than other clusters (Tang et al., 2017). In another study, Michael Caballero used Twitter data to predict the 2020 U.S. Presidential Election by combining sentiment analysis from users critical to the election and aggregate polling and a time series analysis (Caballero 2021). Zhang, Yi, Chen, and He used sentiment analysis on COVID-19 Twitter tweets from eight major North American cities concerning masks, vaccines, and lockdown measures in another research paper. Their study found public sentiment to vary by period and location, but people generally had a favorable view of masks and a negative view about vaccines and lockdown (Zhang et al., 2021).

There have been two significant applications of NLP within the aerospace industry specifically. The first was a paper by Rose, Puranik, and Mavris that focused on identifying clusters related to aviation safety incidents for commercial flights to uncover new trends not evident in existing data labels (Rose et al., 2020). The second, related to sentiment analysis, was a publicly available dataset on Kaggle about Twitter data scraped from major U.S. airlines to analyze how travelers expressed their feelings, which could provide better customer service (Hosseini 2020).

The classifier this paper focuses on is a Recurrent Neural Network, a model designed to work with sequential data, such as text, audio, or time-series data. RNNs can form a much deeper understanding of a sequence, allowing them to be more precise in predicting future series. The main drawback with a basic RNN is the short-term memory problem caused by a strong vanishing gradient. The gradient is exponentially shrunk down by the end of the steps, halting updates to the weight and preventing the network from leaning the effects of earlier inputs (Phi, 2018). The solution for vanishing gradients was solved with two specialized versions of an RNN, Long Short-Term Memory (LSTM) and a Gated Recurrent Unit (GRU). Both LSTM and GRU models use a memory cell to store the activation value of previous words in long sequences. Gates are also used to control the flow of information in the network and learn which essential inputs are stored in a memory unit (Pedamallu, 2020).

Data

Model Testing Data

Computer Science graduate students created the Sentiment140 dataset at Stanford University in 2010 (Go et al., 2010). Sentiment140 was chosen to train the neural network since the dataset contains 1.6 million tweets extracted from Twitter and been annotated to represent the polarity of the tweet. The dataset also includes five other fields of data, the tweet ID, date, flag (the query used), the user of the tweet, and the most important feature being the text of the tweet. A score of 0 represents negative, 2 represents neutral, and 4 represents a tweet with positive sentiment.

Data Collection and Preprocessing

There are various methods to extract data from Twitter, and it is dependent on the use. This library was chosen over the official Twitter API for its ability to easily extract historical tweets from January 2012 to December 2021. The Twitter API currently has restrictions that prevent searches older than seven days ago, while SNScrape could pull historical tweets. The term 'Boeing' was queried while filtering out retweets and including only English tweets to get a wide range of meaningful tweets. Extracted with each tweet was the complete text, id, date, user information (display name, number of followers, number of followings, and statuses count), and finally interaction counts (retweets, replies, likes, and quotes) as seen in Figure 1. Figure 2 describes the numerical columns statically. A total of 6.5 million tweets were extracted covering the past ten years. Appendix Figure A1 shows a Python code snippet demonstrating how tweets were extracted using SNScrape. The next step is to preprocess the text for the algorithm to extract the sentiment behind the tweets more easily. To do this, links and non-alphanumeric values were removed using Regex. Initially, only alpha characters would be included; however, the tweet would lose crucial information about the Boeing aircraft model such as 737. Another improvement was removing mentions with the username instead of only the '@' symbol to remove unnecessary words for the model to process. Finally, default stopwords were removed, and then the entire tweet text was lemmatized before appending back to the original dataframe. Figure 3 compares the original tweet to the processed text for the first ten tweets.

```
RangeIndex: 6539154 entries, 0 to 6539153
Data columns (total 19 columns):
 #   Column            Non-Null Count    Dtype
---  ------            --------------    -----
 0   Tweet Id          6539154 non-null  int64
 1   Datetime          6539154 non-null  object
 2   Text              6539154 non-null  object
 3   Username          6539152 non-null  object
 4   Displayname       6537842 non-null  object
 5   User Description  2434662 non-null  object
 6   User Link         1604099 non-null  object
 7   User Followers    6539154 non-null  int64
 8   User Following    6539154 non-null  int64
 9   User Statues      6539154 non-null  int64
 10  User Favorites    6539154 non-null  int64
 11  User Location     4963744 non-null  object
 12  Replies           6539154 non-null  int64
 13  Retweets          6539154 non-null  int64
 14  Likes             6539154 non-null  int64
 15  Quotes            6539154 non-null  int64
 16  TextClean         6537919 non-null  object
 17  NLTK_Score        6539154 non-null  float64
 18  RNN_Score         6539154 non-null  float64
dtypes: float64(2), int64(9), object(8)
memory usage: 947.9+ MB
```

Figure 1: Pandas DataFrame of Boeing Tweets

	User Followers	User Following	User Statues
count	6539154	6539154	6539154
mean	69856	2206	114009
std	1087066	12102	252539
min	0	0	0
25%	185	108	7233
50%	806	535	29882
75%	3141	1656	97540
max	130733257	4203331	6962784

	User Favorites	Replies	Retweets	Likes	Quotes
count	6539154	6539154	6539154	6539154	6539154
mean	13183	0	1	3	0
std	45692	8	35	104	6
min	0	0	0	0	0
25%	6	0	0	0	0
50%	454	0	0	0	0
75%	6648	0	0	0	0
max	1820619	11489	31564	56210	5637

Figure 2: Describing Numerical Columns

```
                                                                                                    Text
0        @conspiracyb0t People just blindly believe whatever they are told. They failed to see that every...
1        I don't know about you folks but, I want one of the graduates of this school to pilot my next va...
2             @dezmondOliver @elonmusk Even worse he could ask Boeing to take care of it with their rocket.
3        @benjaminpacini Yes, it is true that as of today, Jeff Bezos cannot order a military strike on E...
4                                                                               @asm_just I work for Boeing
...                                                                                                      ...
6539149  BE: Fair and 43 F at Seattle/Boeing Fld, WA Winds are Northeast at 5.8 MPH (5 KT). The pressure ...
6539150  RT @indoflyer: Mengenang 5 tahun Adam Air KI574, Boeing 737-400, PK-KKW. We hope there is no mor...
6539151  Mengenang 5 tahun Adam Air KI574, Boeing 737-400, PK-KKW. We hope there is no more airline that ...
6539152  @HiLeslee_ nah , lol \nIt's Boeing & I needa use the bathroom -______-\n& grandma's house ? \nI ...
6539153  Boeing: Dividend Dynamo, or the Next Blowup? - Motley Fool: Boeing: Dividend Dynamo, or the Next...

                                                                                               TextClean
0        people blindly believe whatever told failed see everything cover 60 year perfect example ford mo...
1        know folk want one graduate school pilot next vacation boeing 747 could go wrong maybe could nex...
2                                                         even worse could ask boeing take care rocket
3        yes true today jeff bezos cannot order military strike etsy foreign country however clear lockhe...
4                                                                                           work boeing
...                                                                                                      ...
6539149                  fair 43 f seattle boeing fld wa wind northeast 5 8 mph 5 kt pressure 1027 2 mb hu
6539150               rt mengenang 5 tahun adam air ki574 boeing 737 400 pk kkw hope airline cut cost much
6539151                  mengenang 5 tahun adam air ki574 boeing 737 400 pk kkw hope airline cut cost much
6539152                            nah lol boeing needa use bathroom grandma house might call like 12 03 lol
6539153            boeing dividend dynamo next blowup motley fool boeing dividend dynamo next investing brk
```

Figure 3: Original Text and Preprocessing Text

Methods

Methodology Overview

Three sets of experiments were ducted to understand how different hyperparameters affect model performance and process time to ultimately develop the best-performing model to calculate sentiment from processed text. The hyperparameter tuning involved testing 1-layer vs. 2-layer, regularization vs. no regularization, and changes in directionality. Each set was initially chosen to include 12 models, each tuning the architecture and hyperparameters of a Simple RNN, LSTM, and GRU models, created using the Python TensorFlow library. However, due to computational constraints preventing them from being computationally feasible several 2-layer LSTM and GRU models were not implemented. The NLTK SentimentIntensityAnalyzer package is also tested on the Sentiment140 dataset and used as a benchmark comparison. The best-performing model is used to assign a valence-based compound score representing the polarity of sentiment for all tweets related to Boeing over the last ten years.

Implementation and Programming of RNN Model

A total of 24 experiments were conducted on variations of an RNN, LSTM, and GRU model. The Python TensorFlow Keras library was used to create all of the RNN models. Appendix Figure A2 and A3 shows how to implement a 2-layer bidirectional LSTM with regularization. The regularization technique chosen was dropout with a hyperparameter value of 0.3 while using 64 hidden units in all layers. Dropout represents the probability of a randomly selected node dropping out of the network before the fully connected layer. Figure 4 on the next page shows the accuracy and loss for train/test/validation for LSTM and GRU.

#	Type	Train Acc.	Train Loss	Val Acc.	Val Loss	Test Acc.	Test Loss	Process Time
B1	1-Layer LSTM Unidirectional (no dropout)	0.7729	0.4695	0.7556	0.4995	0.7497	0.5073	31:36
B2	1-Layer LSTM Bidirectional (no dropout)	0.7650	0.4802	0.7541	0.5001	0.7506	0.5055	39:58
B3	1-Layer LSTM Unidirectional (dropout)	0.7726	0.4717	0.7591	0.4926	0.7577	0.4941	33:34
B4	1-Layer LSTM Bidirectional (dropout)	0.7731	0.4682	0.7581	0.4957	0.7544	0.5008	41:51
B5	2-Layer LSTM Unidirectional (no dropout)	0.7711	0.4726	0.7552	0.4961	0.7503	0.5074	52:16
B6	2-Layer LSTM Bidirectional (no dropout)	0.7859	0.4490	0.7583	0.4989	0.7538	0.5147	2:28:11
B7	2-Layer LSTM Unidirectional (dropout)	0.7788	0.4635	0.7598	0.4925	0.7581	0.4975	2:07:12
B8	2-Layer LSTM Bidirectional (dropout)	0.7773	0.4644	0.7569	0.4985	0.7566	0.4982	3:12:10

Figure 4: Benchmark Metrics for an LSTM Networks

Appendix Figure A4 and A5 shows the benchmark metrics for a RNN and GRU network. The model with the best overall performance and reasonable process time was model B3, with a test set accuracy of 75.77% and a processing time of 33 minutes. Model B3 was a 1-Layer Uni-Directional LSTM unit with dropout applied. The model with the best performance based solely on test set accuracy was model B7, a 2-Layer Uni-Directional LSTM unit with dropout used. Model B7 had an accuracy of 75.81%; however, it took over 2 hours to process.

The confusion matrix for the overall winning model – B3 – is shown in Appendix Figure A6, and the training vs. validation set accuracy by epoch graph is shown in Figure 5 below.

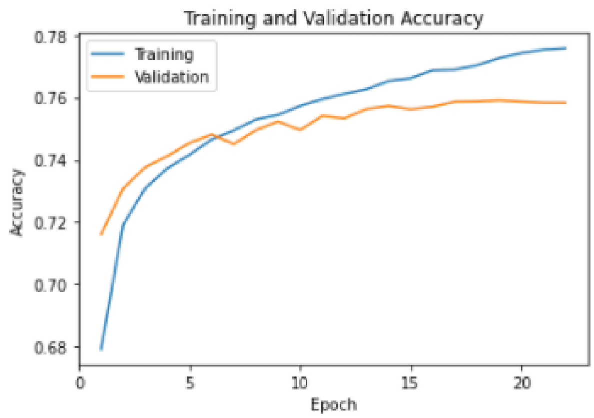


Figure 5: Training vs. Validation Set Accuracy by Epoch

To compare the results to the SentimentIntensityAnalyzer package, the compound score was scaled to match the original score range representing the polarity of the tweet. The accuracy of the package was 63.76%. The confusion matrix for the SentimentIntensityAnalyzer is shown in Appendix Figure A7.

Results

Figure 6 shows the scores for the first ten tweets, and at an initial glance, look reasonably accurate. Both the RNN and SIA model were used to score each extracted tweet. Out of 6.5 million tweets, 1.3M were tagged as neutral, 2.3M tagged as positive, and 2.9M tweets as unfavorable, while Figure 7 shows a histogram representing the distribution of scores for both models

	TextClean	NLTK_Score	RNN_Score
0	people blindly believe whatever told failed see everything cover 60 yea...	5.5135	4.739391
1	know folk want one graduate school pilot next vacation boeing 747 could...	1.4520	1.918240
2	even worse could ask boeing take care rocket	5.1290	2.077522
3	yes true today jeff bezos cannot order military strike etsy foreign cou...	9.3400	4.878884
4	work boeing	5.0000	3.406874
5	happy belated birthday bobby angel 85 year young still going like boeing	8.6755	8.938408
6	well ya think boeing	6.3660	7.271252
7	friend visit seattle wrote asking give ride boeing tour said san diego ...	8.2430	5.308849
8	swa3834 n7841a boeing 737ng 719 w southwest airline squawk 0766 4 2 mi ...	5.0000	3.188233
9	baby boeing 747	5.0000	5.429285

Figure 6: SentimentIntensityAnalyzer Results

Figure 7: Histogram of Score Distribution

To expand on this research, tweets related to Airbus over the same period were also used to compare by similarly using the RNN model to score sentiment. There were around 3M tweets with the mention of Airbus from the last 10 years.

Analysis and Interpretation

The sentiment scores from the RNN model are aggregated by day and a 14-day moving average is used to analyze how public opinion towards Boeing has shifted over the last 10 years as show in Figure 8 with a corresponding graph of the daily volume of tweets.

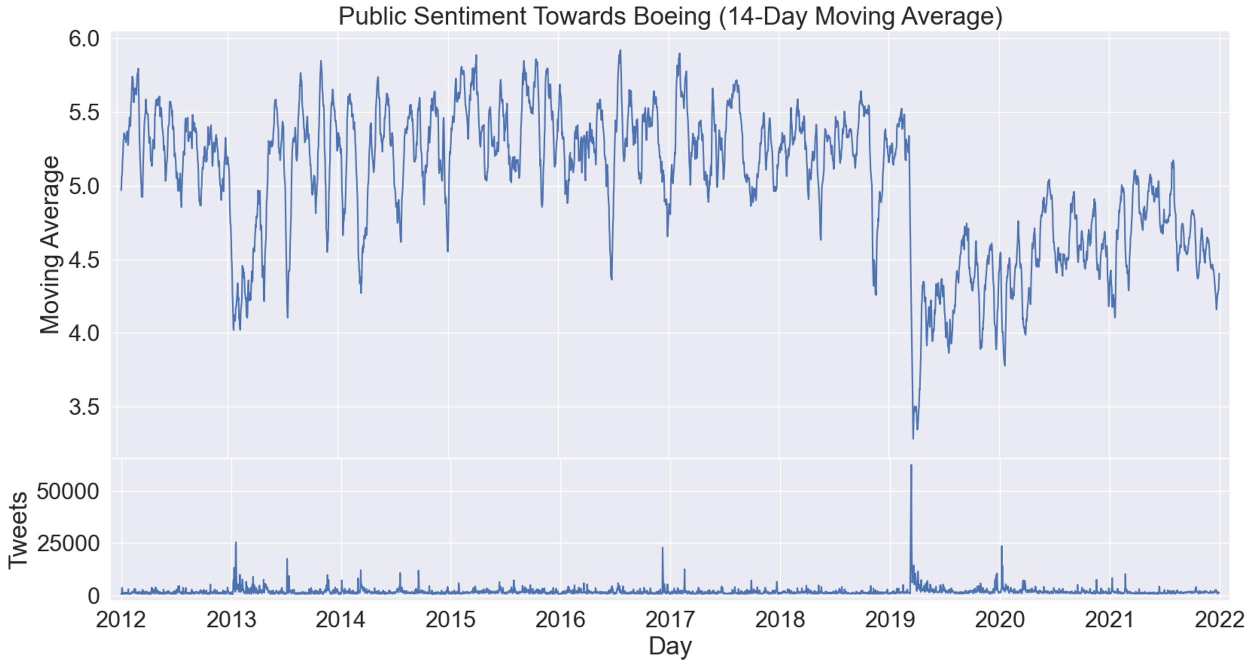


Figure 8: 14-Day Moving Average of RNN Public Sentiment with Volume Plot

The events of the two separate Boeing 737 Max crashes had varying effects on public sentiment. After the Lion Air Crash in October 2018, sentiment dropped from a high average of 5.5 to 4.3 before quickly rebounding to the pre-crash score. However, after the Ethiopia Airlines crash in March 2019, public sentiment crashed to an all-time low of 3.2 after the crash was determined to be a preventable issue from Boeing. In the weeks and months following the crash, the sentiment score was slower to improve and did not immediately rebound back to normal like previously. Along with those two events, there were other significant spikes in number of tweets worth exploring. Figure 9 shows the days with the most posted tweets and the corresponding sentiment score.

Day	Tweets	Score
2019-03-13	62493	3.176652
2019-03-12	49526	2.939629
2019-03-11	31962	2.705522
2019-03-14	26242	3.237184
2013-01-17	25366	3.428244
2013-01-16	24375	3.786016
2020-01-08	23617	2.438014
2016-12-06	22871	4.516320
2013-07-06	17550	3.100851
2019-03-10	15315	2.339139
2019-03-21	14400	3.642371
2020-01-10	14208	3.473920
2016-12-07	13977	5.134306
2013-01-18	13445	4.049262
2013-01-11	13252	3.843436
2017-02-17	12512	6.174497

Figure 9: Days with Most Tweets and Average Sentiment Score

To understand the most trending tweets, a word cloud map was generated for tweets on days with a high number of posted tweets and can be found in Appendix D. The word cloud map was used to highlight trending topic on Twitter for the following dates chosen from Figure 13 in chronological order:

- January 16th, 2013 – The FAA issued an emergency airworthiness directive to ground all Boeing 787s after an All-Nippon Airways emergency landing caused by battery overheating issues.
- July 6th, 2013 – Asiana Airlines Flight 214, a Boeing 777-200ER, crashed on final approach into SFO, killing 3 passengers and injuring 187. The cause of the accident was pilot error, a mismanagement of the airplane's final approach by the flight crew.
- December 6th, 2016 – President Trump threw criticism about Boeing regarding the next generation of Air Force One by tweeting, "Boeing is building a brand new 747 Air Force One for future presidents, but costs are out of control, more than $4 billion. Cancel order!"

- February 17th, 2017 – President Trump became the first sitting president to visit Boeing's campus in North Charleston, South Carolina, where he attended the unveiling of Boeing's new 787-10 Dreamliner.
- October 30th, 2018 and March 11th, 2019 represent the day after the tragic 737 Max crashes of Lion Air Flight 610 and Ethiopian Airlines Flight 302 respectively.
- Finally, January 8th, 2020 represents the day after Ukraine International Airlines Flight 752 was shot down shortly after takeoff from Tehran killing all 176 passengers and crew on board.

The events between December 2016 and February 2017 are examples of events that provide sway to public opinion towards Boeing in both directions. The unveiling of the new 787-10 Dreamliner along with having President Trump as a speaker helped negate the effect of previous criticism by President Trump, while also reaching an all-time high in terms of public sentiment. Additional figures with analysis can be found in Appendix B.

Public Opinion Towards Airbus

Since the commercial airliner market is held in a global duopoly between Boeing and Airbus, understanding public opinion related to Airbus can be beneficial to see if Airbus have faced similar issues in the past or have inversely benefitted at the recent downfall of Boeing. The sentiment scores for Airbus related tweets are aggregated by day and a 14-day moving average is applied to see how the public reacted to Airbus over the last 10 years. Figure 10 shows an overlay of public sentiment and tweet volume for Boeing and Airbus on the next page. There were five significant spikes in the number of tweets worth exploring. Figure 11 shows the days with the heaviest tweet volume and the corresponding sentiment score.

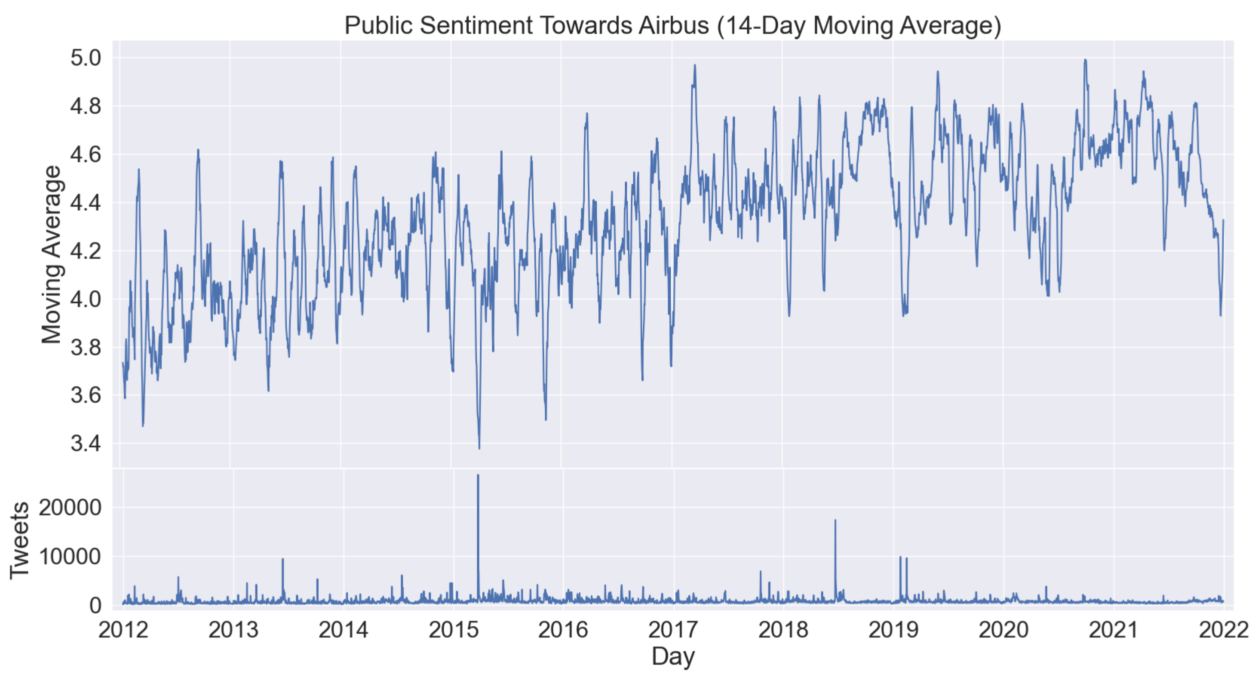


Figure 10: Public Sentiment Towards Airbus

Day	Tweets	Score
2015-03-24	26585	2.013526
2018-06-22	17380	3.620024
2019-01-24	9879	3.561838
2019-02-14	9614	2.633025
2013-06-14	9463	4.977307

Figure 11: Days with Most Tweets and Average Sentiment Score

A word cloud map was generated for tweets on these days can be found in Appendix E and used to highlight trending topics on Twitter for the dates chosen from Figure 11 in chronological order:

- June 14th, 2013 – Airbus A350 successfully completes its maiden test flight on schedule; the A350 was designed to be a direct competitor to Boeing's 787 Dreamliner.
- March 24th, 2015 – Germanwings (a Lufthansa subsidiaries) Flight 9525, an Airbus A320-211, was deliberately crashed by the co-pilot killing all 150 passengers and crew on board.
- June 22nd, 2018 – On the previous day, Airbus released an announcement demanding the UK government details of a Brexit Deal. Otherwise, Airbus warned that it would reconsider long-term investments and leave Britain.
- January 24th, 2019 – On the previous day, Airbus and its CEO Tom Enders released a video message stating if there was a no-deal Brexit, Airbus would have to make potentially very harmful decisions for the UK economically.
- February 14th, 2019 – Airbus announced that due to lack of substantial A380 backlog, the company would halt deliveries in 2021. Airbus delivered the last A380 to Emirates in December of 2021.

Additional figures related to Airbus can be found in Appendix C.

Boeing vs. Airbus Public Sentiment

Finally, in an effort to compare public sentiment towards Boeing and Airbus, Figure 12 overlays the two historical sentiment graphs.

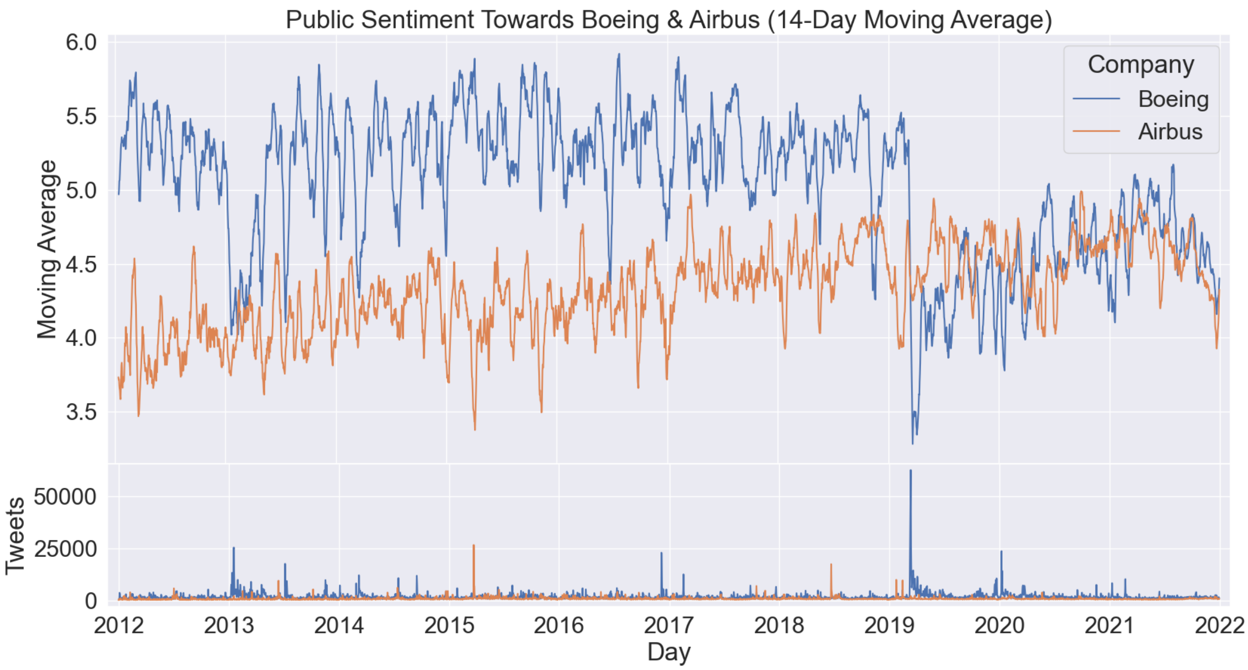


Figure 12: Public Sentiment Towards Boeing & Airbus

Up until March 2019, Boeing had maintained a much stronger public outlook, though Airbus had worked towards gradually improving public sentiment. After the second 737 Max crash, views towards Boeing turned negative and struggled to return even after the return of the 737 Max in December of 2020; unlike after the grounding of the 787 in January 2013.

Conclusion

The Boeing Company is currently facing its most challenging period in recent history, starting with the 737 Max crashes in October 2018 and March 2019. Given the significant outrage and frustration towards Boeing over the past few years, it is crucial to understand how public opinion has changed over the years before determining the best course of action to improve. By inspecting historical events that caused an abrupt change in public opinion – both positive and negative – the PR team would discover ways to improve Boeing's public outlook. This included tracking public sentiment towards Airbus to understand if they face similar issues in the commercial airliner market. Twitter provided the ideal platform to collect data comprised of real-life conversations by various people and organization in order to track the public's opinion over the years. A Recurrent Neural Network (RNN) model was developed to analyze sentiment from Twitter data after extensive hyperparameter testing.

As expected, it was found that Boeing had maintained a much stronger public outlook until the second 737 Max crash in March 2019. During this time, Airbus had worked to gradually improve public perception, but faced similar PR crises. It is not an easy task to improve public opinion, especially in recent years that have been plagued with the COVID-19 pandemic, affecting international travel. Fortunately, as seen with events such as the unveiling of the Boeing 787 Dreamliner in February 2017 or the on-schedule test flight of the Airbus A350 in June 2013, it is possible for Boeing to redirect a positive outlook in the future. It will surely take time to regain the public's trust from recent events from the 737 Max crashes. With a re-certified 737 Max that started flying again in December 2020, as well as focusing PR efforts on successful program milestones such as the upcoming 777X program, Boeing can make really positive changes and return to being the aviation giant it once was.

Directions for Future Work

Using historical events can be useful in understanding specific events and timelines that had the biggest impact on public sentiment and ways to improve going forward. However, it would be more beneficial to track real-time sentiment in order for the right PR team to get on top of issues and issue announcements to the public before getting out of hand. The next step would to modify the process of data collection to be continual and provide a real-time dashboard monitor public sentiment and be able to identify trending topics related to Boeing. Early detection for a disaster would allow public relation teams to minimize damage to the company's reputation, such as holding press conferences, apologies, campaigns.

References

Appel, Orestes, Francisco Chiclana, Jenny Carter, and Hamido Fujita. 2016. "A hybrid approach to the sentiment analysis problem at the sentence level." *Knowledge-Based Systems* 108: 110-124. https://doi.org/10.1016/j.knosys.2016.05.040.

Caballero, Michael. 2020. "Predicting the 2020 U.S. Presidential Election With Twitter." *2nd International Conference on Soft Computing, Artificial Intelligence and Machine Learning* 2021: 53-65 https://arxiv.org/abs/2107.09640.

Desai, Rashi. 2021. "How to Scrape Millions of Tweets Using Snscrape." Medium DataSeries. Published June 21st, 2021. https://medium.com/dataseries/how-to-scrape-millions-of-tweets-using-snscrape-195ee3594721.

Go, A., Bhayani, R., & Huang, L. (2010). A Twitter sentiment analysis tool. Sentiment140. Retrieved November 20th, 2021, from http://help.sentiment140.com/.

Hosseini, Soheii. 2020. "Twitter U.S. Airline Sentiment Analysis." Towards Data Science. Published June 14th, 2020. https://towardsdatascience.com/twitter-us-airline-sentiment-analysis-91caa7a22a93.

Lane, Hobson, Cole Howard, and Hannes Hapke. 2019. "Chapter 4: Finding Meaning in Word Counts (Semantic Analysis)." In *Natural Language Processing in Action*. Shelter Island, NY: Manning Publications.

Morris, Rob. 2016 "Insight from Flightglobal: Flight Fleet Forecast's Single-Aisle Outlook 2016-2035." Flight Global. Published November 19th, 2020. https://www.flightglobal.com/insight-from-flightglobal-flight-fleet-forecasts-single-aisle-outlook-2016-2035/121497.article.

Pedamallu, H. (2020, November 30th). RNN vs. GRU vs. LSTM. Medium. Retrieved October 29th, 2021, from https://medium.com/analytics-vidhya/rnn-vs-gru-vs-lstm-863b0b7b1573.

Pennington, J., Socher, R., & Manning, C. D. (2014). GloVe: Global Vectors for Word Representation. Glove: Global vectors for word representation. Retrieved November 27th, 2021, from https://nlp.stanford.edu/projects/glove/.

Phi, M. (2020, June 28th). Illustrated guide to LSTM's and GRU's: A step-by-step explanation. Medium. Retrieved October 28th, 2021, from https://towardsdatascience.com/illustrated-guide-to-lstms-and-gru-s-a-step-by-step-explanation-44e9eb85bf21.

Ramsundar, B., & Zadeh, R. B. (2018). TensorFlow for Deep Learning. O'Reilly Media, Inc.

Roldós, Inés. 2020. "Track Brand Sentiment – Why It's Important." MonkeyLearn. Published March 27th, 2020. https://monkeylearn.com/blog/brand-sentiment/.

Rose, Rodrigo, Tejas Puranik, and Dimitri Mavris. 2020. "Natural Language Processing Based Method for Clustering and Analysis of Aviation Safety Narratives." *Aerospace* 7, no. 10: 143. https://doi.org/10.3390/aerospace7100143.

Tang, LiYaning, Yiming Zhang, Fei Dai, Yoojung Yoo, Yangqiu Song, and Radhey Sharma. 2017. "Social Media Data Analytics for the U.S. Construction Industry: Preliminary Study on Twitter." *Journal of Management in Engineering* 33, no 6. https://doi.org/10.1061/(ASCE)ME.1943-5479.0000554.

Zhang, Qihuang, Gace Y. Yi, Li-Pang Chen, and Wenqing He. 2021. "Text Mining and Sentiment Analysis of COVID-19 Tweets." *ArXiv*. https://arxiv.org/abs/2106.15354.

Appendix A

```python
# importing libraries and packages
import snscrape.modules.twitter as sntwitter
import pandas as pd

# Creating list to append tweet data
tweets_list = []

# Using TwitterSearchScraper to scrape data and append tweets to list
for i, tweet in enumerate(sntwitter.TwitterSearchScraper('Boeing since:2018-10-01 until:2021-07-20 lang:en').get_items()):
    if not i % 1000: #number of tweets you want to scrape
        print(f"Scraping, {i} results so far on {tweet.date.strftime('%Y-%m-%d')}")
    tweets_list.append([tweet.id, tweet.date, tweet.content, tweet.user.username, tweet.user.displayname, tweet.user.followersCount,
tweet.user.friendsCount, tweet.user.statusesCount, tweet.user.favouritesCount, tweet.user.location, tweet.replyCount, tweet.retweetCount,
tweet.likeCount, tweet.quoteCount, tweet.retweetedTweet])

# Creating a dataframe from the tweets list above
tweets_df = pd.DataFrame(tweets_list, columns=['Tweet Id', 'Datetime', 'Text', 'Username', 'Displayname', 'User Followers', 'User Following',
'User Statues', 'User Favorites', 'User Location', 'Replies', 'Retweets', 'Likes', 'Quotes', 'Retweet Original'])
tweets_df['Datetime'] = tweets_df['Datetime'].dt.tz_localize(None)
tweets_df.to_excel('boeing_tweets_all.xlsx')
```

Figure A1: Python Code Snippet to Extract Tweets from snscrape

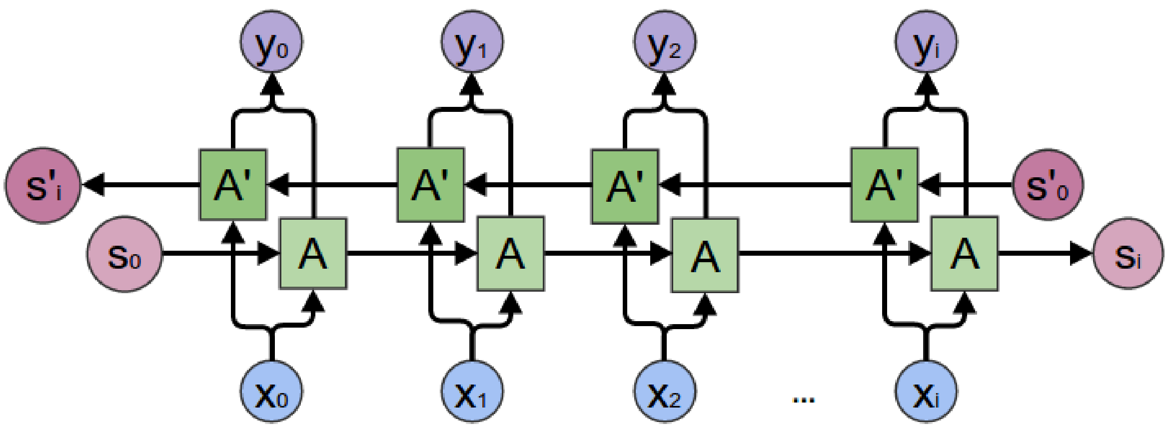
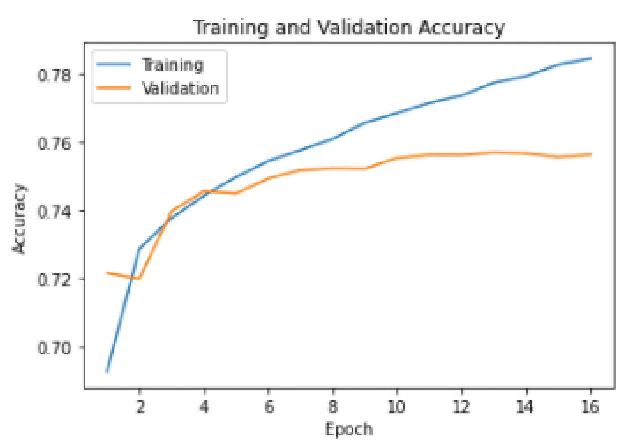


Figure A2: Architecture of Bidirectional LSTM Layer

```
modelB8 = Sequential([
    embedding_layer,
    Bidirectional(LSTM(units=64, return_sequences=True)),
    Dropout(0.3),
    Bidirectional(LSTM(units=64)),
    Dropout(0.3),
    Dense(units=64, activation='relu'),
    Dense(units=1, activation='sigmoid')
])
runModel(modelB8)
```

```
Model: "sequential_25"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
embedding (Embedding)        (None, 50, 50)            25148750
_________________________________________________________________
bidirectional_14 (Bidirectio (None, 50, 128)           58880
_________________________________________________________________
dropout_18 (Dropout)         (None, 50, 128)           0
_________________________________________________________________
bidirectional_15 (Bidirectio (None, 128)               98816
_________________________________________________________________
dropout_19 (Dropout)         (None, 128)               0
_________________________________________________________________
dense_50 (Dense)             (None, 64)                8256
_________________________________________________________________
dense_51 (Dense)             (None, 1)                 65
=================================================================
Total params: 25,314,767
Trainable params: 166,017
Non-trainable params: 25,148,750
_________________________________________________________________
```

```
Training Accuracy: 0.7773
Training Loss: 0.4644
Validation Accuracy: 0.7569
Validation Loss: 0.4985
Testing Accuracy: 0.7566
Testing Loss: 0.4982
Process Time: 3:12:10.192417
Model Accuracy Trend
```



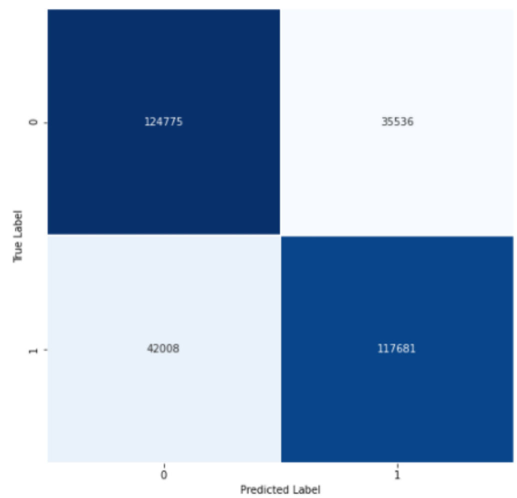
Appendix Figure A3: Architecture/Results of 2-Layer Bidirectional LSTM with Regularization

#	Type	Train Acc.	Train Loss	Val Acc.	Val Loss	Test Acc.	Test Loss	Process Time
A1	1-Layer RNN Unidirectional (no dropout)	0.7533	0.5009	0.7413	0.5189	0.7280	0.5383	07:18
A2	1-Layer RNN Bidirectional (no dropout)	0.7492	0.5070	0.7349	0.5292	0.7303	0.5321	08:19
A3	1-Layer RNN Unidirectional (dropout)	0.7488	0.5105	0.7442	0.5135	0.7430	0.5154	21:06
A4	1-Layer RNN Bidirectional (dropout)	0.7436	0.5163	0.7369	0.5232	0.7313	0.5336	08:56
A5	2-Layer RNN Unidirectional (no dropout)	0.7517	0.5031	0.7402	0.5211	0.7350	0.5275	11:24
A6	2-Layer RNN Bidirectional (no dropout)	0.7633	0.4863	0.7406	0.5199	0.7352	0.5311	25:34
A7	2-Layer RNN Unidirectional (dropout)	0.7357	0.5269	0.7388	0.5222	0.7244	0.5393	15:56
A8	2-Layer RNN Bidirectional (dropout)	0.7444	0.5144	0.7404	0.5184	0.7383	0.5238	35:27
A9	2-Layer RNN Uni/Bi directional (no dropout)	0.7678	0.4780	0.7396	0.5258	0.7358	0.5281	29:04
A10	2-Layer RNN Uni/Bi directional (dropout)	0.7160	0.5524	0.7276	0.5371	0.7265	0.5374	15:45
A11	2-Layer RNN Bi/Uni directional (no dropout)	0.7665	0.4808	0.7394	0.5244	0.7362	0.5280	26:21
A12	2-Layer RNN Bi/Uni directional (dropout)	0.7428	0.5164	0.7370	0.5279	0.7366	0.5242	24:47

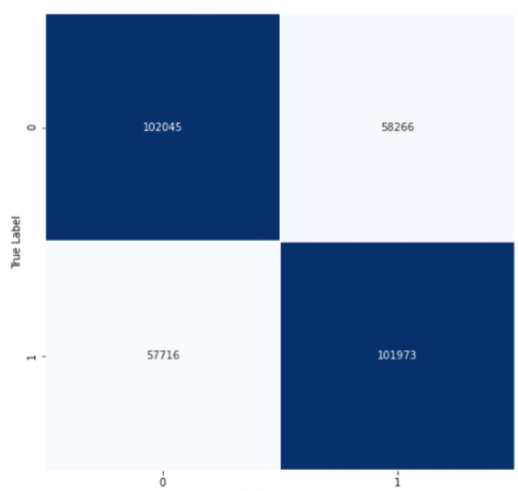
Appendix Figure A4: Benchmark Metrics for a Simple RNN Network

C1	1-Layer GRU Unidirectional (no dropout)	0.7776	0.4631	0.7558	0.4991	0.7534	0.5040	49:26
C2	1-Layer GRU Bidirectional (no dropout)	0.7773	0.4632	0.7560	0.4986	0.7483	0.5134	1:40:21
C3	1-Layer GRU Unidirectional (dropout)	0.7685	0.4787	0.7562	0.4960	0.7550	0.4981	58:34
C4	1-Layer GRU Bidirectional (dropout)	0.7711	0.4750	0.7563	0.4974	0.7522	0.5029	2:10:15

Appendix Figure A5: Benchmark Metrics for a GRU Network



Appendix Figure A6: Confusion Matrix for Winning Model



Appendix Figure A7: Confusion Matrix for SentimentIntensityAnalyzer Package

Appendix B

```
Negative    2877765
Positive    2330268
Neutral     1331121
Name: RNN_Score_Type, dtype: int64
```

Figure B1: Value Counts of Score Type for Boeing Tweets

Week	Tweets	Score
2019-03-11	196477	3.085535
2013-01-14	76542	3.776783
2019-03-18	68064	3.516612
2020-01-06	62414	3.048719
2016-12-05	52429	4.738426
2019-04-01	42223	3.127595
2019-03-25	41679	3.417796
2019-12-16	40212	3.864664
2013-01-07	38358	3.939324
2020-03-23	33668	4.005953

Figure B2: Weeks with Most Tweets and Average Sentiment Score

Week	Tweets	Score
2017-01-23	14295	6.213763
2015-08-31	13631	6.203713
2015-11-16	6979	5.997119
2016-07-18	11306	5.969875
2015-03-23	18935	5.969452

Figure B3: Weeks with Highest Sentiment Score and Number of Tweets

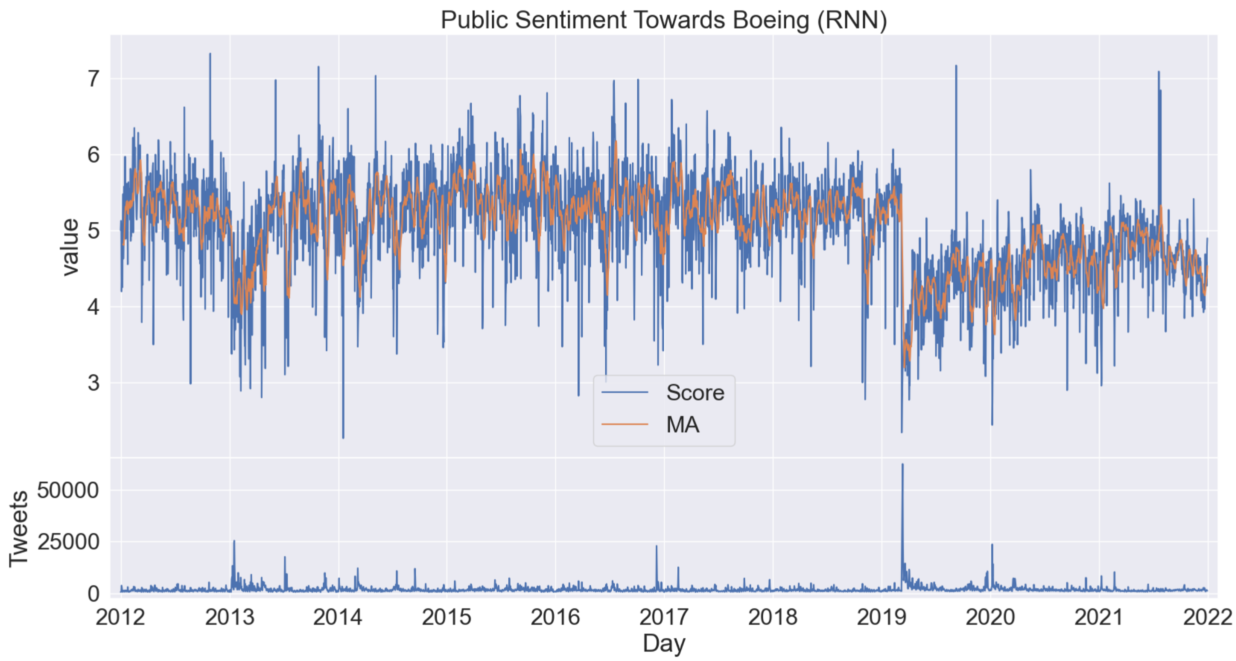


Figure B4: Daily Aggregation of Public Sentiment with Volume Plot and 10-Day Moving Average for the RNN Algorithm

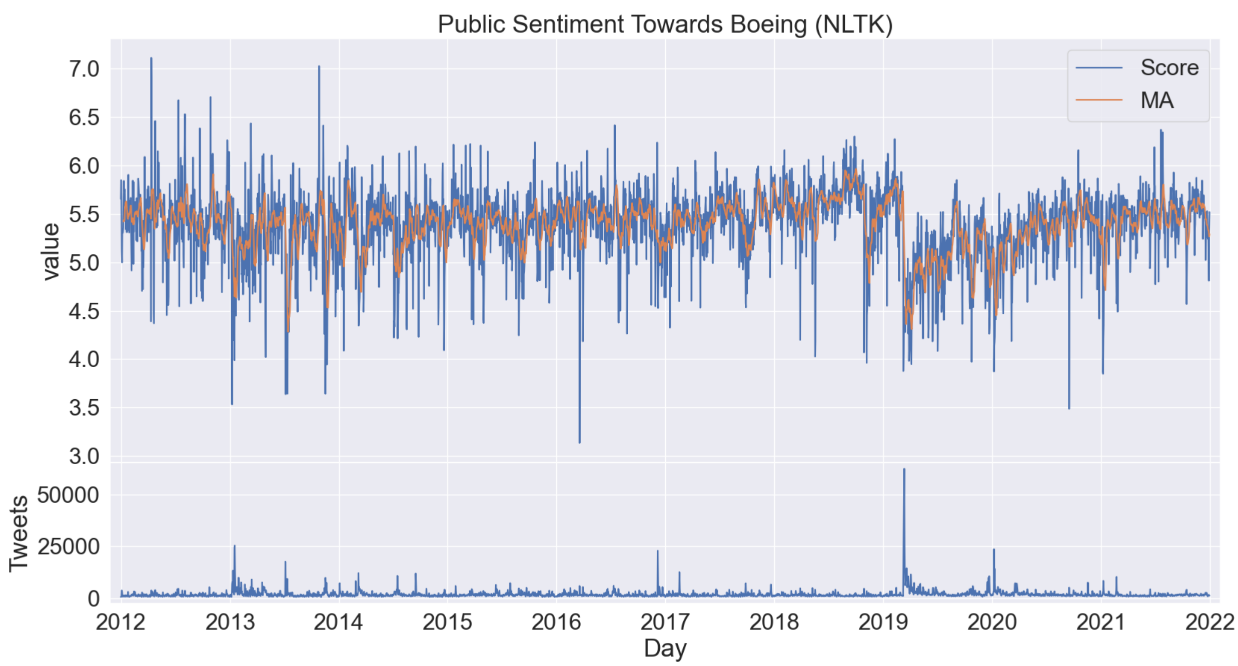


Figure B5: Daily Aggregation of Public Sentiment with Volume Plot and 10-Day Moving Average for the NLTK Algorithm

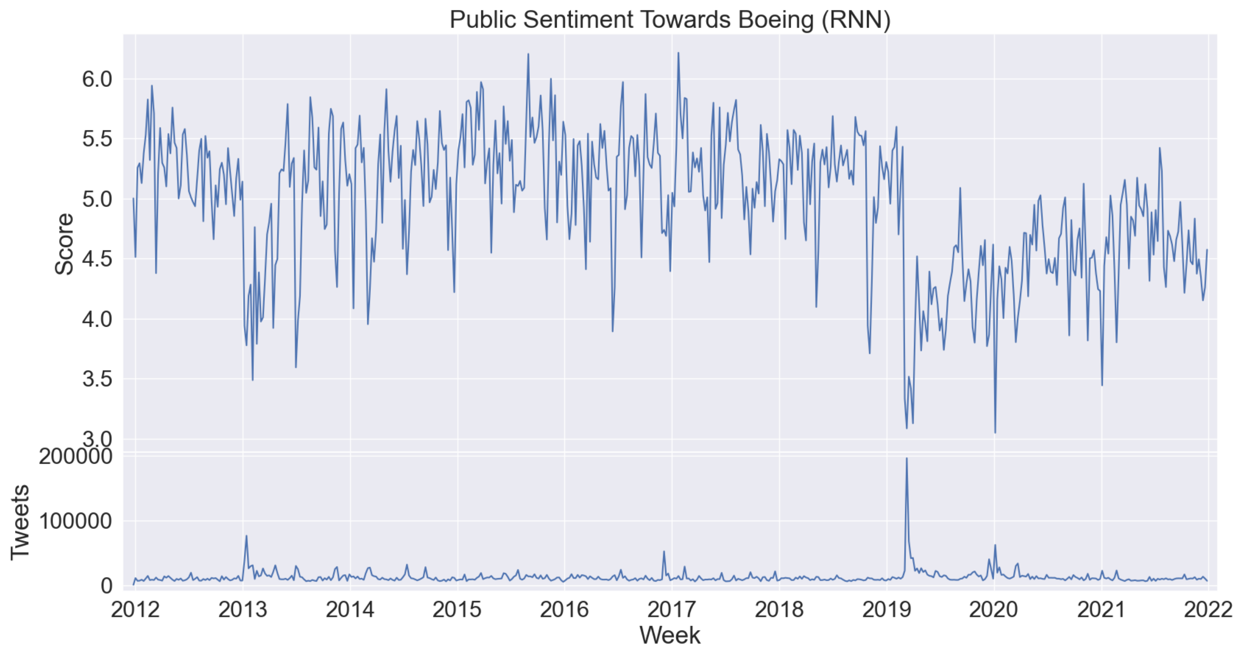


Figure B6: Weekly Aggregation of RNN Public Sentiment with Volume Plot

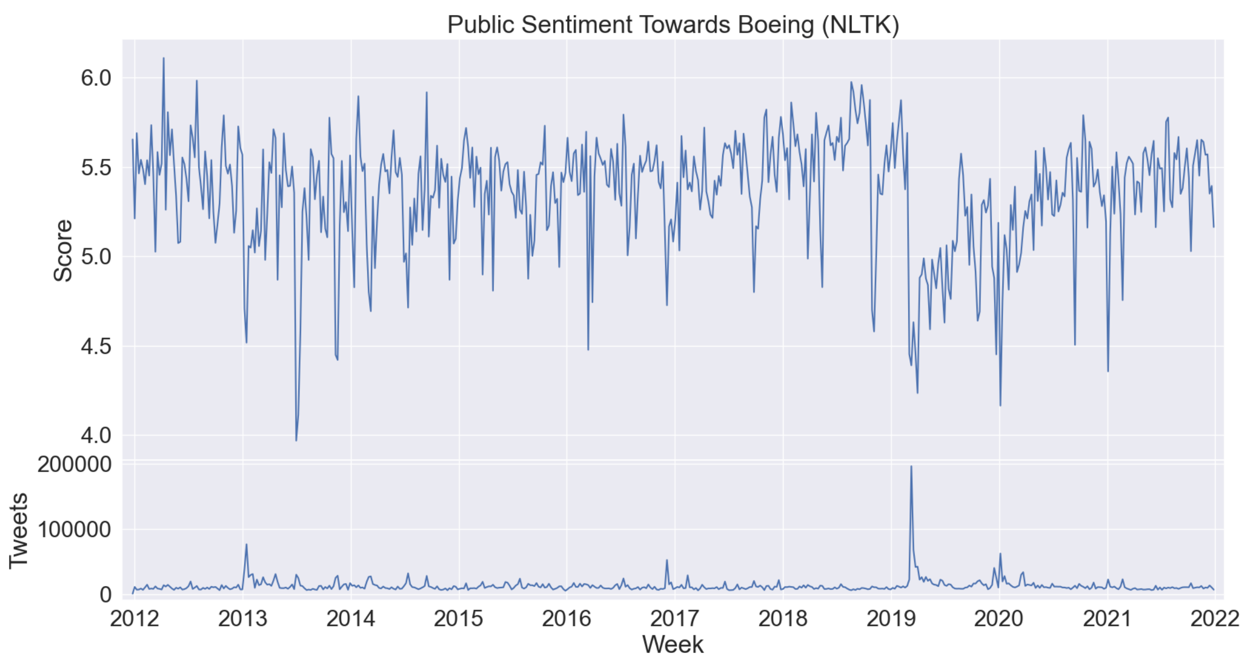


Figure B7: Weekly Aggregation of NLTK Public Sentiment with Volume Plot

Appendix C: Airbus Tweet Figures

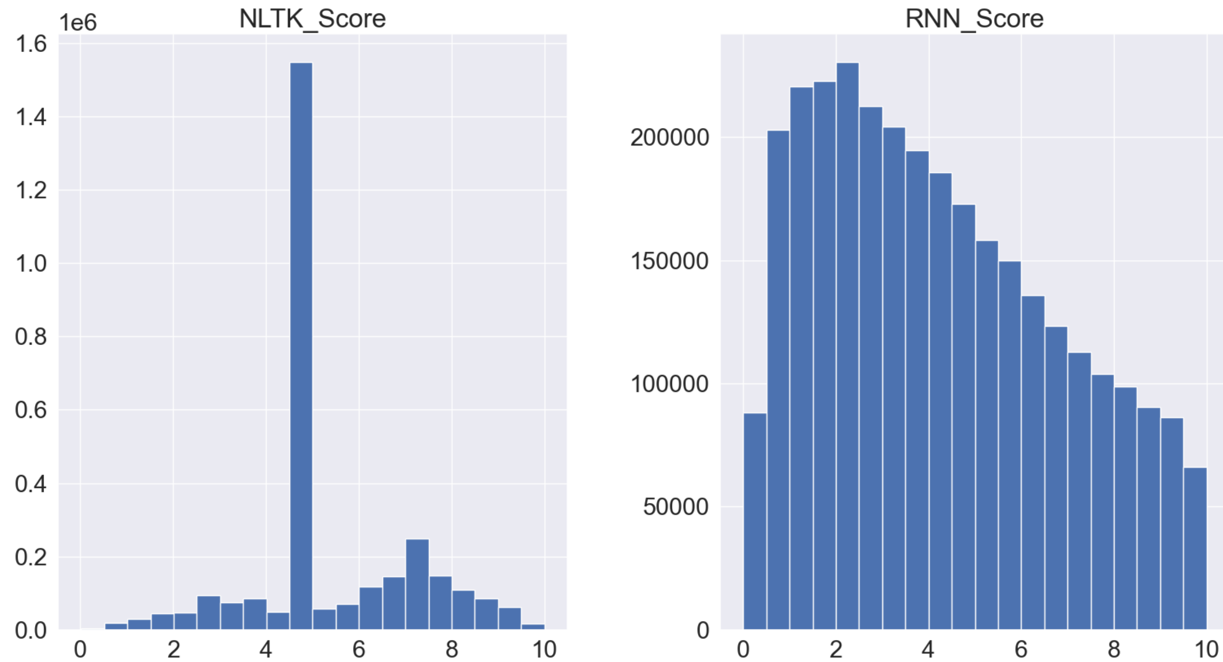


Figure C1: Histogram of Score Distribution for Airbus Tweets

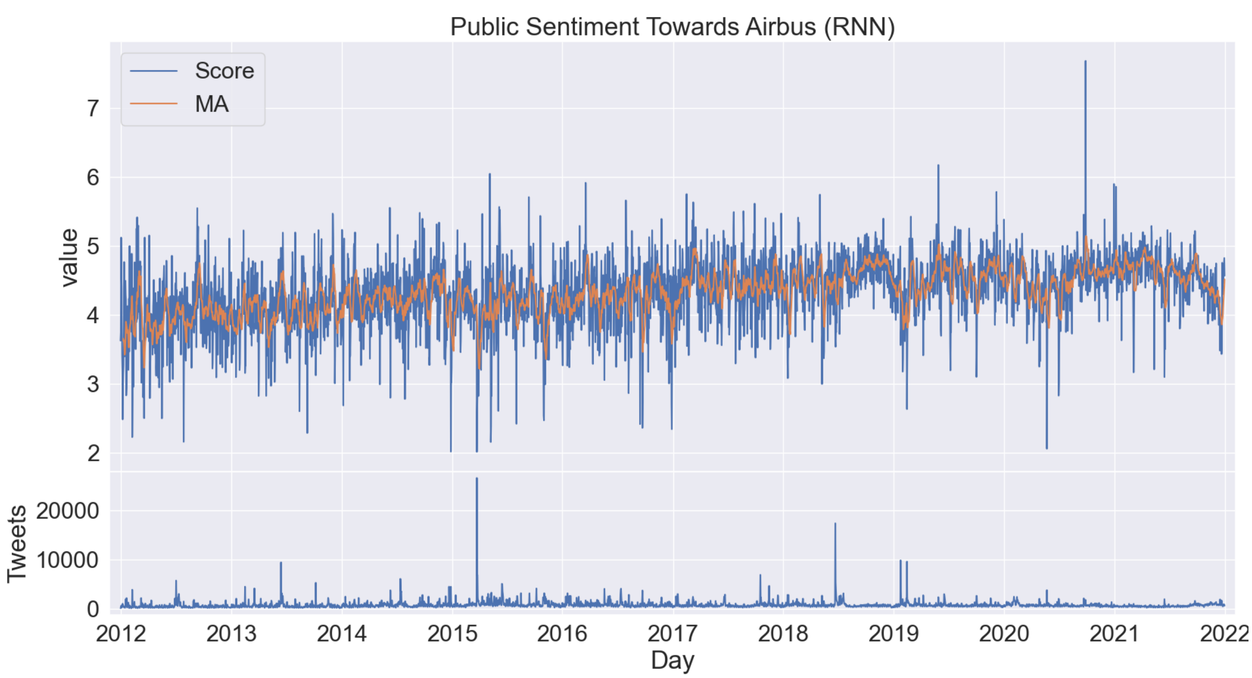


Figure C2: Daily Aggregation of Public Sentiment with Volume Plot and 10-Day Moving Average for the RNN Algorithm for Airbus

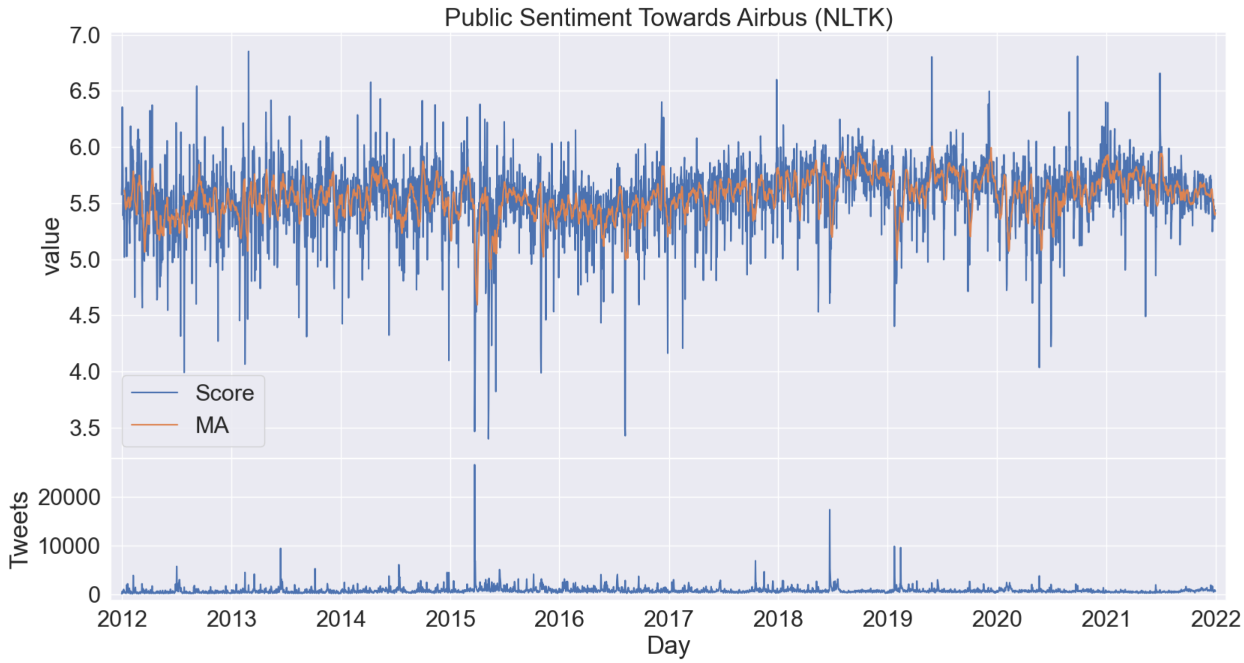


Figure C3: Daily Aggregation of Public Sentiment with Volume Plot and 10-Day Moving Average for the NLTK Algorithm for Airbus

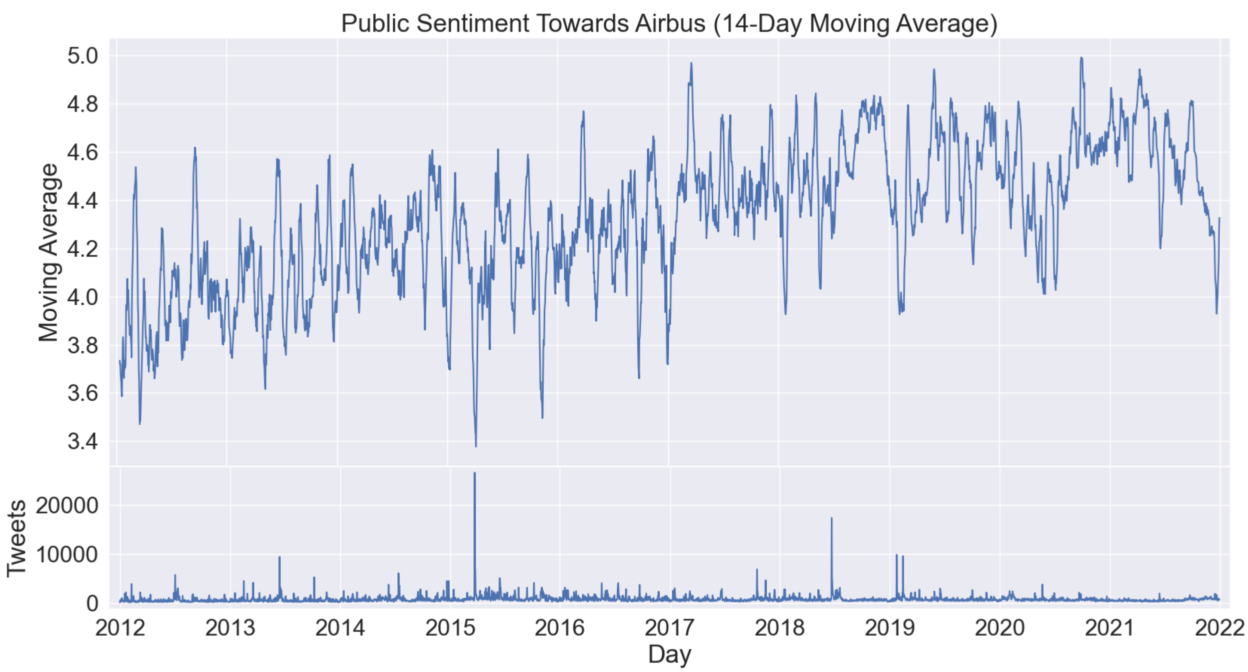


Figure C4: 14-Day Moving Average of RNN Public Sentiment with Volume Plot towards Airbus

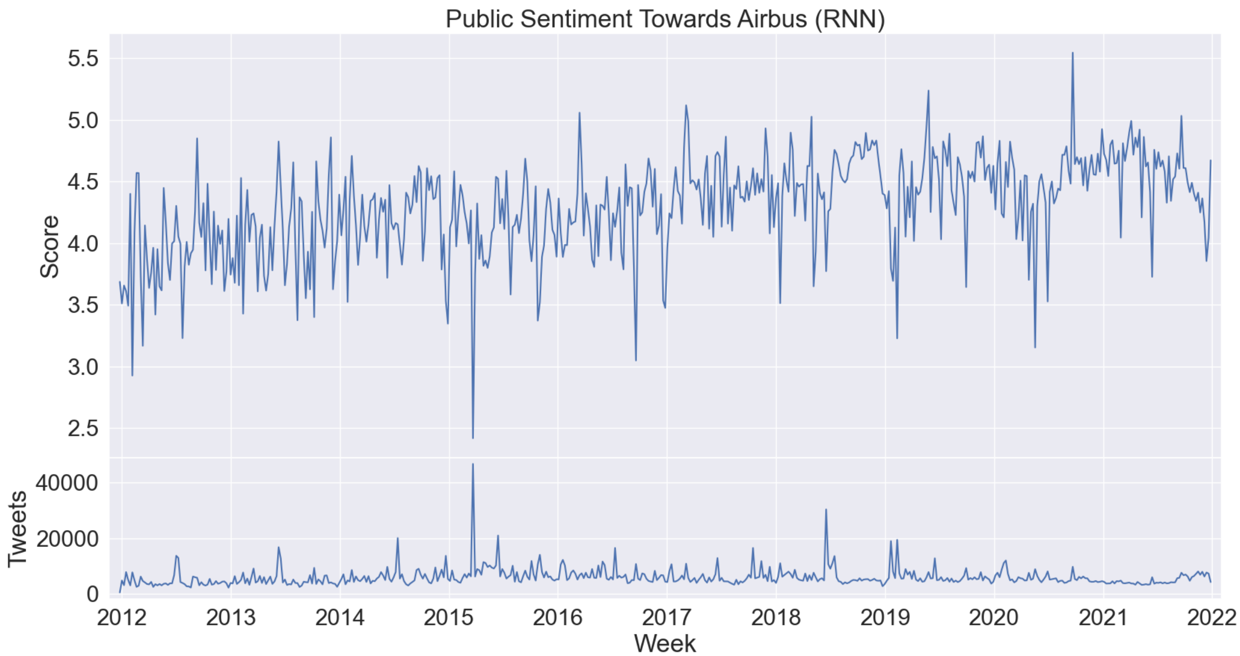


Figure C5: Weekly Aggregation of RNN Public Sentiment with Volume Plot

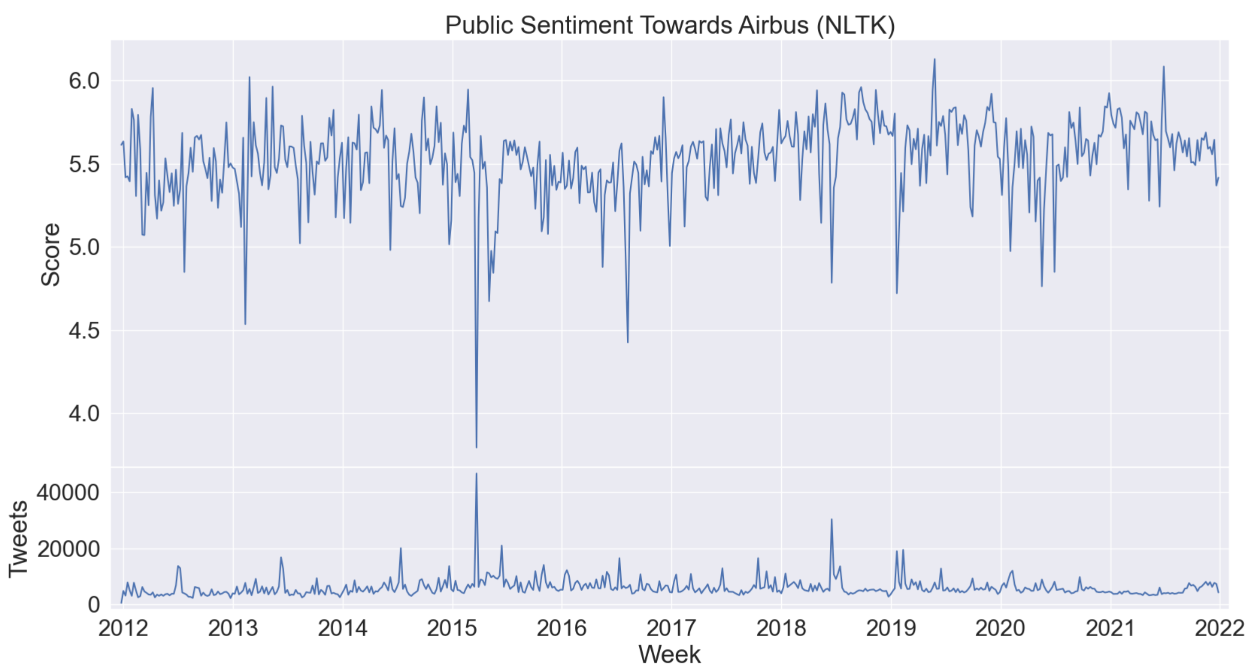
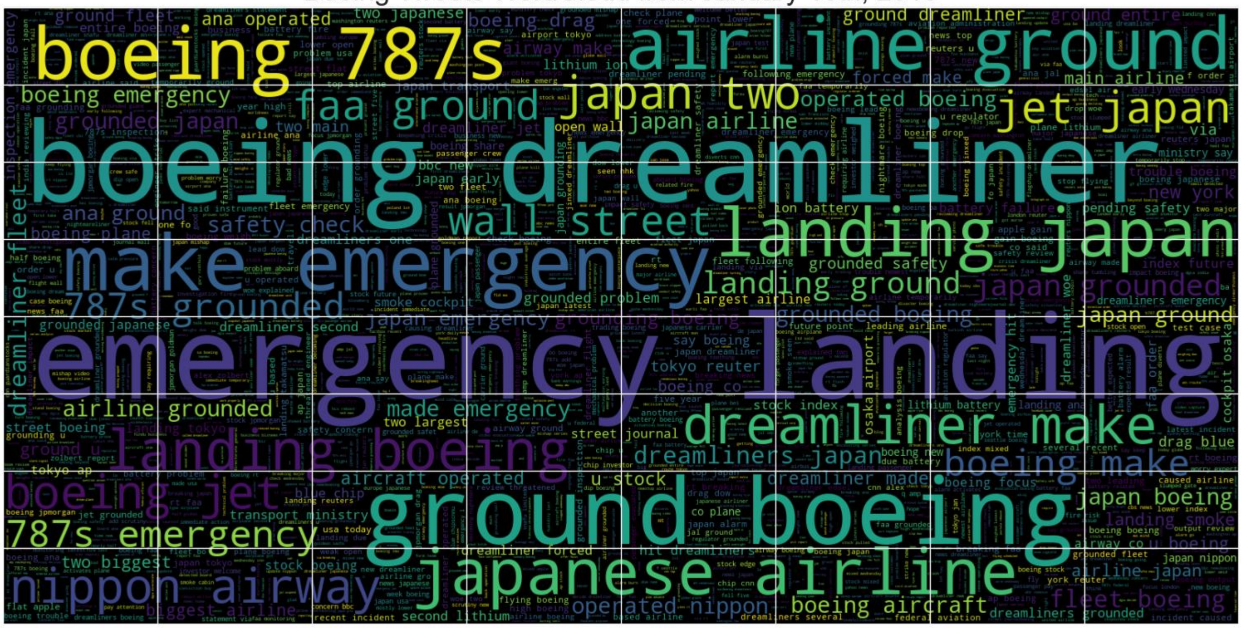


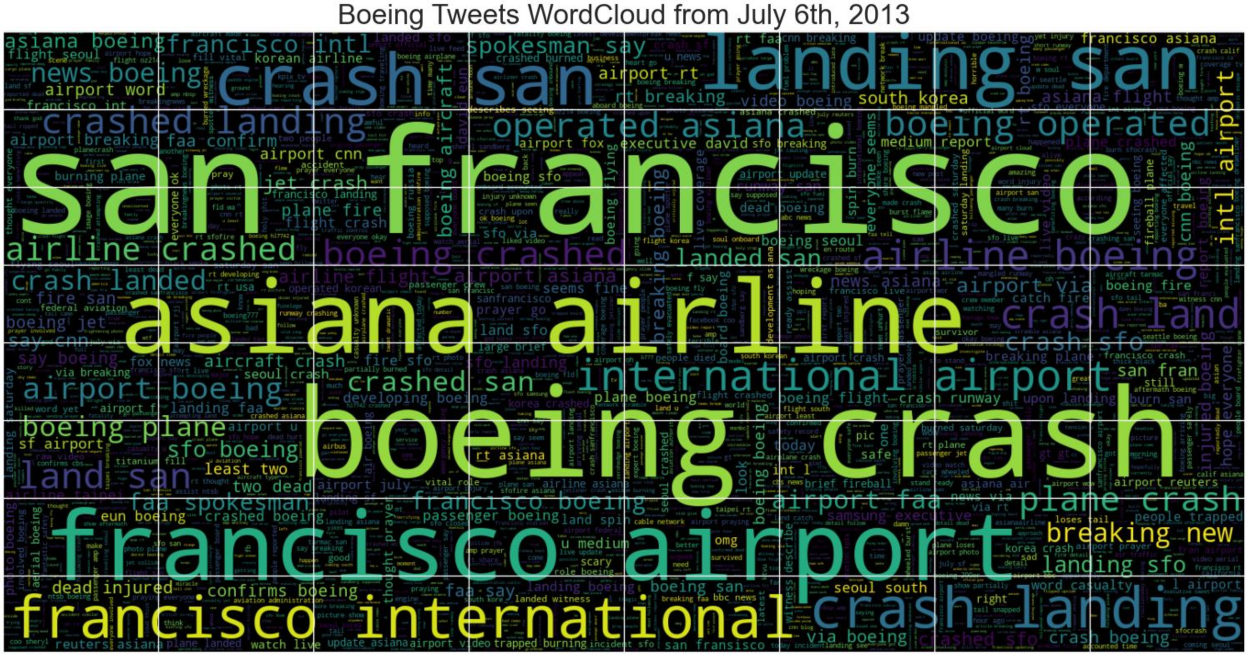
Figure C6: Weekly Aggregation of NLTK Public Sentiment with Volume Plot

Appendix D

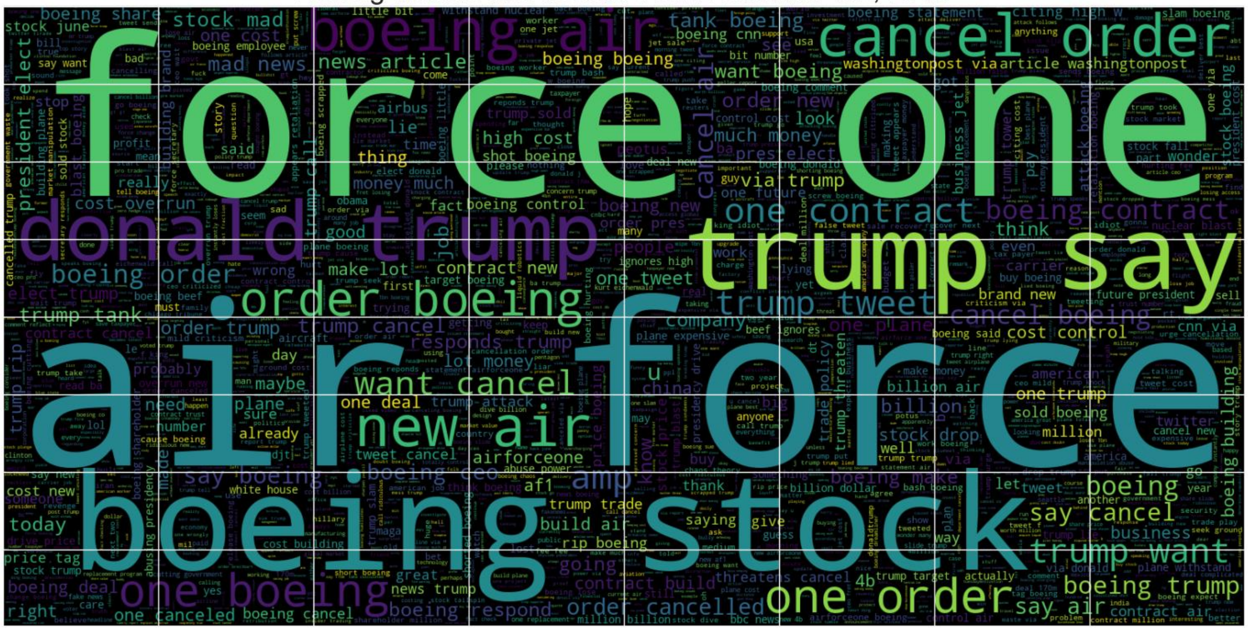
Boeing Tweets WordCloud from January 16th, 2013



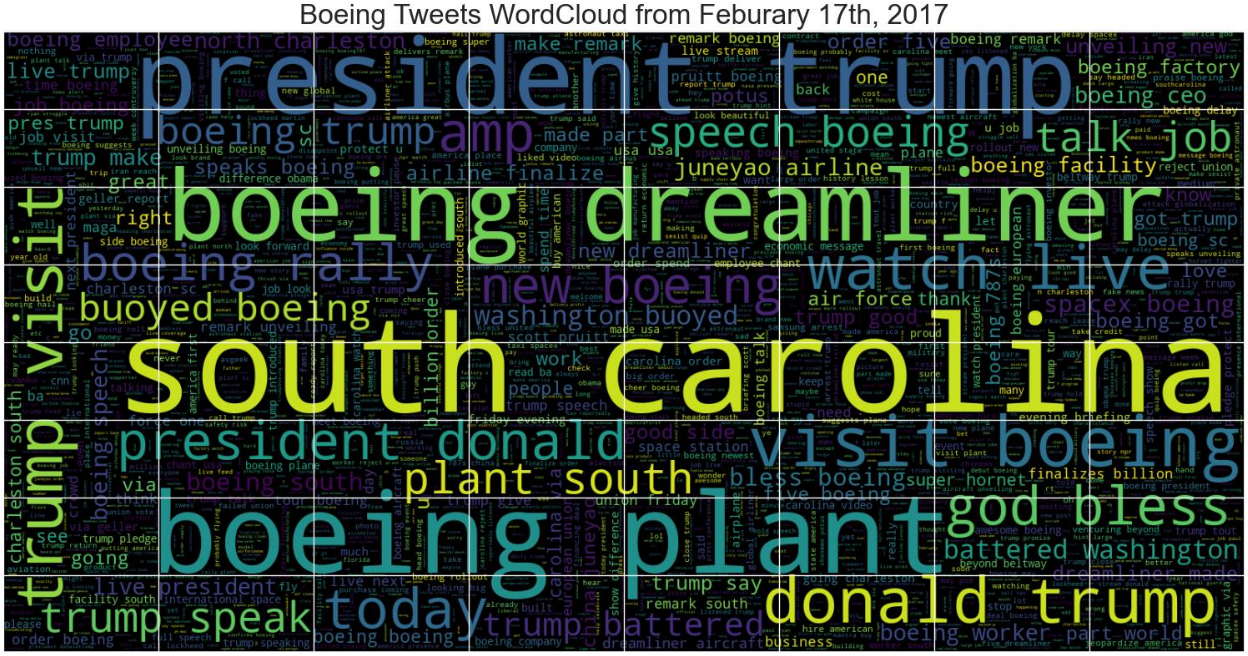
Boeing Tweets WordCloud from July 6th, 2013



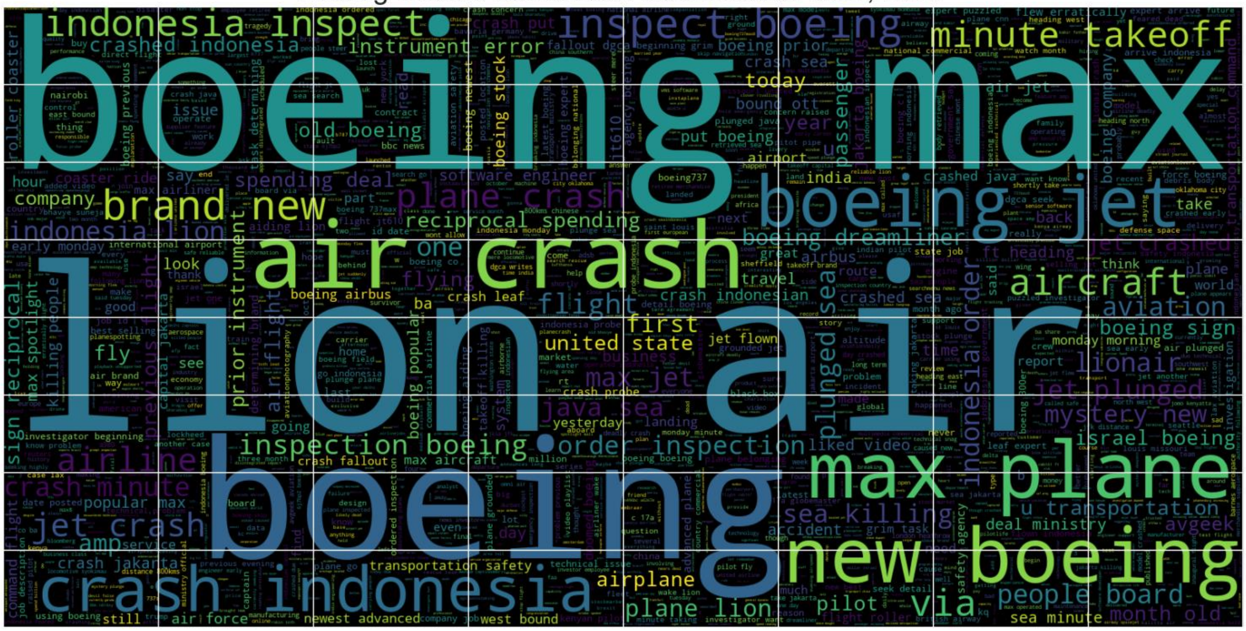
Boeing Tweets WordCloud from December 6th, 2016



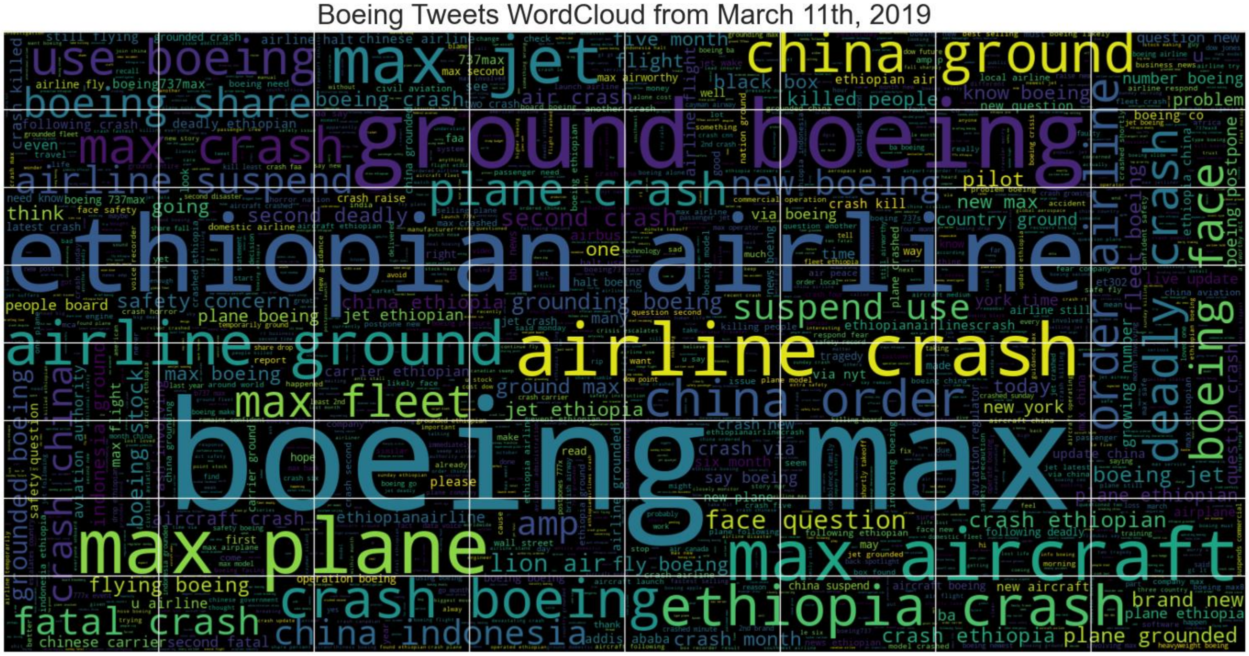
Boeing Tweets WordCloud from Feburary 17th, 2017



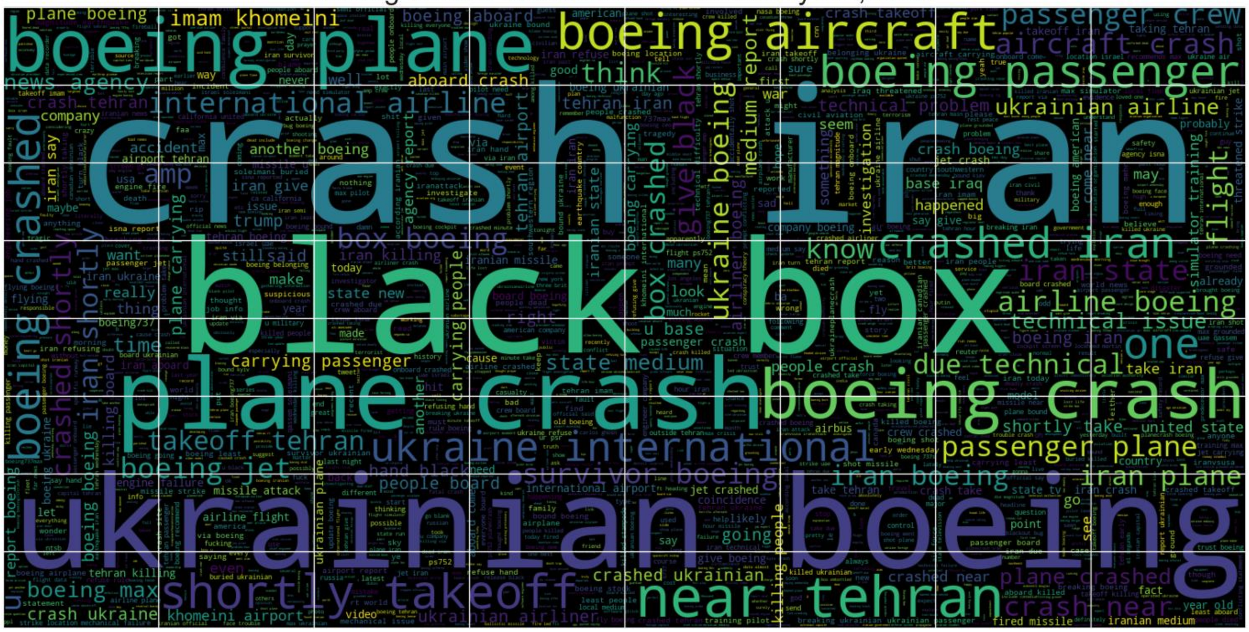
Boeing Tweets WordCloud from October 30th, 2018



Boeing Tweets WordCloud from March 11th, 2019

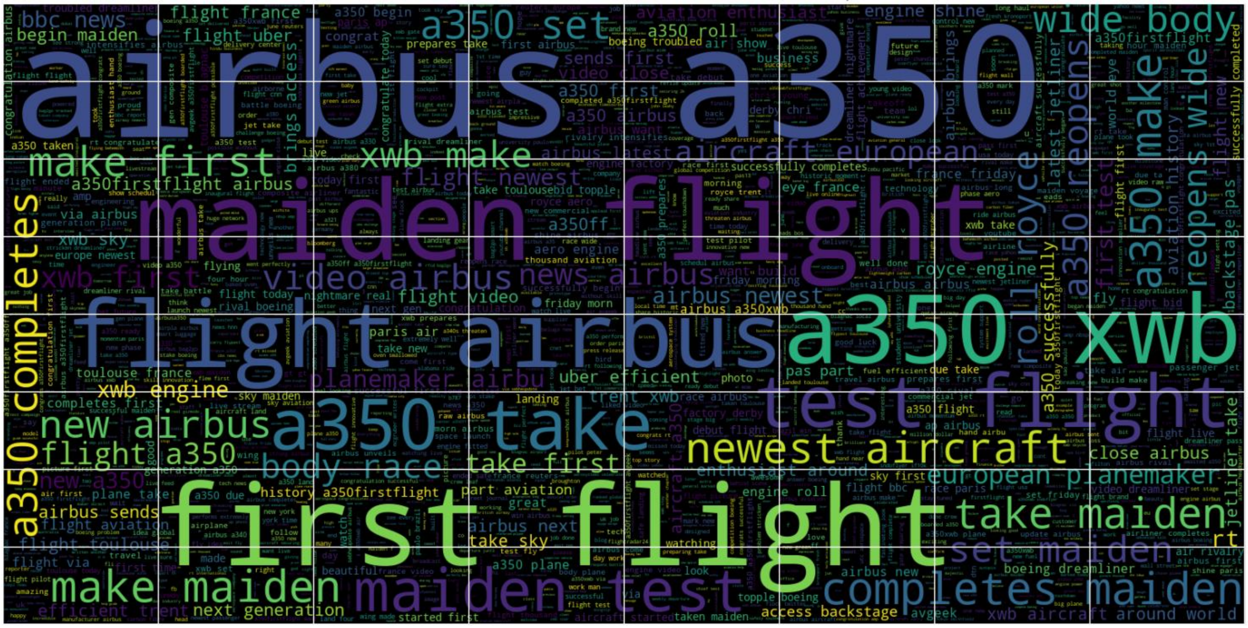


Boeing Tweets WordCloud from January 8th, 2020

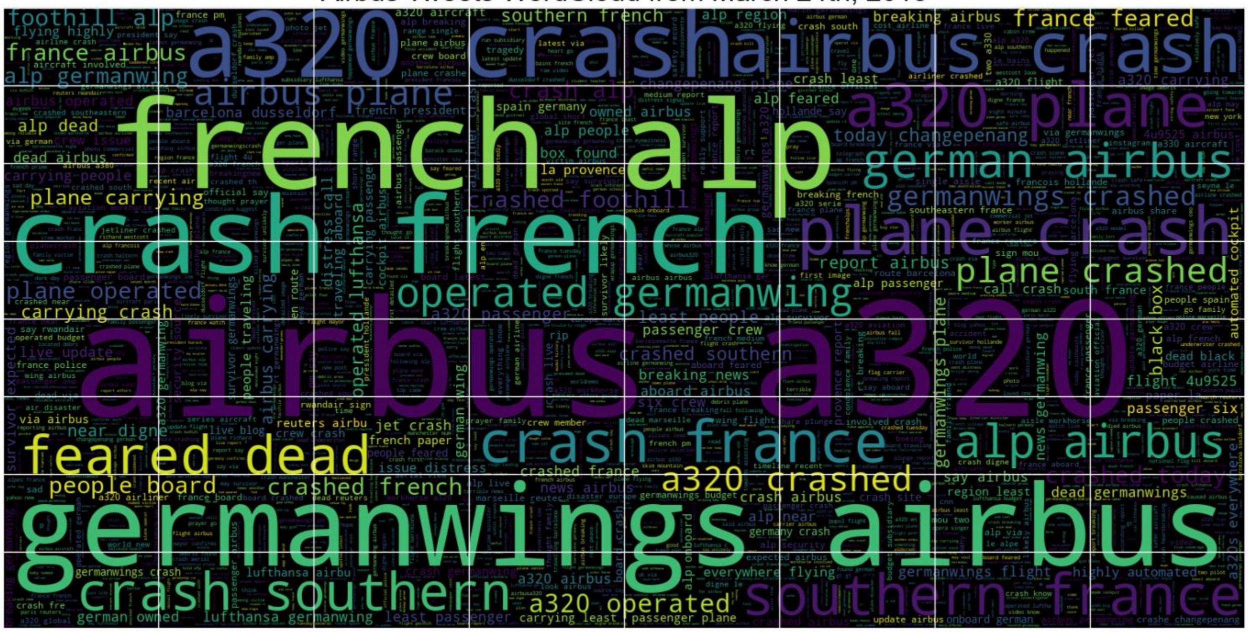


Appendix E

Airbus Tweets WordCloud from June 14th, 2013

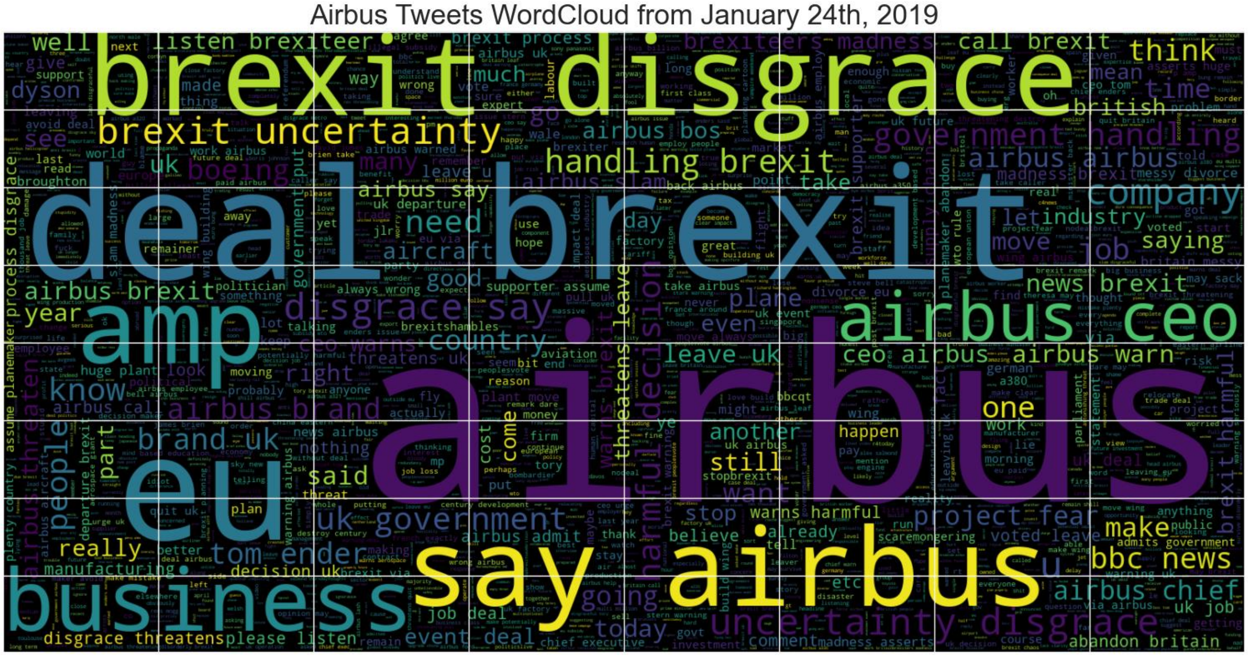


Airbus Tweets WordCloud from March 24th, 2015



Airbus Tweets WordCloud from June 22nd, 2018

Airbus Tweets WordCloud from January 24th, 2019



Airbus Tweets WordCloud from February 14th, 2019