Airplane Sentiment Insights

By Surabhi Kulkarni

Key Numbers (1 of 2)

Count of Airline Sentiment Tweets by Airline

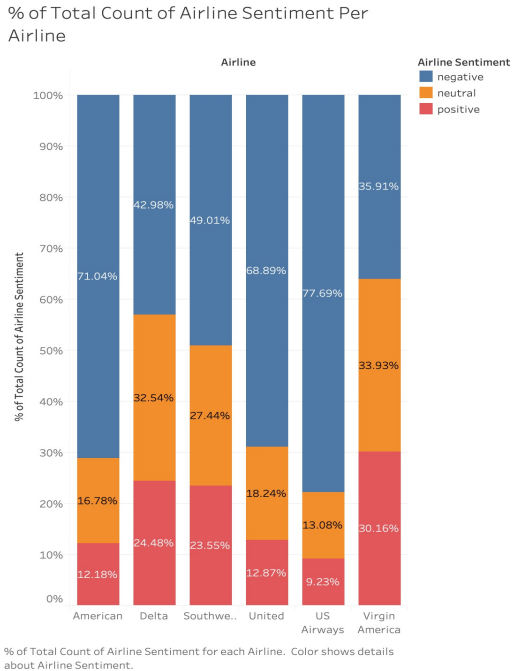
	Airline						
Airline Sent..	American	Delta	Southwest	United	US Airways	Virgin America	Grand Total
negative	1,960	955	1,186	2,633	2,263	181	9,178
neutral	463	723	664	697	381	171	3,099
positive	336	544	570	492	269	152	2,363
Grand Total	2,759	2,222	2,420	3,822	2,913	504	14,640

Count of Airline Sentiment broken down by Airline vs. Airline Sentiment.

The purpose of the data was to scrape tweet sentiment about airlines. The data set has a total of 14,640 instances.

It's clear that the highest prevalence of tweets is of negative sentiment. The airline that was scraped the most was United Airlines and Virgin America has the least instances of tweets overall.

Key Numbers (2 of 2)



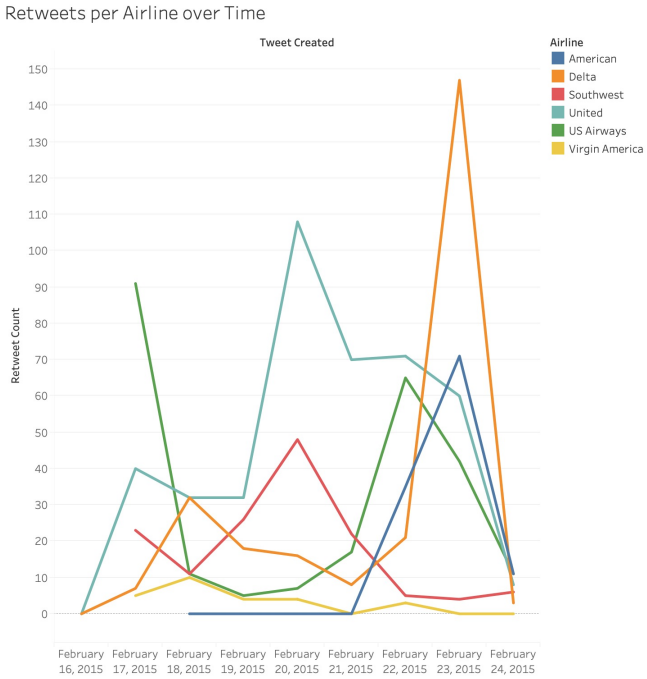
The chart shows the percentage of each type of tweet by sentiment per each individual airline.

Virgin America performed the best by far with the most balanced representation of each type of tweet.

However, it is important to note that overall, there were less tweets recorded for VA, and this statistic could be indicative of a small sample size.

US Airways performed the worst with over 77% of its tweets having negative sentiment.

Relationship between Retweets and Airlines Over Time



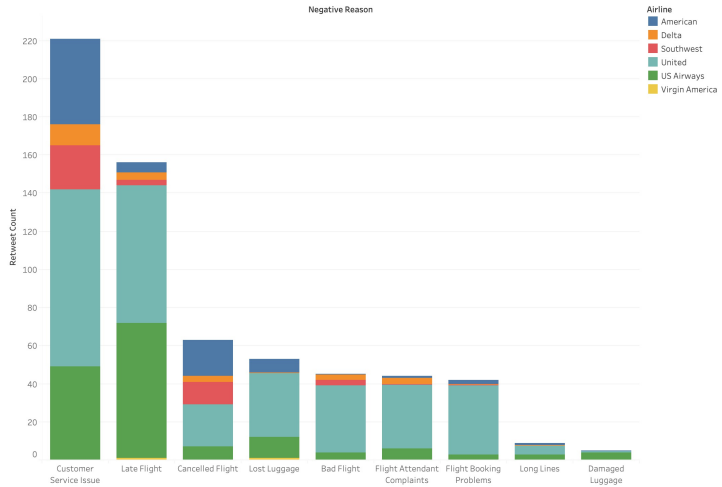
This chart shows the trends of retweets per airlines over time.

Overall, the retweets show no real trend, and are quite unpredictable.

However, the large spikes such as the ones on Feb 20 for United and Feb 23 on Delta suggest that there was a particular event for the respective airline that triggered the retweets.

What was Retweeted the Most?

Negative Reason/RT

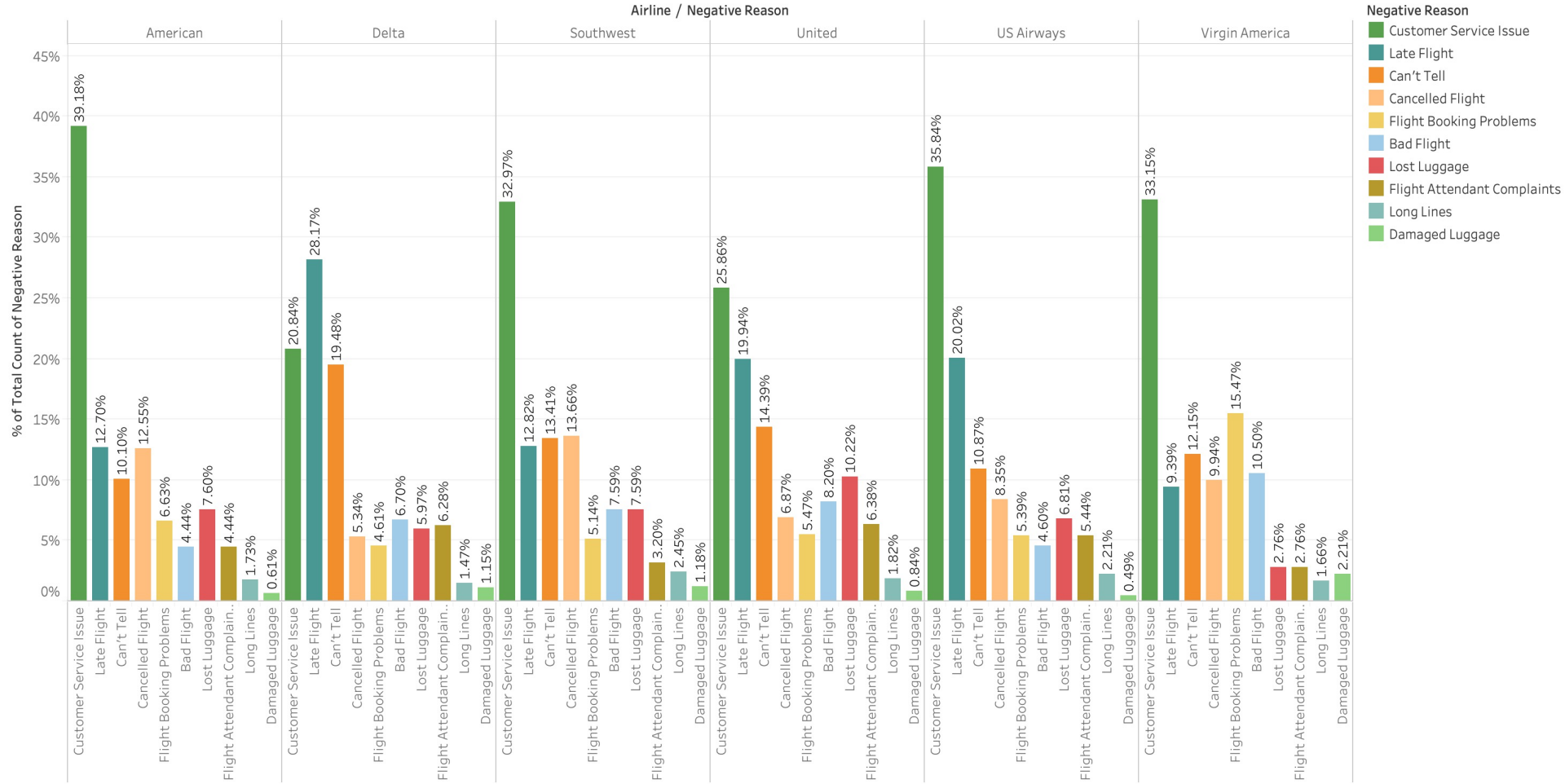


Sum of Retweet Count for each Negative Reason. Color shows details about Airline. The view is filtered on Negative Reason, which excludes Null and Can't Tell.

Customers retweeted and affirmed their hatred for airline companies' customer service, with United Airlines having the largest count of retweets for customer service issues

Following this reason, United and US Airways are the top 2 airlines leading for the second largest category of retweets: Late Flight.

Percentage Counts of Negative Reason Per Airline



% of Total Count of Negative Reason for each Negative Reason broken down by Airline. Color shows details about Negative Reason. The view is filtered on Negative Reason, which excludes Null.

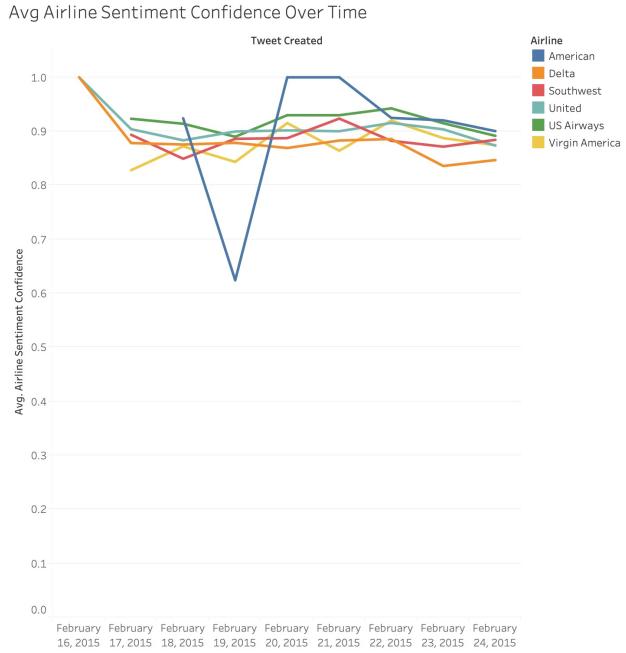
What are the Airlines Doing Wrong?

- When controlled for the number of negative tweets per airline, the chart reveals the percentage per airline that is attributed to each negative reason.

- As seen in all of the charts, the biggest reason for complaint is customer service, except for with Delta who had more issues with a late flight.

- The second most prevalent reason for negative tweets was not as unified per airline and seemed to vary from cancelled flight, late flight, and booking issues.

How Reliable Was The Sentiment Assessment Per Day?

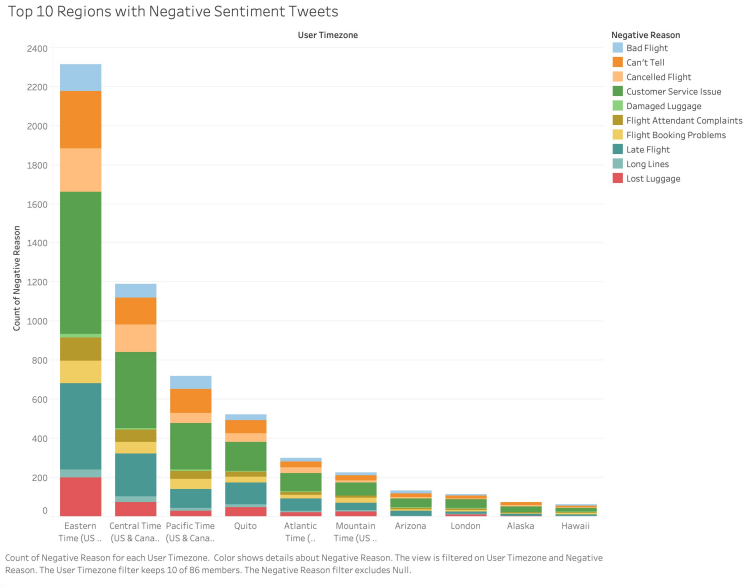


This chart shows the average airline sentiment confidence over the 12 day time period.

For the most part, each airline's assessment was fairly trustworthy as the sentiment confidence was somewhat stable.

However, American Airlines experienced a significant drop in confidence on February 19, 2020. The reason for this was not found in the data set, but maybe this is related to some sort of unique event that happened that the code couldn't confidently classify.

Are the Complaints Concentrated in a Certain Time Zone?



This chart shows the top ten regions with the most negative sentiment tweets. Rather than using location, the time zone was used since the accuracy was greater than

The top region was the Eastern Time Zone so it can be reasonably concluded that people from this time zone are the least happy with flight service.

Unsurprisingly, their biggest complaint was a customer service issue.

Where Should We Go From Here?

- Airline companies should work on customer service and flight timing, especially on the Eastern Time Zone Areas.
- Collect more detailed data not only on the tweets, but also of the customer service process to determine specific areas of improvement.
- Find reasons why the airline company is receiving positive feedback to prioritize their competitive advantage.

Appendix

Data Assessment

- The data set taken at face value is granular and detailed. A README explaining the variables would have greatly improved the clarity
- However, there are several variables that can be used to gather important insights such as sentiment and the negative reason.
- Positive reasons also needed to be added. In addition, the user location needed to be manually cleaned since there were many different captions for the locations.

Analysis Action Steps

- Cleaned data with Tableau data cleaner
- Graphed the general insights by filtering out the data that was unnecessary rather than taking out entire columns
- Analyzed each graph to see if it produced an insight and synthesized the insights.

Where Should We Go From Here? (Unabridged)

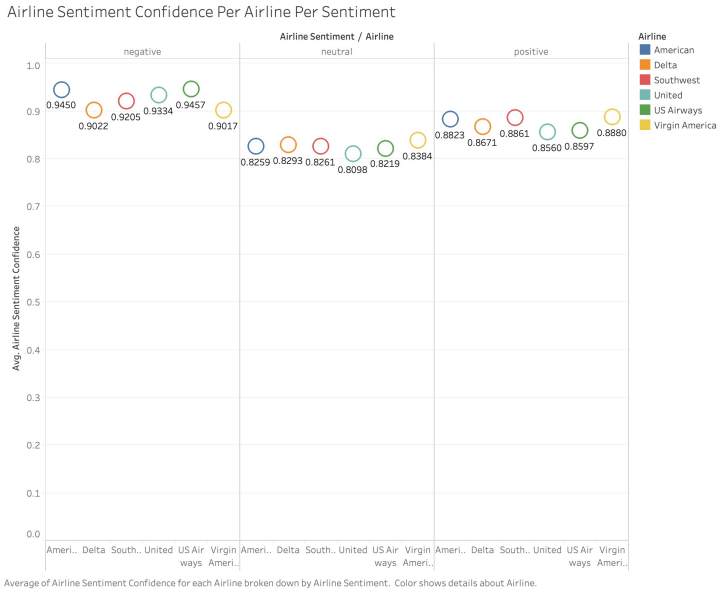
The most clear issue defined in this data set is customer service. Each airline must work on their customer service to provide a better experience. In general, people who are traveling to and from the east coast tend to report more complaints than in other time zones, so it can be reasonably concluded that the customer service experience there is unsatisfactory. There are no discernible reasons for this phenomenon in the data set, but this insight could be borne out of a cultural difference with other regions.

In addition, in order to determine what exactly is the area of improvement within customer service, more detailed data needs to be collected not only on the tweets, but also during the process.

Also, the data set does not provide reasons for why the tweet sentiment is positive. This is crucial information for airline companies so that they can continue to bolster the positive aspects of their service. In addition, more data needs to be collected to contextualize some of the phenomena with the sentiment confidence over time and the existence of maximum complaints from the eastern time zone region.

Extra Charts

How Trustworthy is the Data?



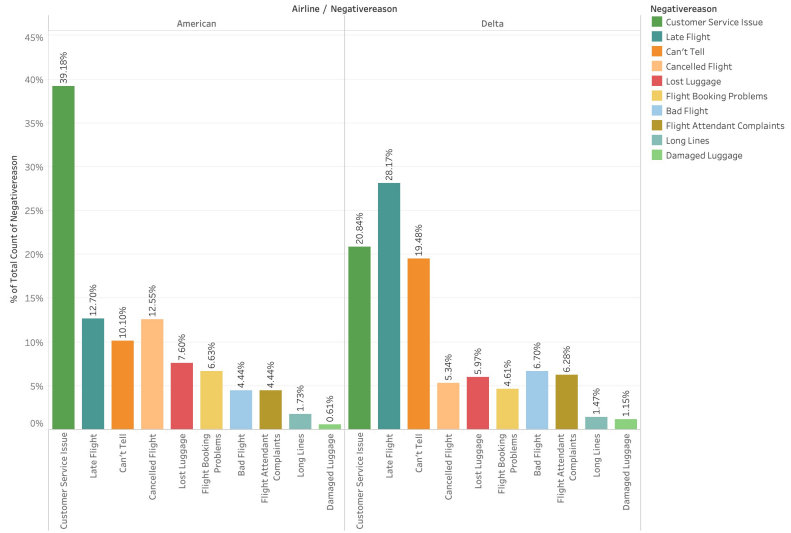
Overall, the data is fairly trustworthy with the average confidence of the airline sentiment being above 0.80.

The most trustworthy data is the data flagged as negative sentiment since all of the average sentiment is above 0.90.

US Airways had the highest average confidence for negative tweets whereas Virgin America had the highest for neutral and positive tweets.

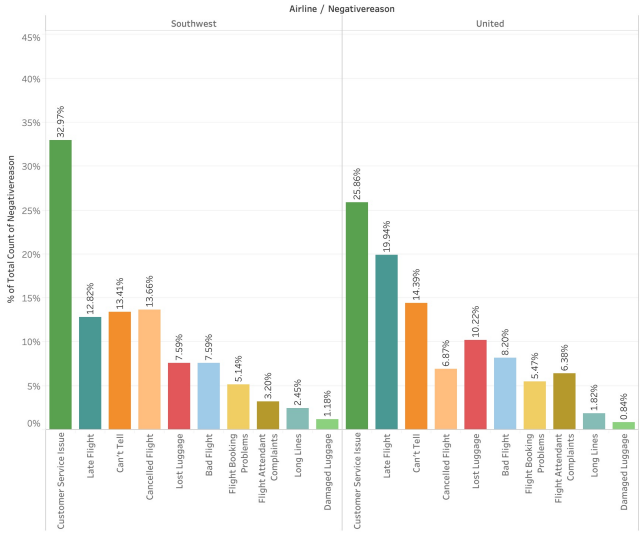
What are the Airlines Doing Wrong? (1 of 2)

Percentage Counts of Negative Reason Per Airline: American and Delta



% of Total Count of Negativereason for each Negativereason broken down by Airline. Color shows details about Negativereason. The view is filtered on Negativereason and Airline. The Negativereason filter excludes Null. The Airline filter keeps American and Delta.

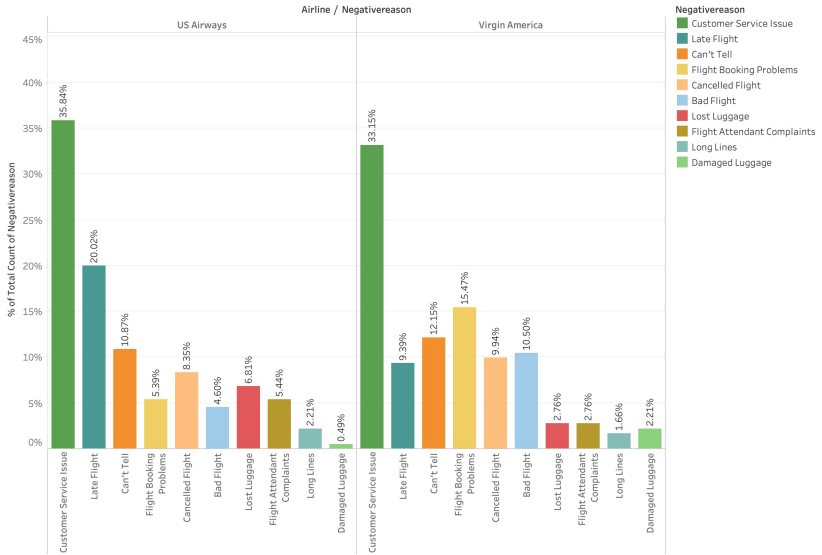
Percentage Counts of Negative Reason Per Airline: Southwest and United



% of Total Count of Negativereason for each Negativereason broken down by Airline. Color shows details about Negativereason. The view is filtered on Negativereason and Airline. The Negativereason filter excludes Null. The Airline filter keeps Southwest and United.

What are the Airlines Doing Wrong? (2 of 2)

Percentage Counts of Negative Reason Per Airline: US Airways and Virgin America



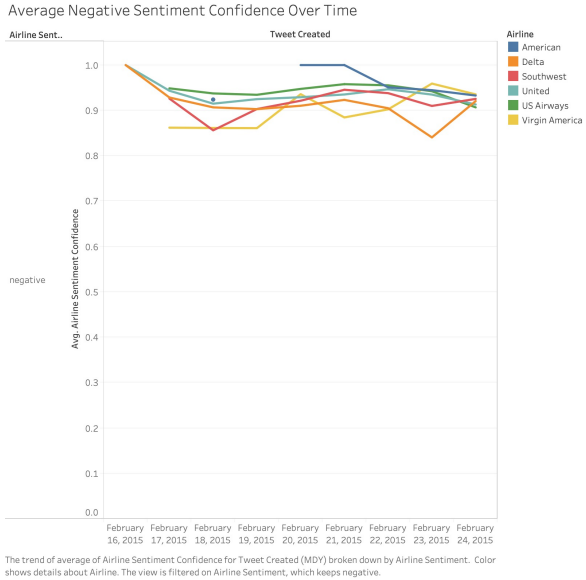
% of Total Count of Negativereason for each Negativereason broken down by Airline. Color shows details about Negativereason. The view is filtered on Airline and Negativereason. The Airline filter keeps US Airways and Virgin America. The Negativereason filter excludes Null.

When controlled for the number of negative tweets per airline, the chart reveals the percentage per airline that is attributed to each negative reason.

As seen in all of the charts, the biggest reason for complaint is customer service, except for with Delta who had more issues with a late flight.

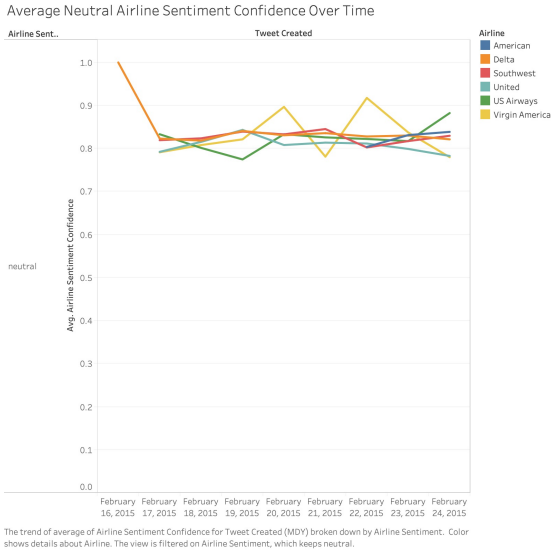
The second most prevalent reason for negative tweets was not as unified per airline and seemed to vary from cancelled flight, late flight, and booking issues.

A Closer Look at Tweet confidence over time



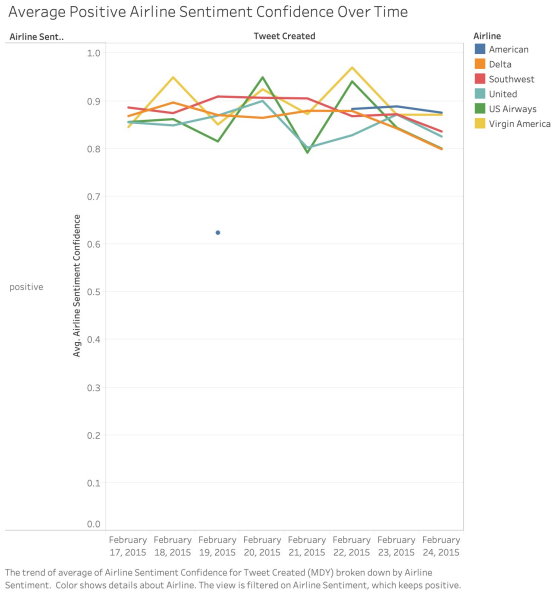
Negative tweet confidence seems to be quite stable and is probably not skewing the average sentiment confidence.

A Closer Look at Tweet confidence over time



Neutral tweet confidence is stable and constant for the most part

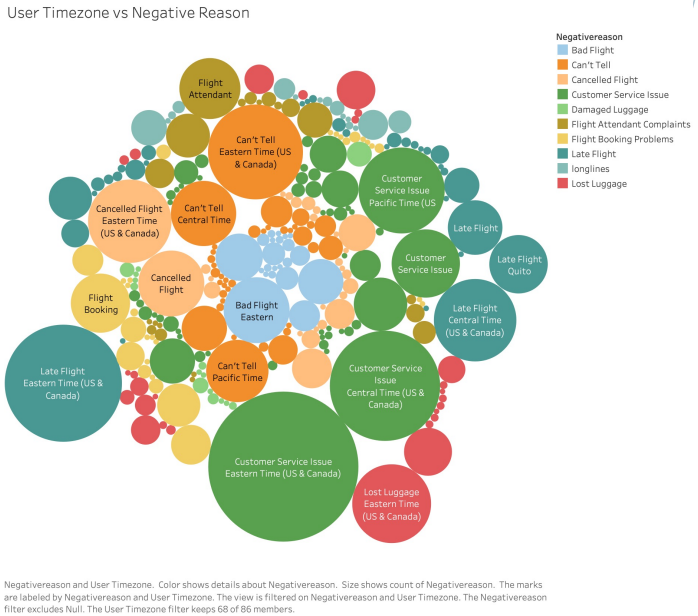
A Closer Look at Tweet confidence over time



The chart indicates that the average sentiment confidence for the American Airlines tweet dropped to just above 0.6.

Reasons are unclear based solely on the data set, but it would be interesting to scrape through news data for the airlines for February 19.

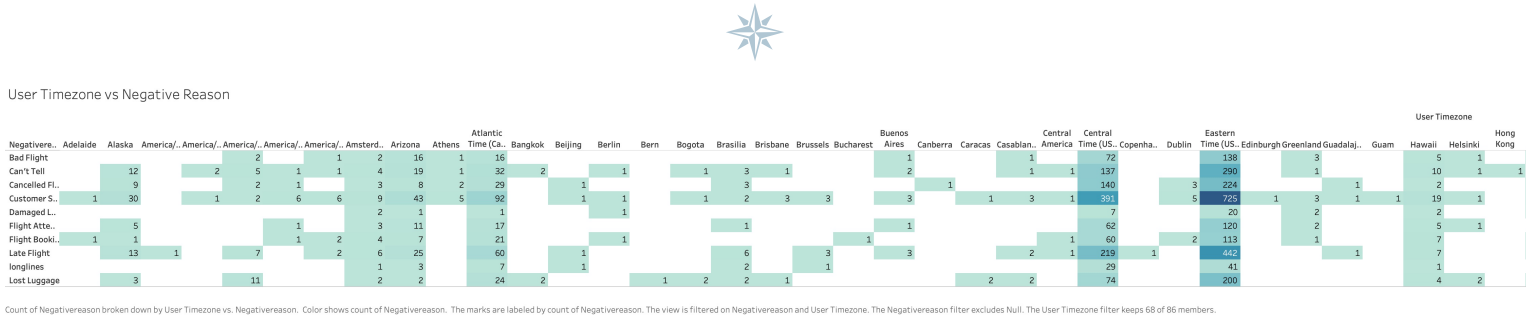
A Closer Look at Tweet confidence over time



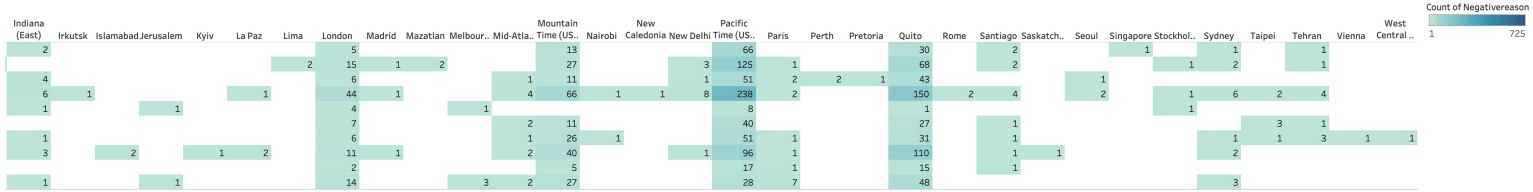
Negativereason and User Timezone. Color shows details about Negativereason. Size shows count of Negativereason. The marks are labeled by Negativereason and User Timezone. The view is filtered on Negativereason and User Timezone. The Negativereason filter excludes Null. The User Timezone filter keeps 68 of 86 members.

A different representation of the chart in slide 9. It was unused because it was hard to compare the different time zones side by side and featured irrelevant information.

User Time Zone vs Negative Reason



Count of Negativereason broken down by User Timezone vs. Negativereason. Color shows count of Negativereason. The marks are labeled by count of Negativereason. The view is filtered on Negativereason and User Timezone. The Negativereason filter excludes Null. The User Timezone filter keeps 68 of 86 members.



Thank You For Reading!