Twitter US Airline Sentiment Analysis

By: Sanjeev Narasimhan

Dataset Assessment/ Context

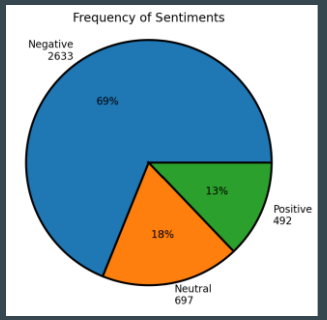
Overview of Dataset

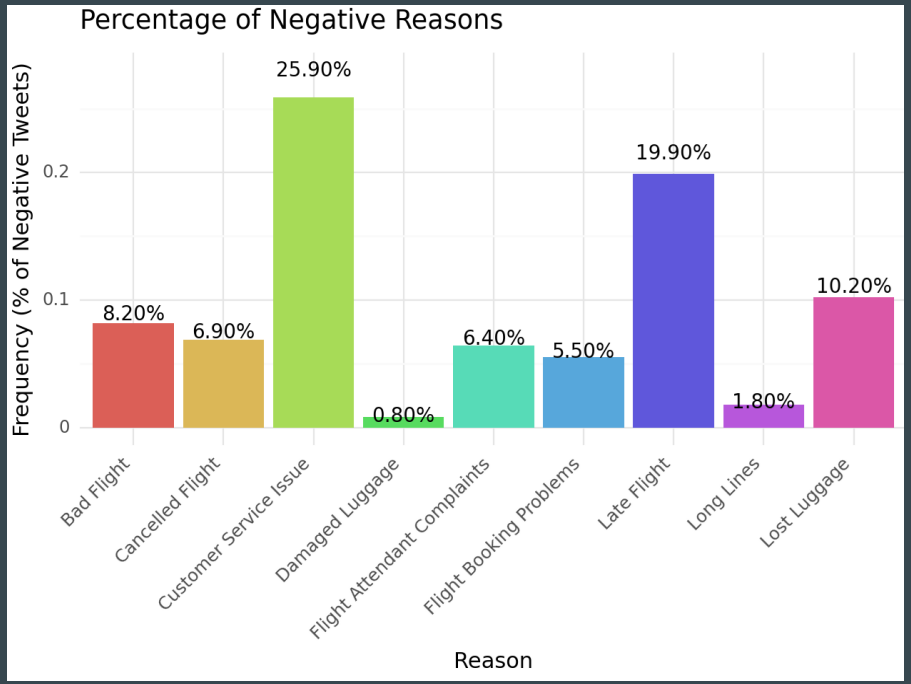
- Dataset consists of 14640 tweets collected from February of 2015 → short period so difficult determine confidence of analysis
- Dataset included many missing values for the gold sentiment based values, tweet coordinate, tweet location, and tweet timezone
- Dataset applied a sentiment analysis to the tweet text to determine whether the sentiment was positive, negative, or neutral, as well as a confidence score
- Breakdown of Tweets per Airline:
    - United: 3822, US Airways: 2913, American: 2759, Southwest: 2420, Delta: 2222, Virgin America: 504
- Since I wasn't creating any learning models the quality of the dataset wasn't impossible to work with as the only other data I needed was in the tweets
    - For this I just pulled a random sample of them and physically read through them to check
- Main Issues I had with Dataset: no documentation of variables, missing/ NULL values, some negative reasons were too general

Understanding Negative Reasons

- I ignored the "Can't Tell" section since the model used to detect the sentiment was unable to find a reasoning (would have to look at text)
- Most reasons are self explanatory, but other require a closer look at the texts themselves
    - Customer Service Issue Common Reasons: issues with website/ applications crashing, late response times to inquiries/ not able to talk to representatives, issues with transferring tickets, rude employees
    - Bad Flight Common Reasons: customer centric mechanical issues (seat, headphone jack, entertainment system), spotty wifi, issues with seats (seat placement/ amount of space) issues with food/ drink options, and some flight delays
    - Flight Booking Problems Common Reasons: rewards points not working, website crashing/ no confirmation emails unhelpful staff at airport

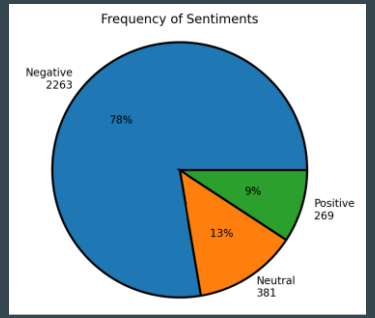
United Airlines Takeaways/ Insights

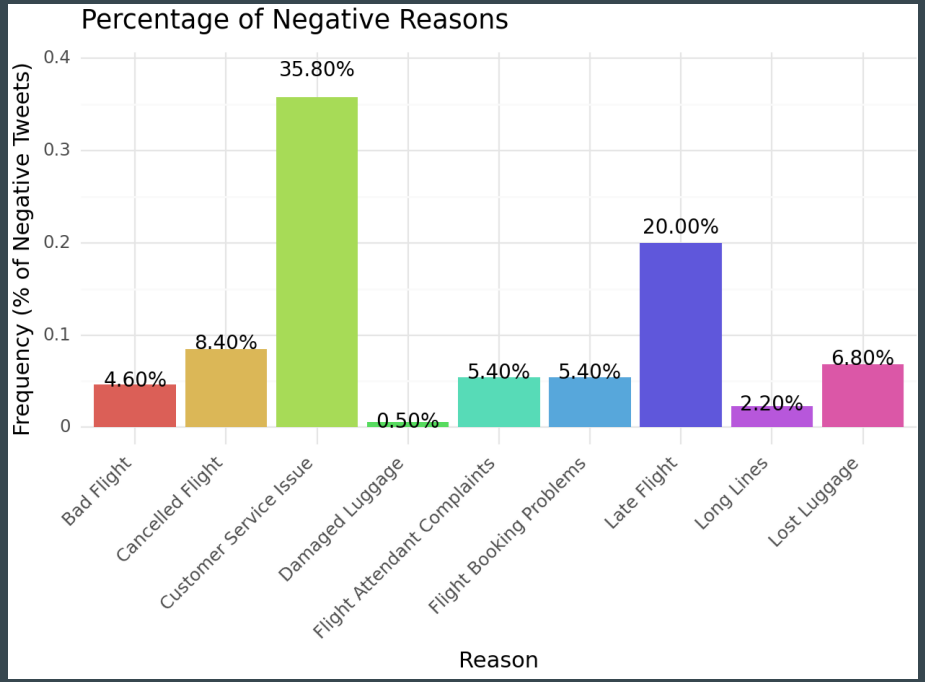




- Of the 3822 tweets centered around United Airlines, 2633 (69%) were of a negative sentiment
- United Airlines clearly has a significant problem with customer service, and looking at some sample tweets, common problems include:
  - Little to no communication (delayed communication) with representatives, rude employees (similar to flight attendant complaints), and website problems
- United also seems to have operational challenges with their late/ cancelled flights, and lost luggage problems
- For a company that has a comparatively larger market share, they have a lot of customer service and operational based complaints
  - They should consider allocating more budget towards fixing these issues through training and improving their operations processes

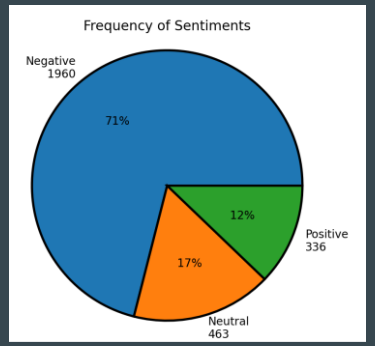
US Airways Takeaways/ Insights

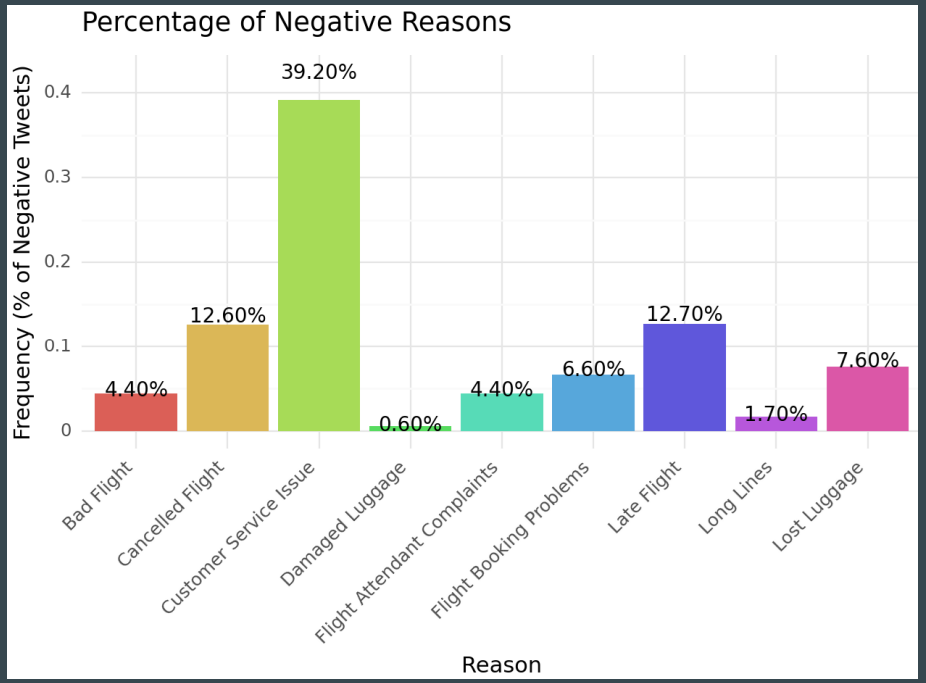




- Of the 2913 tweets centered around US Airways, 2263 (78%) were of a negative sentiment
- Similar to United, US Airways clearly has a significant problem with customer service, and looking at some sample tweets, common problems include:
  - Little to no communication (delayed communication) with representatives, rude employees (similar to flight attendant complaints), and website problems
- US Airways also seems to have operational challenges with their late/ cancelled flights, but less issues with lost luggage
- For US Airways, the biggest problem that they face is their customer service
  - Tweet contents regarding customer service, often included references to other airlines having a much better customer experience → US Airways needs to prioritize making improvements there
  - Prioritize resource allocation to Customer Service and then Operation problems

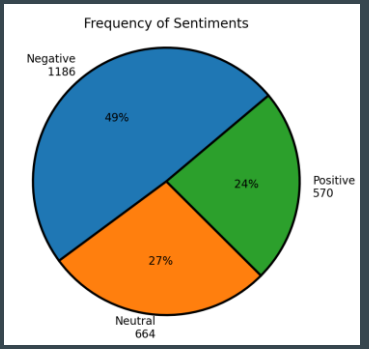
American Airlines Takeaways/ Insights

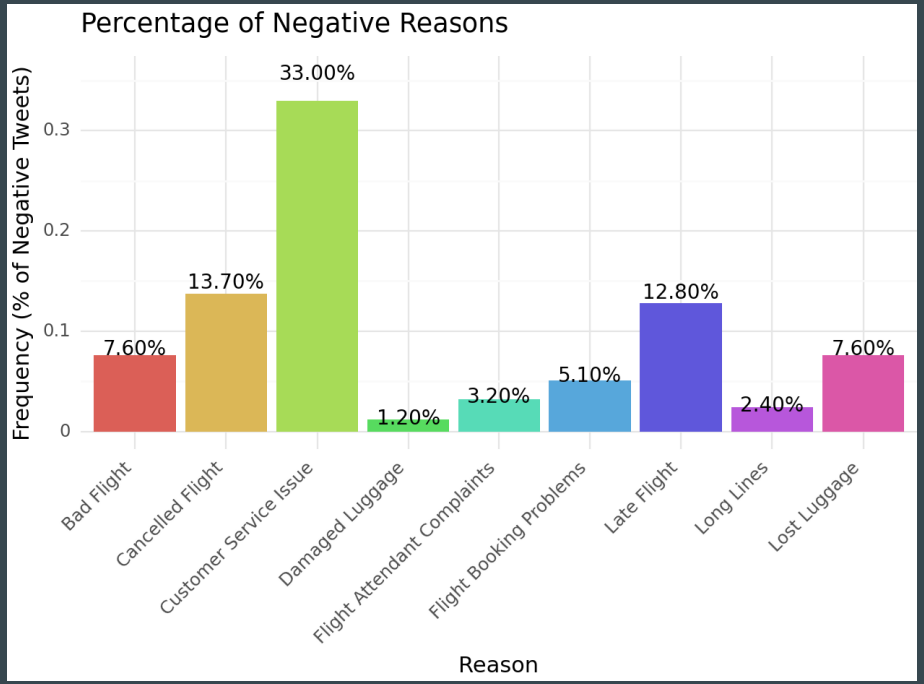




- Of the 2759 tweets centered around American Airlines, 1960 (71%) were of a negative sentiment
- American Airlines clearly has a significant problem with customer service as it makes up ~40% of the negative tweets, and looking at some sample tweets, common problems include:
  - Little to no communication (delayed communication) with representatives, rude employees (similar to flight attendant complaints), and website problems
- Similar to US Airways, the biggest problem that American Airlines faces is their customer service and then cancelled flights after that
  - Prioritize resource allocation to Customer Service as most tweets are created because of bad customer service
  - Also prioritize resources to cancelled flights issues → significant problem as they hurt the airline's public view more than other operational problems

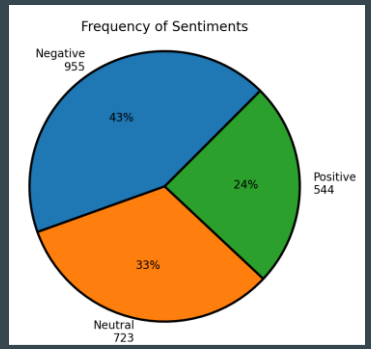
Southwest Airlines Takeaways/ Insights

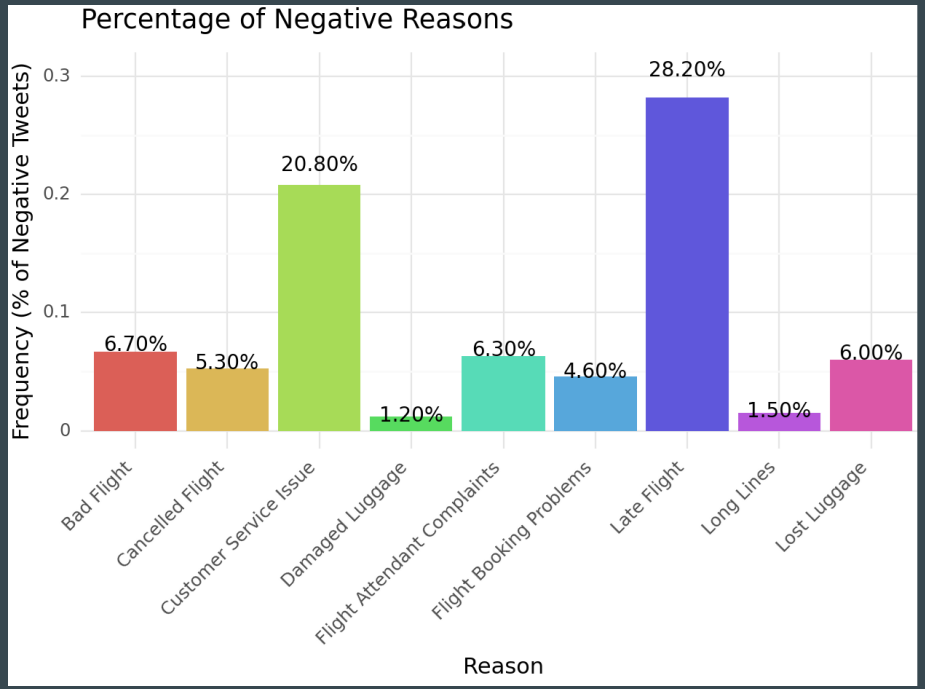




- Of the 2420 tweets centered around Southwest Airlines, 1186 (49%) were of a negative sentiment → among lowest proportion
  - This paints a comparatively positive picture of how Southwest's consumers view them
- Southwest Airlines trends are very similar to American Airlines → clearly has a significant problem with customer service, and looking at some sample tweets, common problems include:
  - Little to no communication (delayed communication) with representatives, rude employees (similar to flight attendant complaints), and website problems
- The customer service of Southwest makes up the greatest proportion of negative tweets also has the highest proportion of cancelled flights
  - Prioritize resource allocation to Customer Service as most tweets are created because of bad customer service
  - Also prioritize resources to cancelled flights issues → significant problem as they hurt the airline's public view more than other operational problems

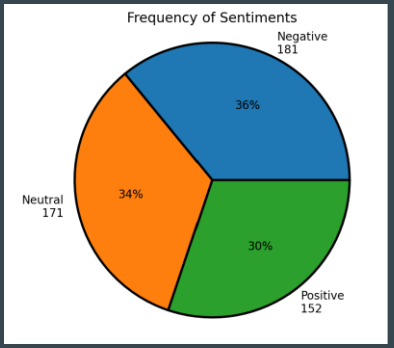
Delta Airlines Takeaways/ Insights

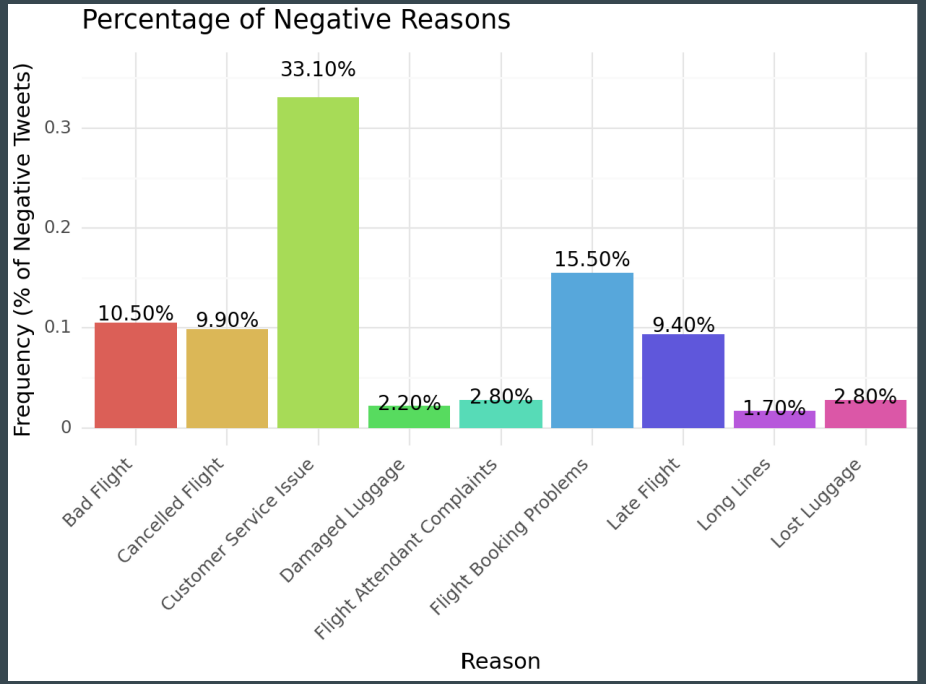




- Of the 2222 tweets centered around Delta Airlines, 955 (43%) were of a negative sentiment → among lowest proportion
  - This paints a comparatively positive picture of how Delta's consumers view them
- Delta Airlines has one of the most different trends
- Some problems with customer service, and looking at some sample tweets, common problems include:
  - Little to no communication (delayed communication) with representatives, rude employees (similar to flight attendant complaints), and website problems
- The issue responsible for the highest proportion of complaints is with late flights, but surprisingly not cancelled flights → might have to do with a specific few late flights having people with more of a social media presence (also had highest retweet)
  - Prioritize resource allocation improving processes to minimize operational problems
  - Allocation of resources to customer service should come after

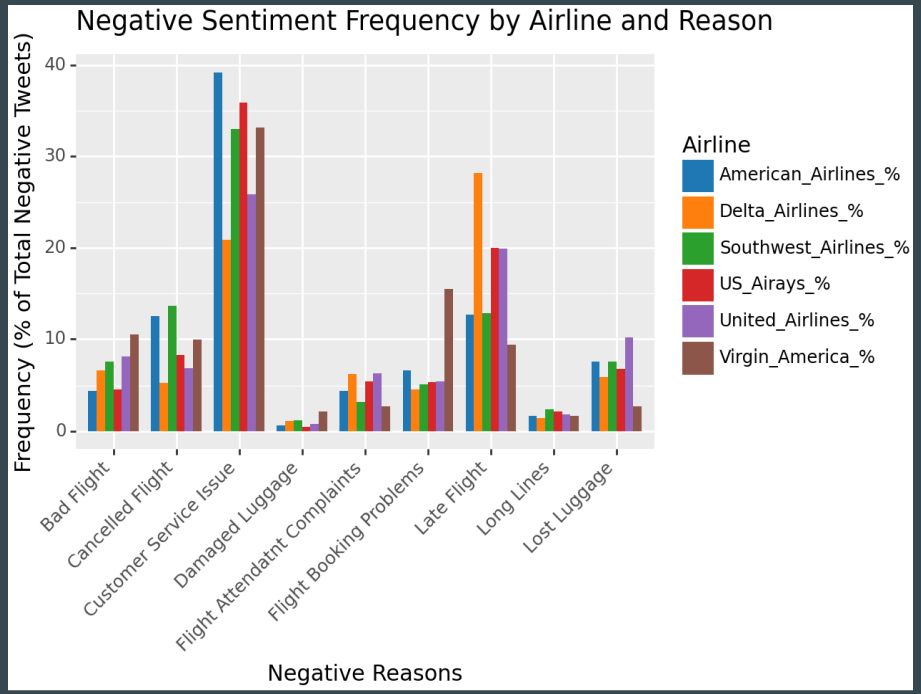
Virgin America Takeaways/ Insights





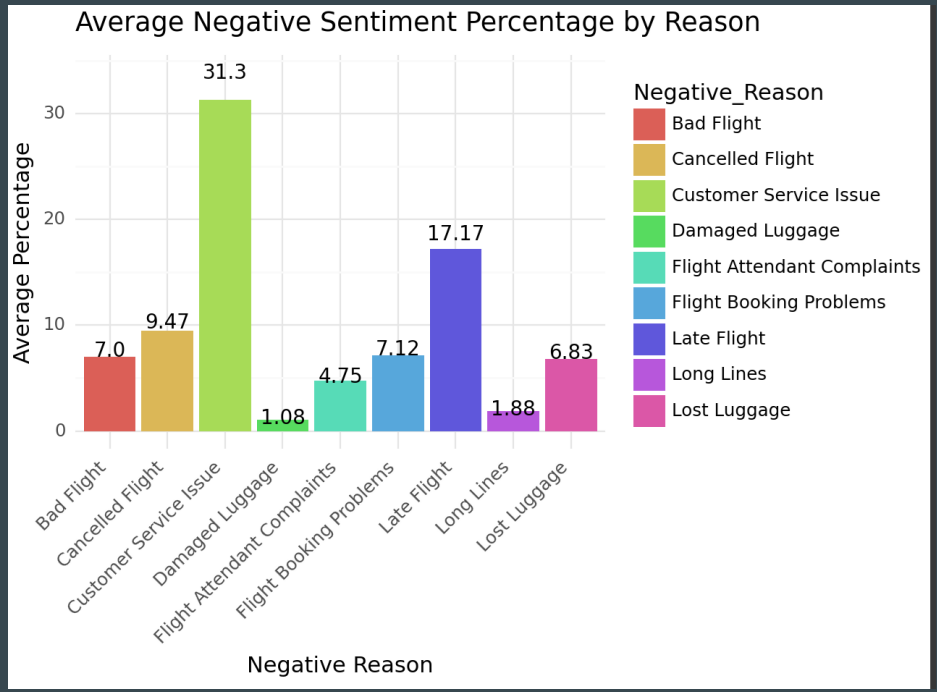
- Of the 504 tweets centered around Virgin America, 181 (36%) were of a negative sentiment → lowest proportion
  - Virgin America has the most balanced distribution of tweet sentiments
- Virgin America has a significant problem with customer service, and looking at some sample tweets, common problems include:
  - Little to no communication (delayed communication) with representatives, rude employees (similar to flight attendant complaints), and website problems
- Looking at the tweets, many of the “bad flight” reasons had to do with space on the plane, cleanliness, and broken amenities
- The customer service of Virgin America makes up the greatest proportion of negative tweets
  - Prioritize resource allocation to Customer Service as most tweets are created because of bad customer service
- Flight booking problems consisted of many website errors with seat selection/ saving not working
  - Allocate resources to fix bugs/ improve user experience on website

Comparison of Airline Trends



- By comparing each airlines percentage of negative complaints by proportion, we can visualize which airlines perform better in the different areas of customer service and operational processes
- Some notable spikes/ maxima throughout the airlines/ reasons:
  - American Airlines had the greatest proportion of customer service issues from their negative tweets
  - Virgin America had the greatest proportion of flight booking problems (by a considerable margin)
  - Delta had the greatest proportion of late flights (also by a considerable margin)
- Aside from those spikes, the distribution of airline negative tweet proportion by reason is relatively consistent

Key Insights



- By comparing the values for each airline to this benchmark of average percentage value, we can determine which airlines are performing, above, at, or below industry average when it comes to managing these kinds of complaints
- Since this data was pulled from tweets over the course of a one month period, we cannot extrapolate these findings with high confidence
    - We should monitor these trends and see the average deviation between the reason percentages over a larger time frame
- While some perform better than others, every airline should allocate more resources to customer service: improving training, increase hiring/ potentially hire more people as representatives to improve communication with customers
- Furthermore, all airlines should strive to minimize their issues with operational processes: improve website/ fix bugs, decrease number of delayed/ late/ cancelled flights that are not due to weather

Executive Summary

In the analysis of sentiment data related to various airlines, we have gathered valuable insights into passenger experiences and identified areas for potential improvement. The analysis of negative sentiment reasons by airline reveals several noteworthy findings:

- Customer Service Issues: Across all airlines, "Customer Service Issue" emerges as a common concern, representing a significant portion of negative sentiment (customer service issues (~31.3%), flight attendant complaints (~4.75%), and some aspects of booking problems (~7.12%) / bad flights (~7%)).
    - This indicates a pressing need for airlines to enhance customer service interactions and allocate more resources to support the hiring and training of employees and representatives when communicating and working with customers
- Operational Challenges: "Late Flight" (~17.17%), "Cancelled Flight" (~9.47%), and “Flight Booking Issues” (~7.12%) are also prevalent across airlines, which suggest potential operational challenges require attention to improve airline punctuality, reduce non-weather related flight disruptions, and enhance online booking process efficiency through updating/ fixing their websites.

The average sentiment percentage provides a comprehensive view of the overall negative sentiment landscape, and by benchmarking the various airlines with the industry averages, companies can gain insights into their relative market standing and allocate resources accordingly.

In summary, this analysis empowers airlines to enhance their customer experiences, streamline operations, and focus on areas that matter most to passengers. By addressing these concerns and continually monitoring sentiment data over time, airlines can work towards improving overall satisfaction and maintaining a positive brand image in the commercial airline industry.

Appendix (Action List)

1) After opening up the dataset the first few actions included reading through what each column represented and returning the number of NULL values for each column → With this I determined which columns to drop and if it was worth dropping any rows as well (decided not to since I was not developing any learning models)
2) Once I had a decent understanding of the columns and data I was working with, I determined that there weren't many models that I could make that would be necessarily effective
    a) I considered making my own sentiment analysis model but decided against since we had one provided and didn't want to fiddle around with fine tuning the model to have high accuracy when I couldn't benchmark it with the existing sentiment analysis model's
3) From there I created a dataframe for each airline, returned the avg and max retweets per airline to see if there would be any significant values that would indicate a particularly bad instance since the timeframe of the dataset is smaller → values were not particularly significant
4) Once that was setup I began the main analysis that I had planned for each airline:
    a) I returned a pie chart with the frequency of each sentiment → provides a basis for how customers view the company online
    b) I created a separate data frame that returned the frequency of each negative sentiment, as well as the proportion is was to the total number of negative sentiments → since some airlines have more tweets than others, this standardizes the values as a proportion rather then count frequency
    c) I then returned a barchart that plotted the negative reasons with their respective proportion of the total negative tweets (per airline)
5) After the above steps were completed for each airline I created a clustered bar chart that returned each airlines proportion of negative tweets based on the negative reason → allowed for direct comparison of airlines and the percentage frequency of different complaints
6) I then created another bar chart that returned the "industry average" values based on the average proportion of reason/ negative tweets across all airlines → allows for benchmarking airlines with industry average
7) With all of this done I began making the slides and pulling the insights from the raw tweets and the visualizations created