X (Twitter) Sentiment Analysis for Top Airlines in the USA

Analysis, findings and recommendations

Presented by: Aviators
Joseph Kinuthia, Raphael Kariuki, Emily Owiti, Paul Muriithi, Sylvia Muchiri & James Mungai

Problem Statement

- **Social media generates abundant data,** presenting both opportunities and challenges for organizations
- This data **can offer unparalleled insights** into customer perceptions, preferences and feedback
- Many organizations are **yet to develop** frameworks and strategies to analyze and interpret this data
- Insights from this data holds the **potential to benefit various domains** e.g., business operations, marketing strategies, customer service management & other
- **Our stakeholders** (United Airlines, US Airways, American Airlines, Delta Airlines and Southwest Airlines) have requested us to analyze data from social media and showcase how the companies can utilize such data to generate insights for use in driving improvement in business operation & customer service

Context - Airline Industry Overview

Key trends & Industry characteristics:

- **Digital transformation and technology:** Increasing adoption e.g., with deployment of apps, self service, AI chatbots & data analytics
- **Sustainable and environmentally friendly practices:** Significant focus on clean & efficient fuel usage. Adds pressure on profit & brand perception
- **Partnerships and alliances:** Many programs to expand networks, improve cost-efficiency & offer a variety of travel options
- **Health and safety measures**: Adopting stringent health and safety protocols to regain traveler confidence adding to pressure on profit
- **Airline Industry profit Margins:** ~ 2-5%. High impacted by fuel prices, inflation & price sensitive demand
- **Other characteristics**: Critical infrastructure & very visible to the public: Both of these impact brand perception and customer loyalty

Scaled technology adoption critical to future going concern & sustainability

Project Objectives

Main Objective:

1. Perform analysis on raw tweets & extract public sentiment & as well as build a chatbot solution

Specific Objectives:

1. To build a model that classifies raw tweets into the 3 sentiment classes for future use
2. To visualize the top drivers for each sentiment to help management target service delivery improvement
3. To create & deploy a chatbot to monitor customer feedback on X that provides real time responses

Data sources

CrowdFlower: Airline Twitter Sentiment - dataset by crowdflower | data.world

CrowdFlower (now part of Figure Eight Inc.) is a crowdsourcing platform that sources a diverse online workforce to perform tasks such as data labelling, content moderation, sentiment analysis and more. The site also provides public datasets to enable analytics for various purposes

Our dataset was sourced from X in February 2015 and contains tweets & retweets from the public on all major airlines in the USA. The data contains 14,640 rows and 20 columns

Data analysis approach

1. Loading the data to pandas and analyzing the dataframes
2. Cleaning the data by checking & handling:
    - Duplicates
    - Missing data
    - Anomalies
    - Invalid data
    - Text preprocessing
    - Performing exploratory analysis & creating visualizations
3. Modelling
4. Drawing conclusions and making recommendations

Research Questions (RQ)

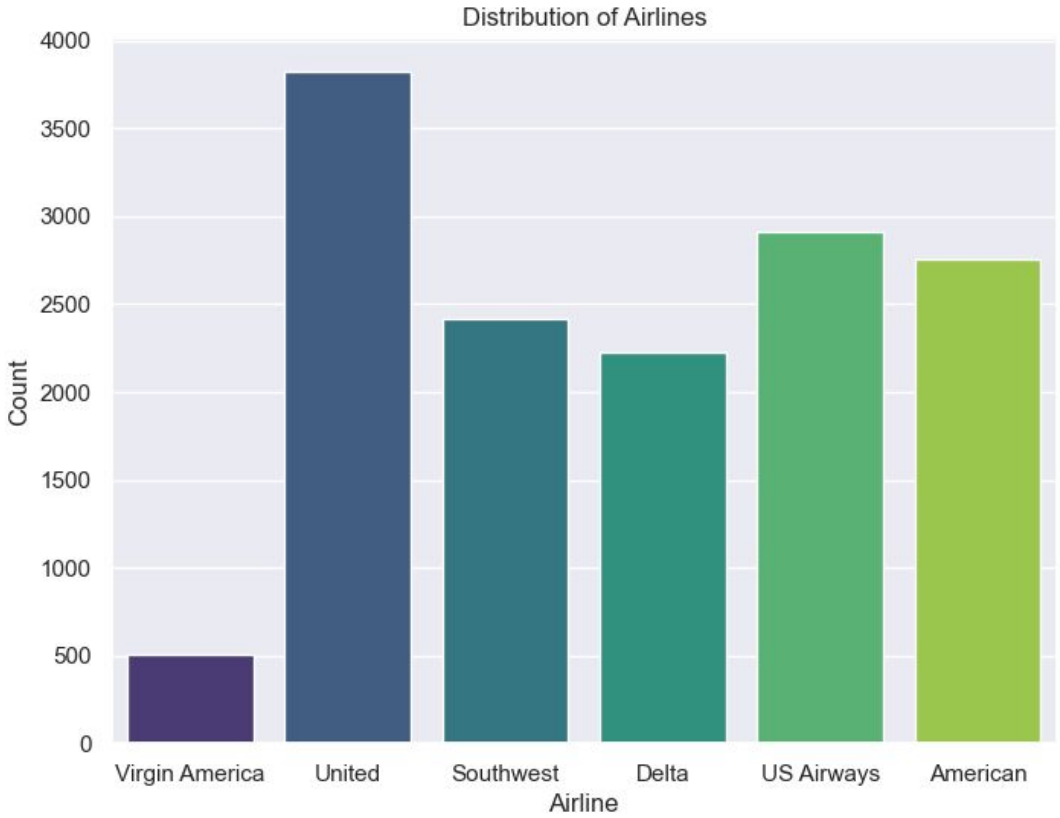
The project will be answer the below research questions:

1. What are the major airlines represented in tweets per the dataset?
2. What are the predominant sentiments expressed regarding major U.S. airlines?
3. How does the sentiment compare between the various airlines in our dataset?
4. What are the most common reasons for negative sentiments?
5. What are the most common reasons/terms used in positive sentiments?

Observations & Conclusions

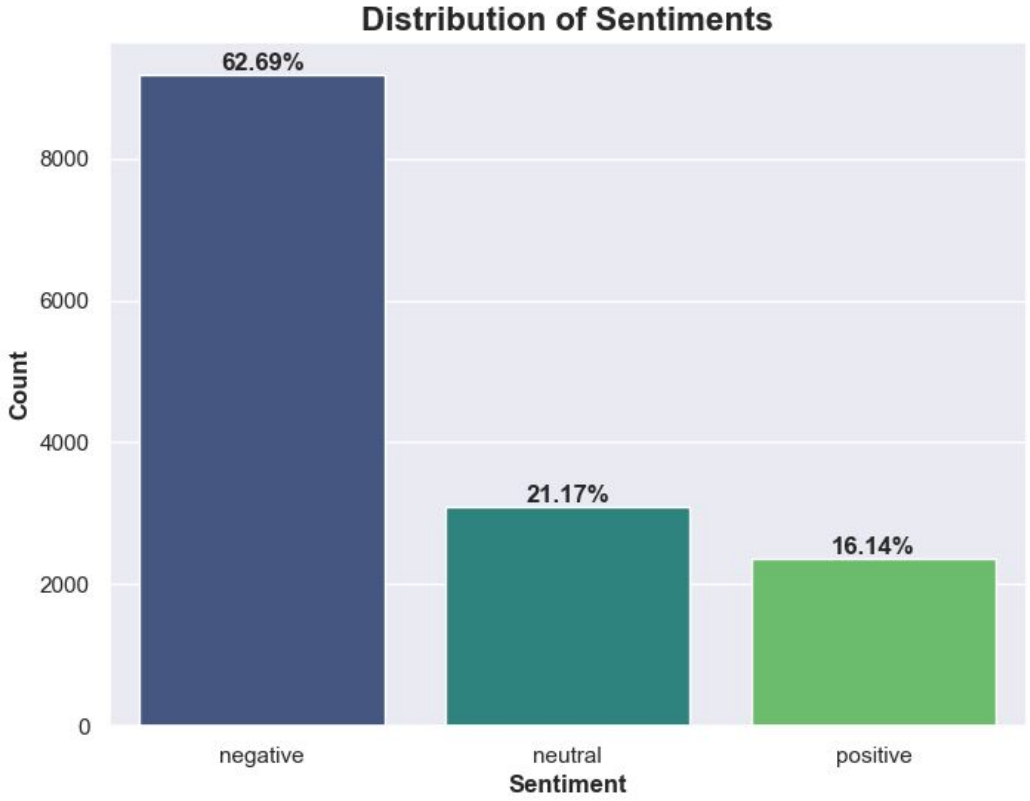
See next pages

RQ1 - What are the major airlines represented in tweets per the dataset?



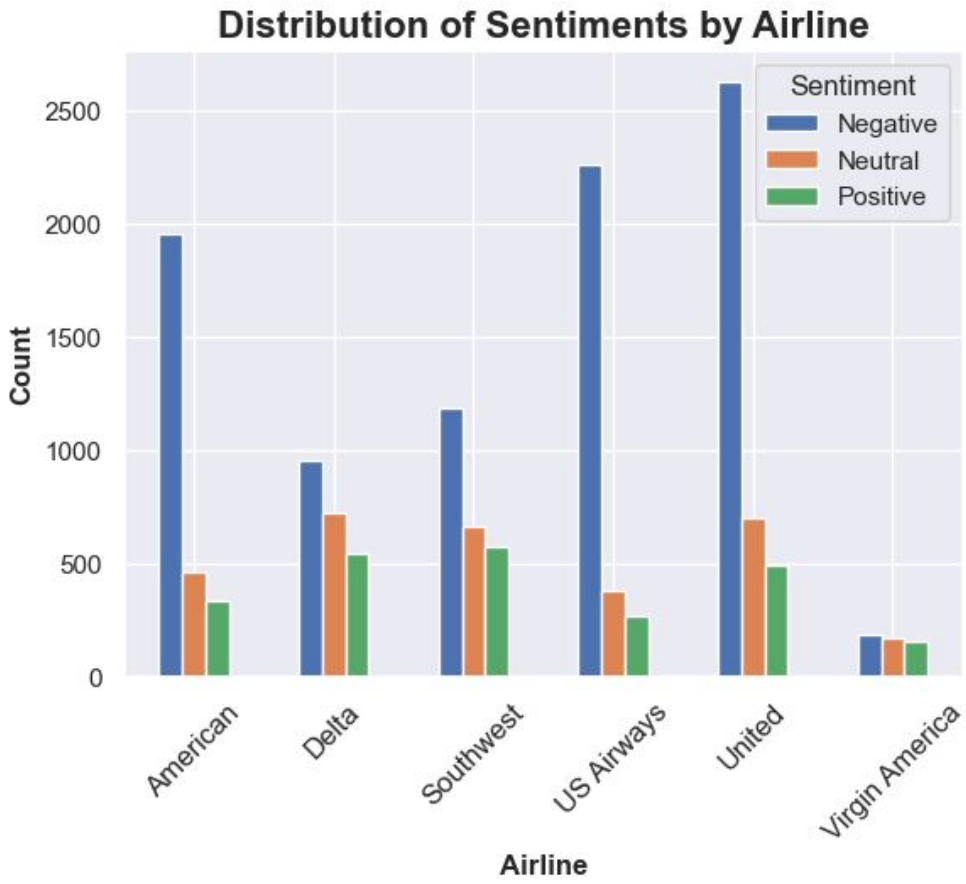
- This chat shows the distribution of tweets amongst the top six airlines in America that were subject of discussion on X in the time period of February 2015
- United had the highest share of tweets followed by US Airways & American

RQ2 - What are the predominant sentiments expressed regarding major U.S. airlines?



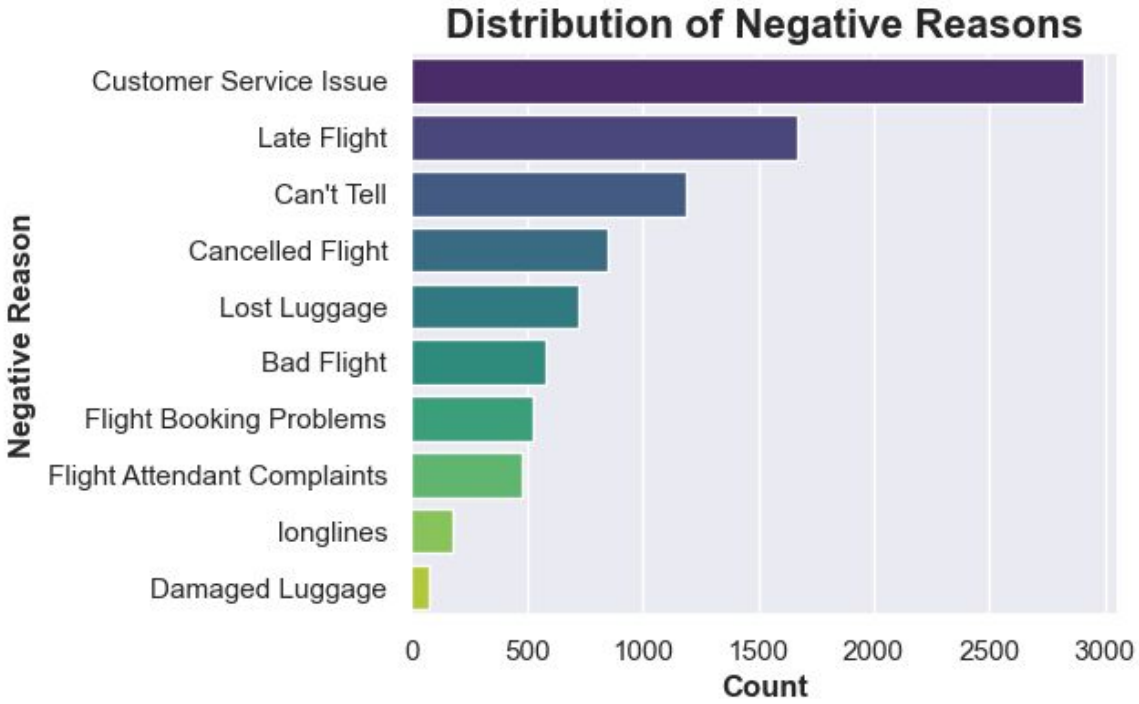
- Public sentiment categorized into: Positive, Neutral & Negative
- Negative sentiment has the highest score making 62.7% of total sentiment & positive sentiment has the least %
- These metrics point to the need root cause analysis to enable decisions that improve customer experience

RQ3 - How does the sentiment compare between the various airlines in our dataset?



- Negative sentiment highest across airlines
- United, US Airlines & American have the highest proportion of negative sentiment in comparison to their respective positive & neutral sentiments

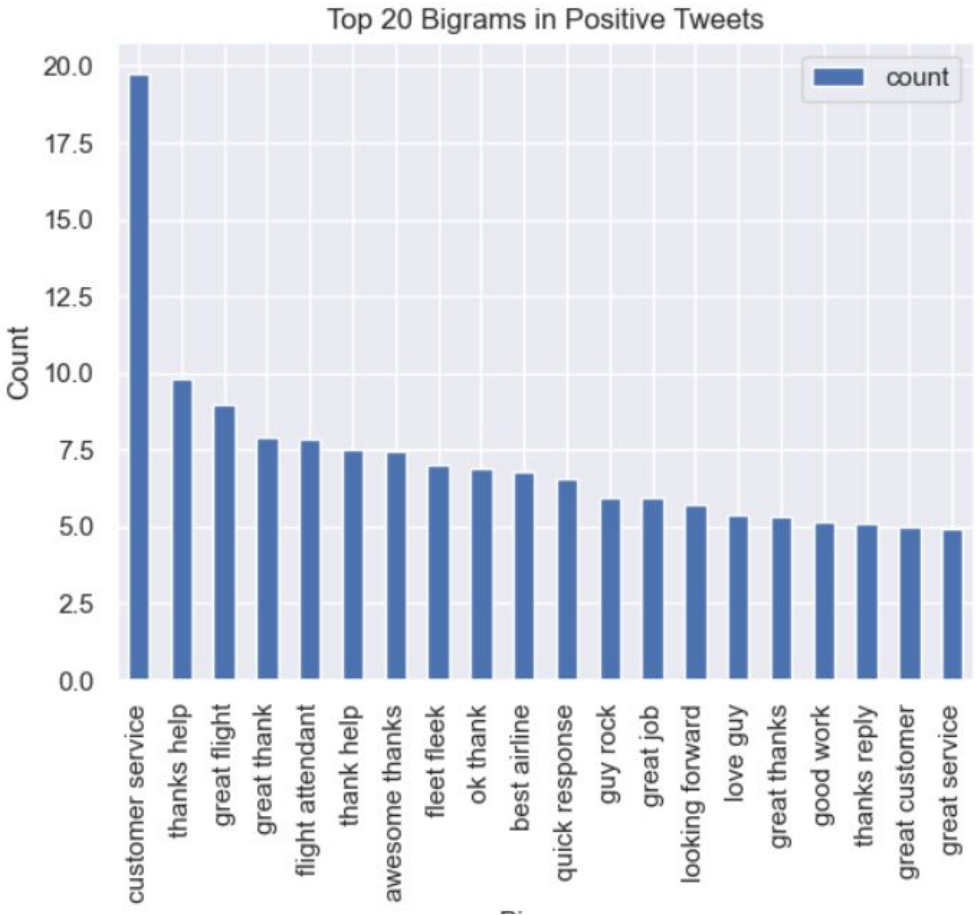
RQ4 - What are the most common reasons for negative sentiments?



See negative sentiment distribution per airline in appendix

- The top 5 reasons include: customer service issues, late & cancelled flights & lost luggage
- 3rd highest category is ’Can’t tell’ needs further investigation to understand real driver(s)
- Observation: February is a cold month with heavy snows that causes massive flight cancellations driving customer dissatisfaction

RQ5 - What are the most common reasons/terms used in positive sentiments?



- *The bigram shows a distribution of the most common positive words across the positive tweets*
- *Positive tweets are dominated by expressions of appreciation e.g., “thank you”, “great flight”, “best airline”showing appreciation for service delivery per customer expectation*

Conclusions

Following observations in previous section, we made the following recommendations:

1. *Sentiment distribution:* Negative sentiment represents 62.7% of the total sentiments highlighting the need to address customer concerns & improving their satisfaction
2. *Sentiment distribution across airlines:* Analysis revealed insights on sentiment distribution across airlines that can be used for benchmarking & strategic play
3. *Common drivers for negative sentiment:* Includes: Customer Service, late flights, flight cancellations showing potential opportunity to improve customer experience

Recommendations

1. Address flight punctuality: Opportunity to implement flight operations, scheduling & contingency plans
2. Invest in luggage handling: Opportunity to upgrade luggage handling systems to minimize incidents and ensure a smoother travel experience
3. Enhance online booking systems: Opportunity to streamline online booking systems to make it easier and more user-friendly
4. Proactively monitor social media: Opportunity to deploy scalable solutions (e.g. chatbots) to monitor & respond promptly and address concerns or complaints as they occur

Next steps

1. Continuous monitoring: Implement chatbot integration to X to continuously monitor and analyze customer sentiments in real-time and provide feedback

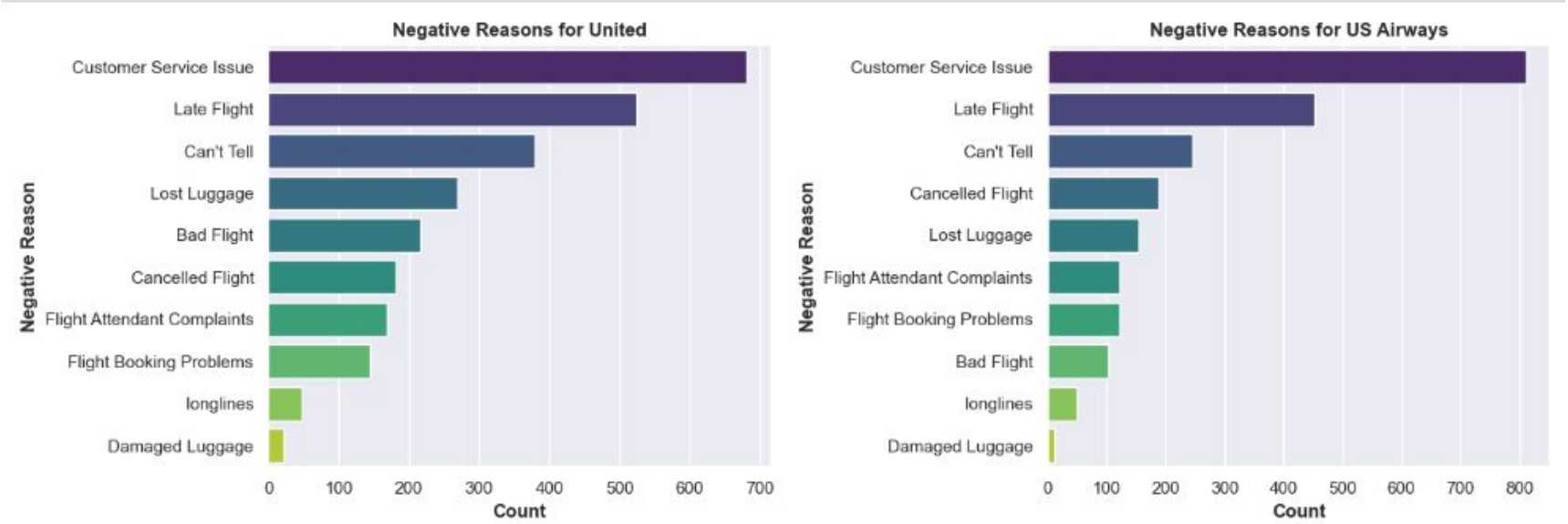
Thank you!

Presented by: Aviators

Appendix

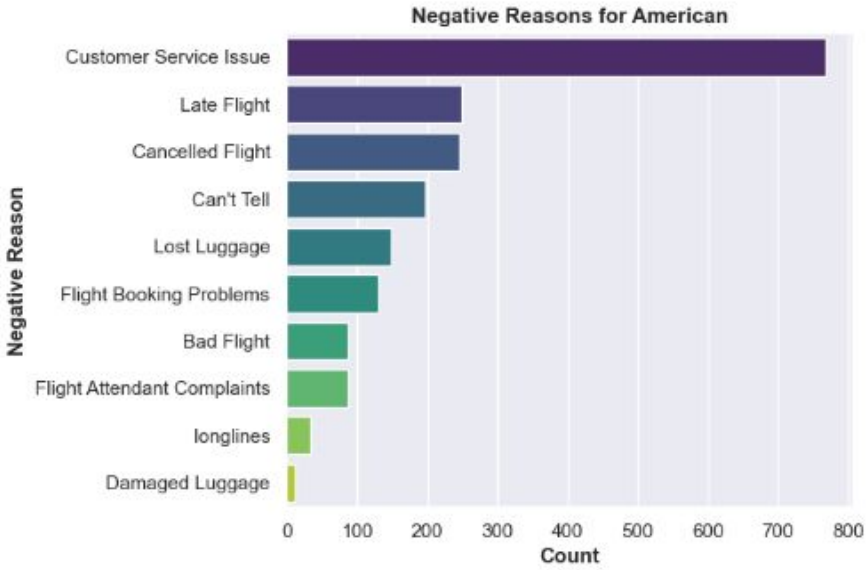
See next pages

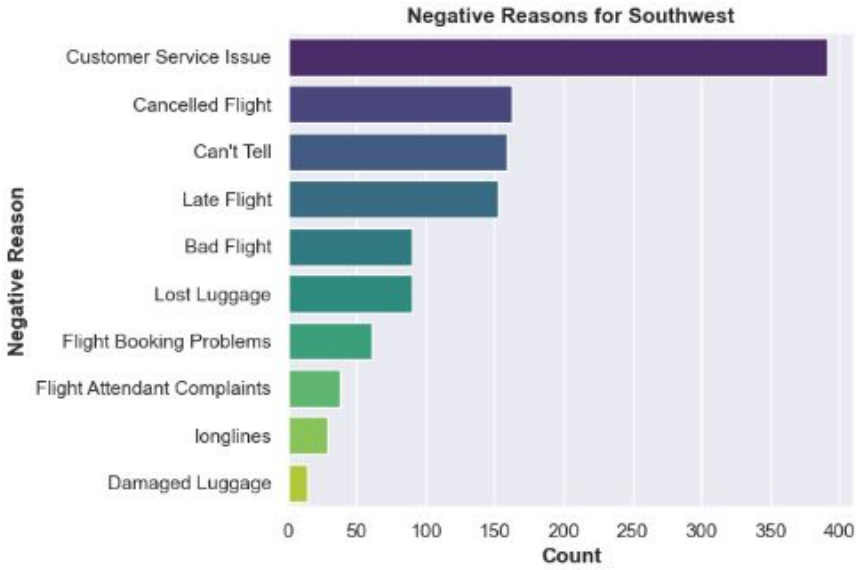
Top Reasons & Drivers for Negative Sentiment - United & US Airways



- Top reasons for negative sentiment consistent with all airlines' top reasons

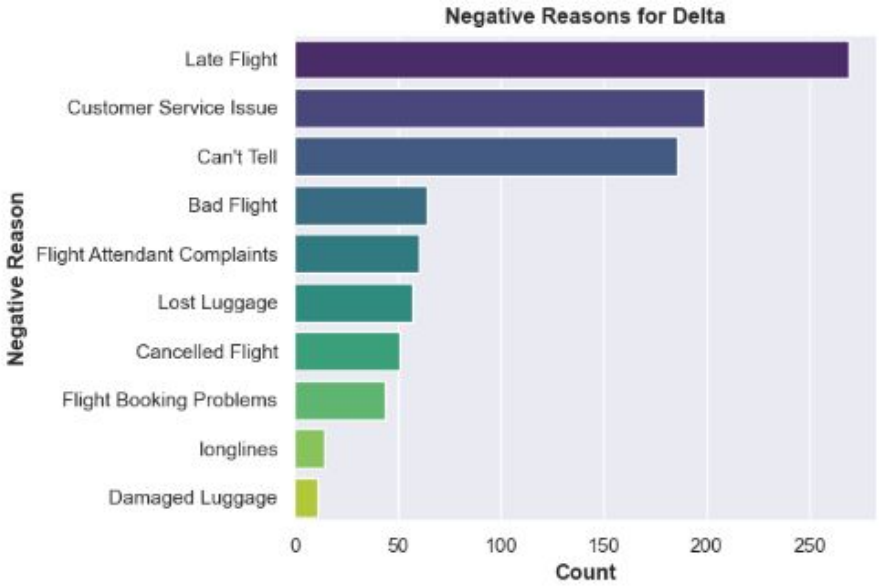
Top Reasons & Drivers for Negative Sentiment - American & Southwest

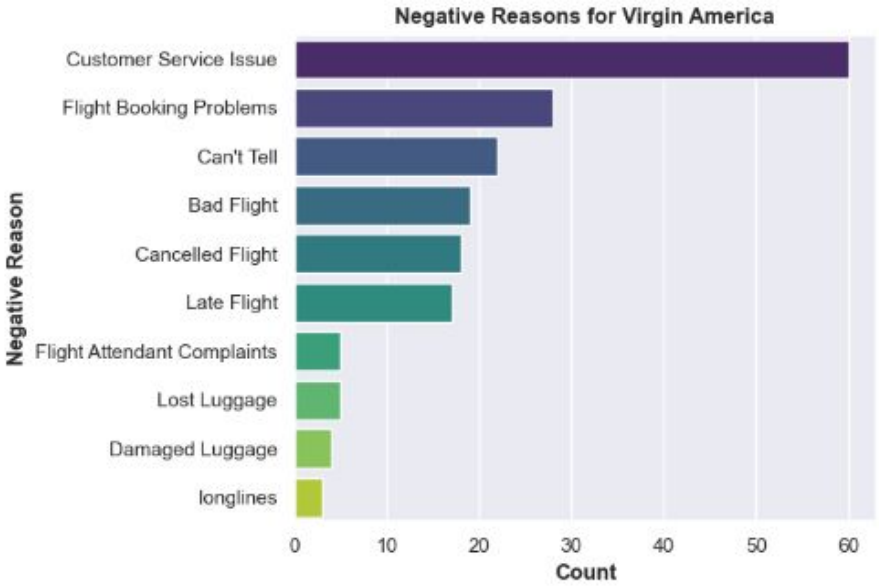




- Top reasons for negative sentiment consistent with all airlines' top reasons

Top Reasons & Drivers for Negative Sentiment - Delta & Virgin America





- For Delta & Virgin we see a slight variation in the top five drivers for negative sentiment, including the order - when compared to the other airlines
- Note: Delta & Virgin made the bottom two airlines in terms of total number of tweets