



Analyzing the Impact of Historical Events on Airline Travel

By Josh Burris, Jack Fitzgerald, and Thomas
Lujan



Background Information

- We were curious about how COVID-19 would affect airline traffic
 - How does it compare to other major events?
- No current datasets exist
 - What other datasets exist?
- Use past datasets to speculate about the present and future
 - How much can we expect it to change?



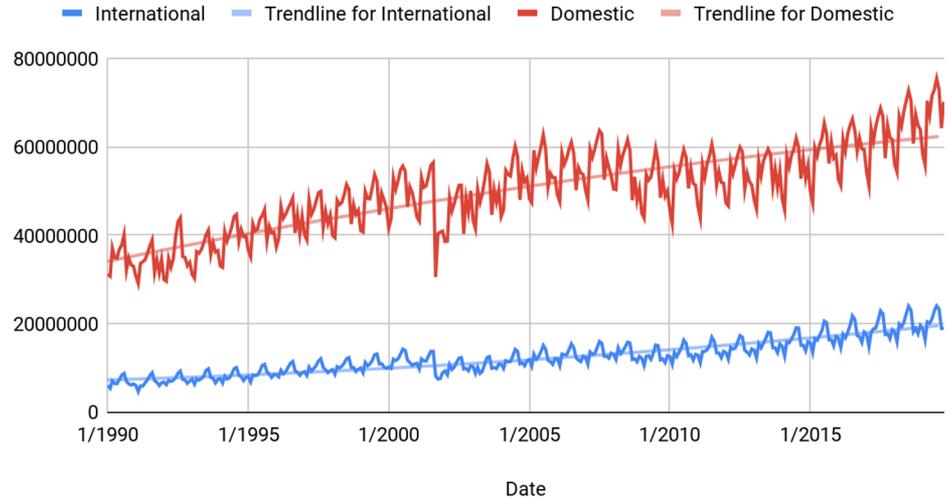
Problem Characterization

- What kind of major events are big enough to affect airlines?
- The datasets used for the analysis include:
 - T-100 Market (domestic and international travel data)
 - Schedule-P1.2 (airline financial data)
 - Reporting Carrier On-Time Performance (delay and cancellation data)
- Analyze statistics such as the total number of passengers, cancellations, and net income
 - Look at statistics for the entire industry or for individual airlines
 - Create visualizations

Verifying Data (Methodology Part 1)

- Expected data points:
 - Major drop in total flyers in September of 2001
 - A large gradual drop from 2008 to 2010
- How long did it take for flying to go back to normal?

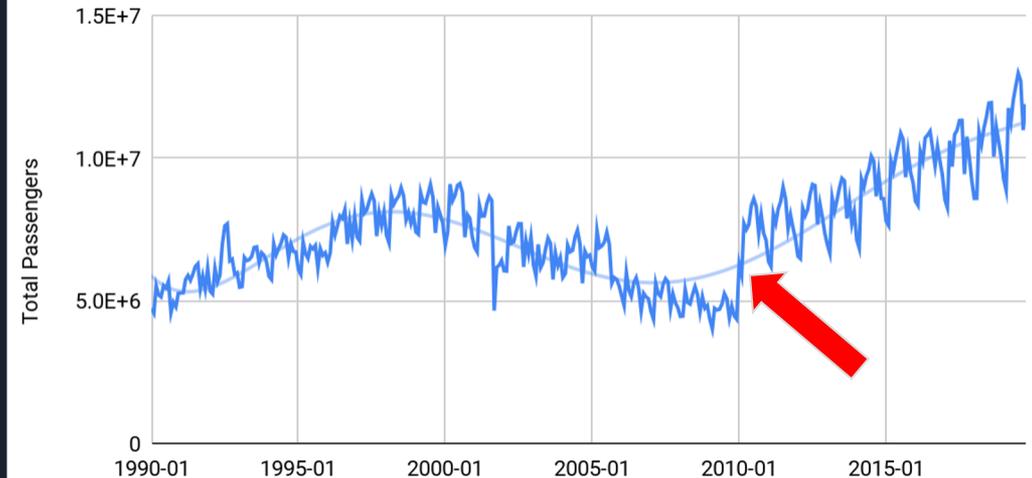
Figure 1: Total # of Passengers For All Airlines



Airline Popularity (Methodology Part 2)

- Analyze individual airlines in the T-100 Market and Schedule-P1.2 data to identify events which impacted only a few of the airlines
 - Useful for identifying events like mergers
- Example: Delta Airlines acquired Northwest in 2008, all operations were combined in 2010
- Created Spark jobs to find statistics for all airlines at once
 - Or for looking at a specific airline

Figure 3: Delta Airlines Total Domestic Passengers



Cancellations (Methodology Part 3)

- Used BTS OST_R data set containing information on delays and cancellations
- Created spark jobs to determine daily cancellations and average delays for years with impactful events (2001: 9/11 attacks, 2008: Recession, 2020: COVID-19)
- Events that threaten passenger health/safety reflect in spikes in cancellations, but events that do not have more long-term effect on air travel.

Figure 5: Cancellations Per Day For 2001

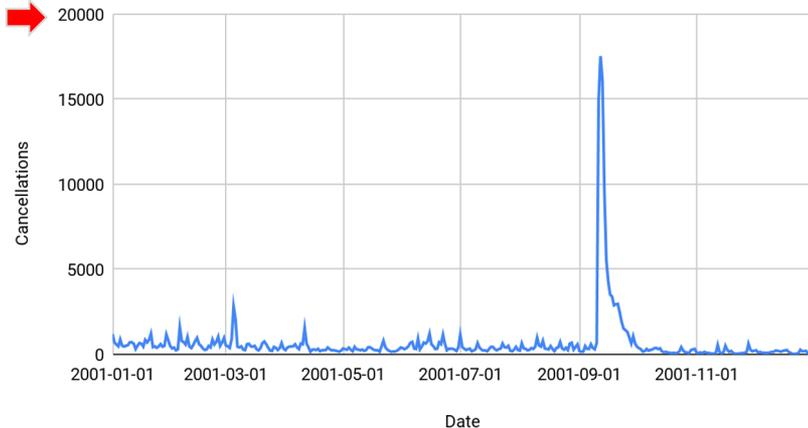
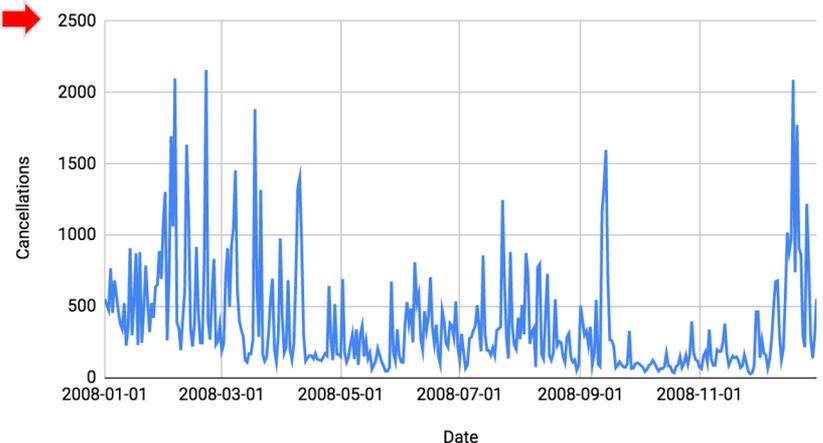


Figure 6: Cancellations Per Day for 2008





Airline Popularity Benchmarks

- Better to use less worker nodes when analyzing smaller datasets like Schedule-P1.2 (~500KB)
- For larger datasets like the T-100 Market (7.1GB), better to use somewhere between 5 - 10 worker nodes



Airline Popularity Benchmarks

- Average turnaround times for TotalNetIncomePerAirline.java:
 - 1 worker node = 12 seconds
 - 2 worker nodes = 14 seconds
 - 3 worker nodes = 21 seconds
- Average turnaround times for DomesticPassengersFreightAndMailPerAirline.java:
 - 2 worker nodes = 51.67 seconds
 - 5 worker nodes = 39.3 seconds
 - 10 worker nodes = 1.1 minutes



Cancellation & Delays Benchmarks

- Average turnaround times for TotalFlights.scala:
 - 1 worker node = 8.5 seconds
 - 3 worker nodes = 8.2 seconds
 - 10 worker nodes = 7 seconds
- Average turnaround times for TotalCancellations.scala:
 - 1 worker nodes = 12 seconds
 - 3 worker nodes = 9 seconds
 - 10 worker nodes = 10 seconds



Insights & Conclusions

Insights

- Events such as economic recession and public opinion do not reflect in travel logistics such as delays and cancellations, but DO have a large effect on airline profitability
- Hoping to gain insight to COVID, but this is an unprecedented external event that is not comparable to previous events. Also there is insufficient (published) current data.

Conclusions

- Government intervention (bailouts) can stop airlines from declaring bankruptcy, but recovery from these events still takes decades to regain normalcy.