# Flight Delay & Cancellation Analysis

Group 6:  Shuchang Liu, Youbin Mo, Ling Hong

# Motivation & Background

Product: Delay & Cancellation Insurance

Project Purpose: Pricing and Maximize profit

Mechanism: Generate prediction model

Key Factors Affects Delay/Cancellation:

- Airlines/Airports/States
- Year/Month

# Dataset

**Flight data**: [OST_R open dataset](#)
- ❏ Date range: 2009-2018, 70 Million flight info
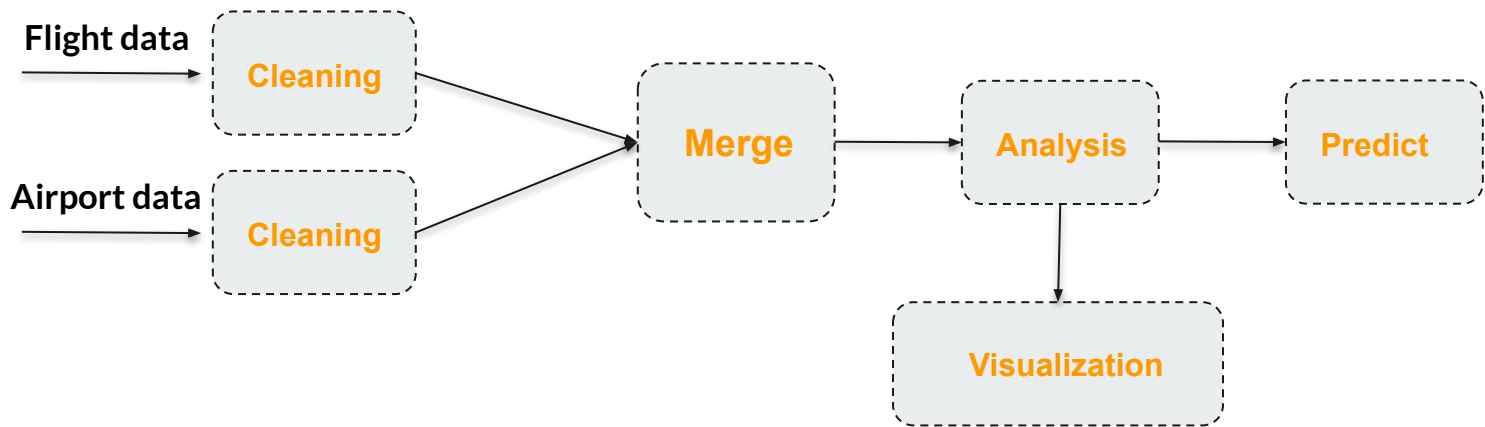
**Airport data:** [OurAirports open dataset](#)
- ❏ Size: 55535, US Airports: 2033

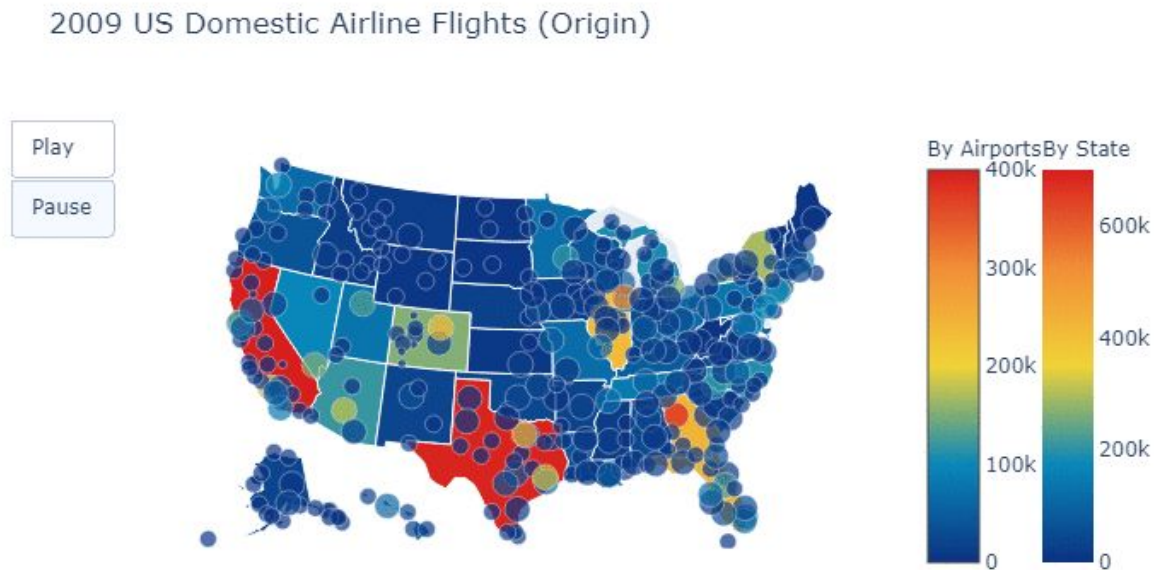**Airline code to Airline name: scrapped from Google**
- ❏ AA => American Airlines, F9 => Frontier Airlines

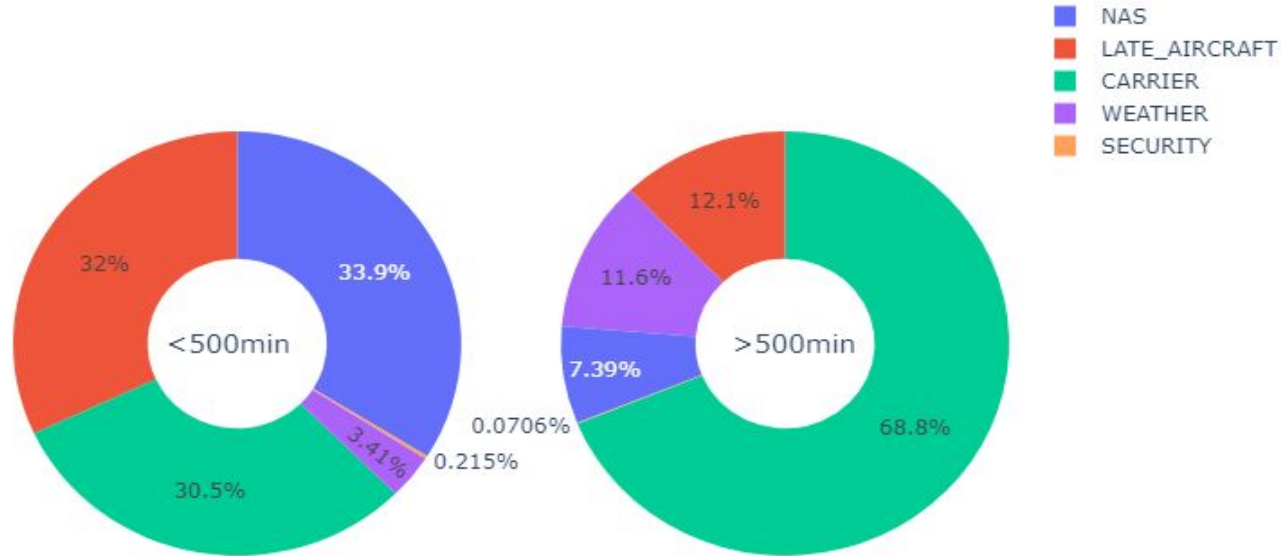**US states & region data**: [Github US states data set](#)

# Methodology

# Overview - Number of Flights (by year)



2009 US Domestic Airline Flights (Origin)

Flights number cause delay?

We find that during 2009-2018, the number of flights in U.S. is very stable.

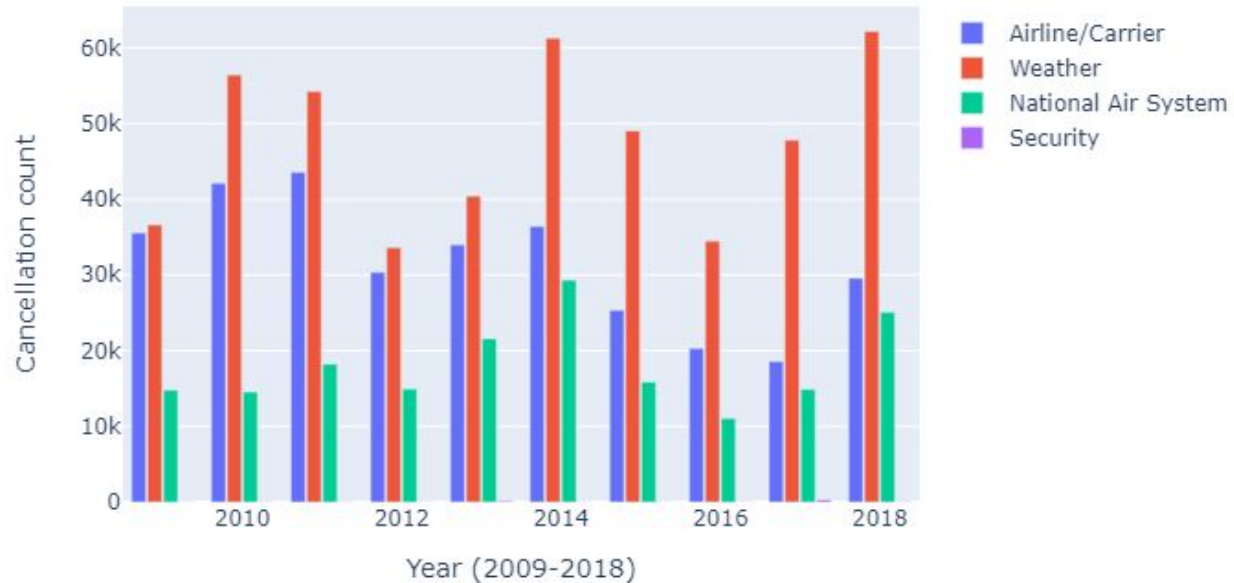# Overview - What is the Delay Reason?



The majority of delayed flights with 500 min or larger delay are caused by **Carrier** while the reasons for smaller delay are evenly distributed.

# Overview - What is the Cancellation Reason?

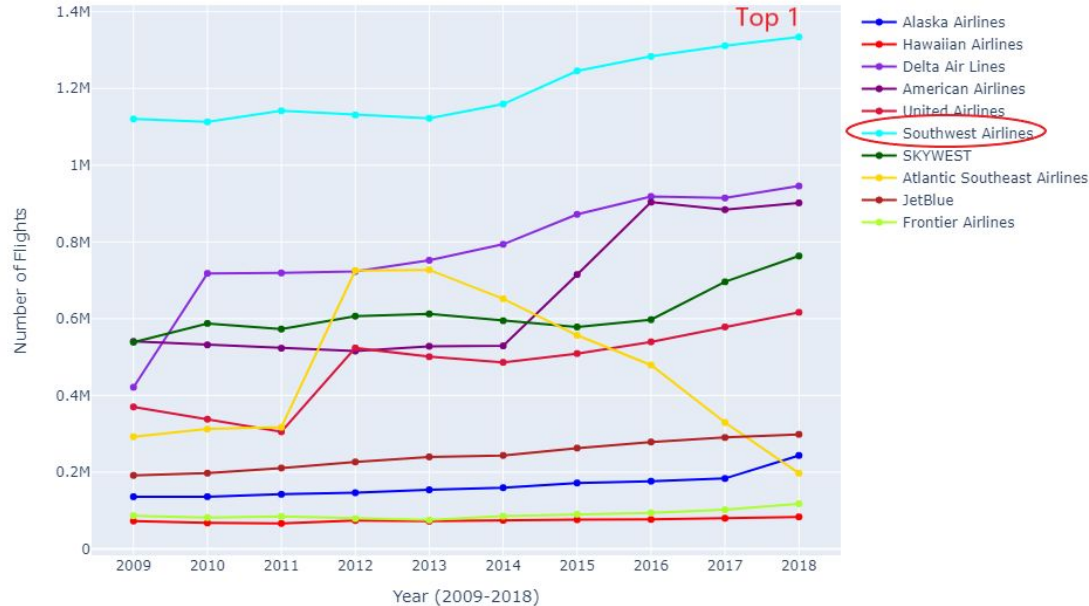What are the main reasons for flight cancellation?    Weather！！



Cancellation Reason Analysis

# Carriers Aspect

Which Airline has the highest flight count?



2009-2018 US Domestic Airlines Flights Number History

# Carriers Aspect

Which Airlines outperform others in terms of delay?
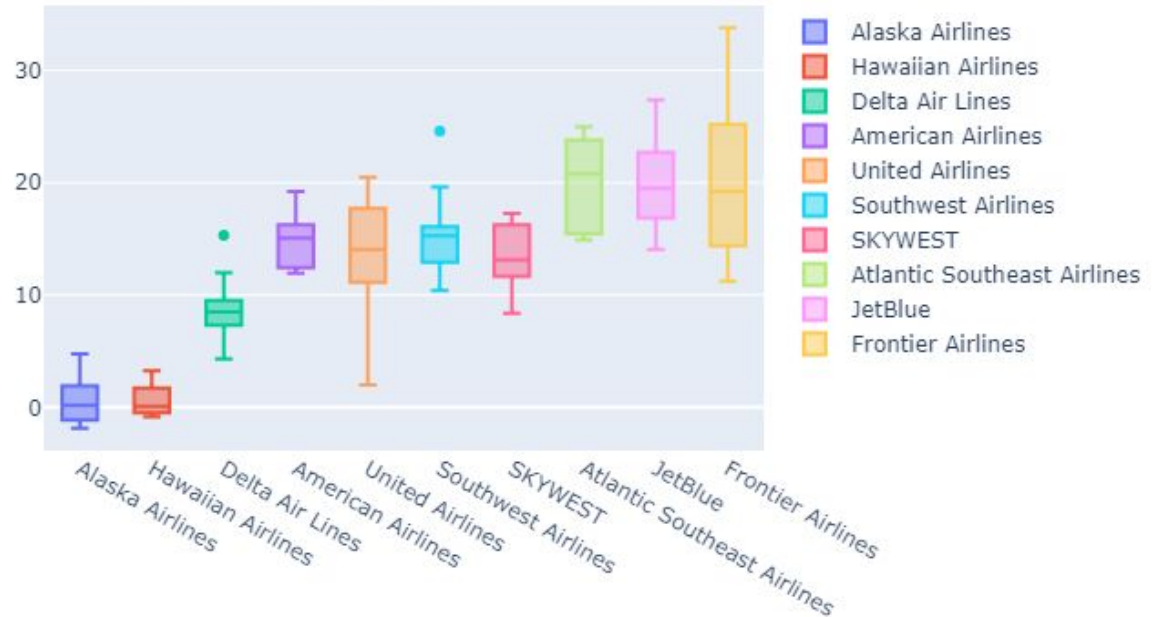


2009-2018 US Domestic Airlines Average Delay History

# Carriers Aspect

*Delay*

**Q: Which Airline will you choose?**

😎 Hawaiian
😎 Alaska Airlines

🙁 Frontier Airlines
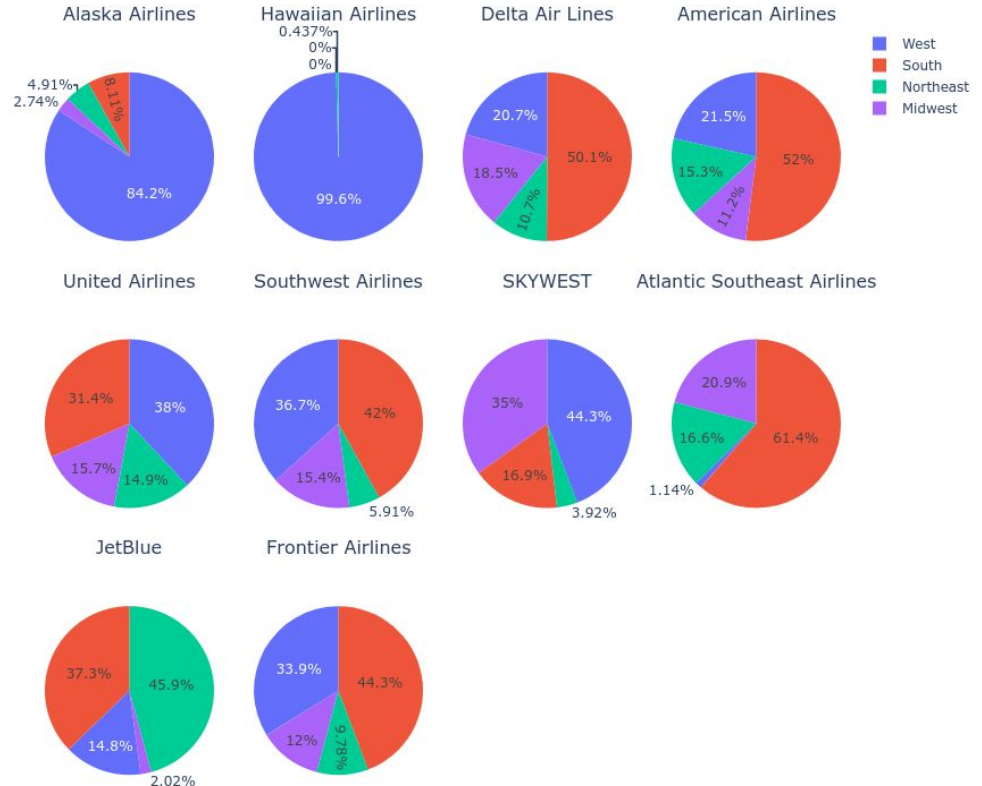


10 US Domestic Airlines Average Delay Box Plot

# Carriers Aspect

*Delay*

Q: Why Alaska & Hawaiian Airlines could reach much lower delay than other airlines?

A: From the chart on the right , we can see that for Alaska and Hawaiian Airlines, the majority of the flights are from/to the West Region. That may be a major reason for their best performance.
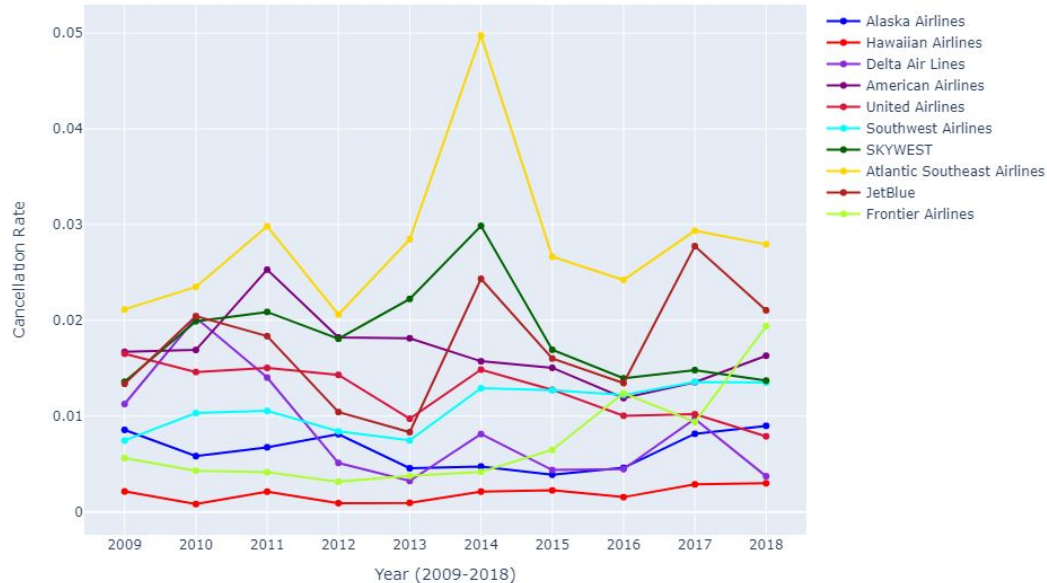


2018 US Domestic Airline Origin and Destination Statistics (by Region)

# Carriers Aspect

*Cancellation*

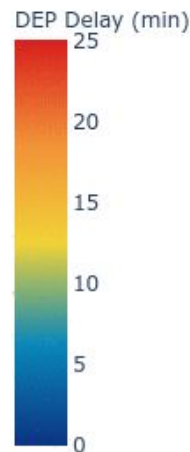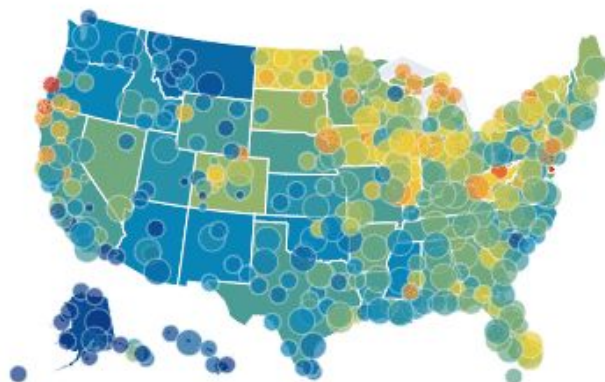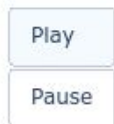2009-2018 US Domestic Airlines Cancellation History



We found

1. Delta Airline and Hawaiian Airline have the smallest cancellation rate.

2. Atlantic Southeast Airlines has high cancellation rate.

# Airports Aspect: Location? Month?

*Delay*



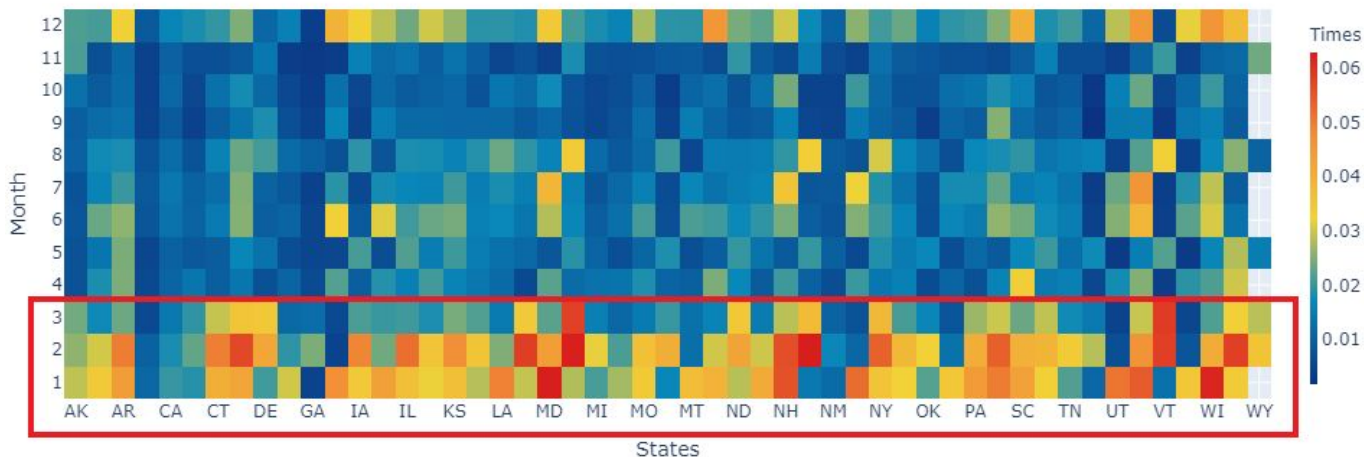January (01) US Domestic Airline Departure Delay (Origin)

Heatmap: Most states have higher average delay in **December, June, July.**
Locations: The delay is much higher in the Central & East area than West in summer.
Features for delay: Weather(month, location)

# Airports Aspect

*Cancellation*



2009-2018 Total Cancellation Rate with respect to month and state

Heatmap: Most states have higher cancellatin rate from **Jan. to Mar**.
X-axis: States, Y-axis: Month, Color: cancellation rate.
Features: weather(month, location)

# Prediction Model

Deep Neural Network

Input features: Flight Date, Carrier, Airport location and altitude

Output: Whether the flight will delay or cancel

Construct: 512*1024*512 Nodes

Training: Dataset from year 2009 to 2017

Test: Dataset 2018

Correctness: Predict 80% of delay and 95% of cancellation

# Future Work

- Use fine-grained weather data to analyze how weather is related to flight delay/cancellation. (wind, temperature, precipitation etc)
- Try another prediction model, like Bayesian Condition

# THANKS!