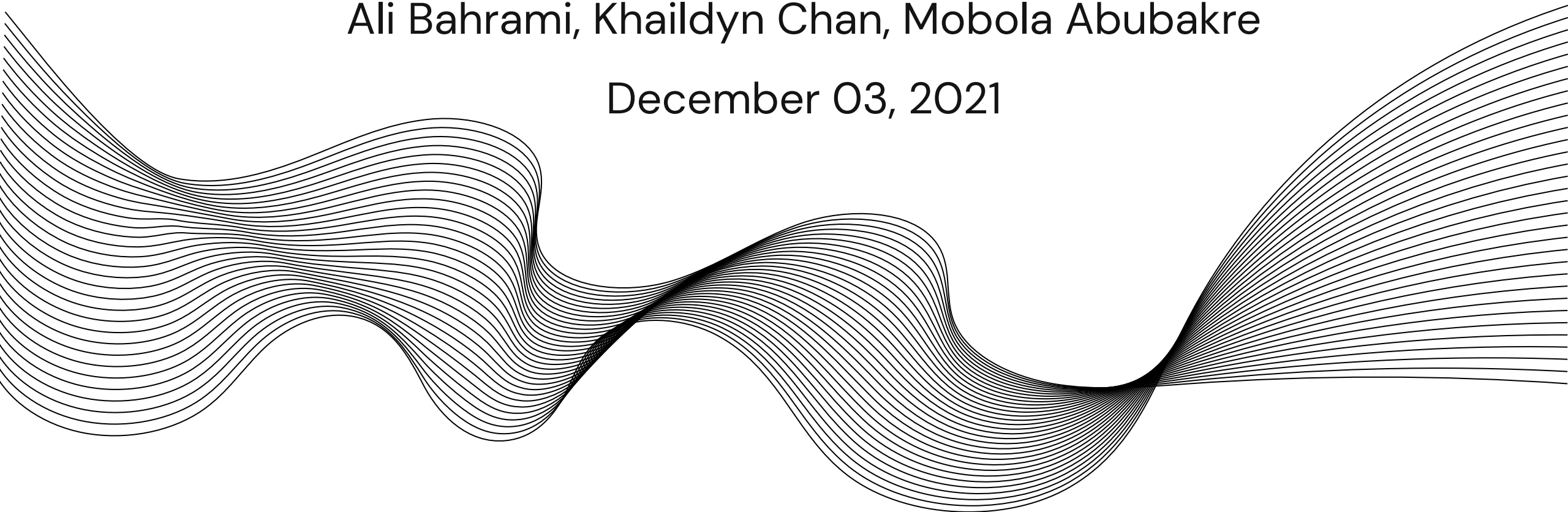


Predicting Flight Delays

Ali Bahrami, Khaildyn Chan, Mobola Abubakre

December 03, 2021



Agenda

1. Project Overview
2. Exploratory Data Analysis
3. Models and Evaluation
4. Conclusion

Why Predict Flights Delays?

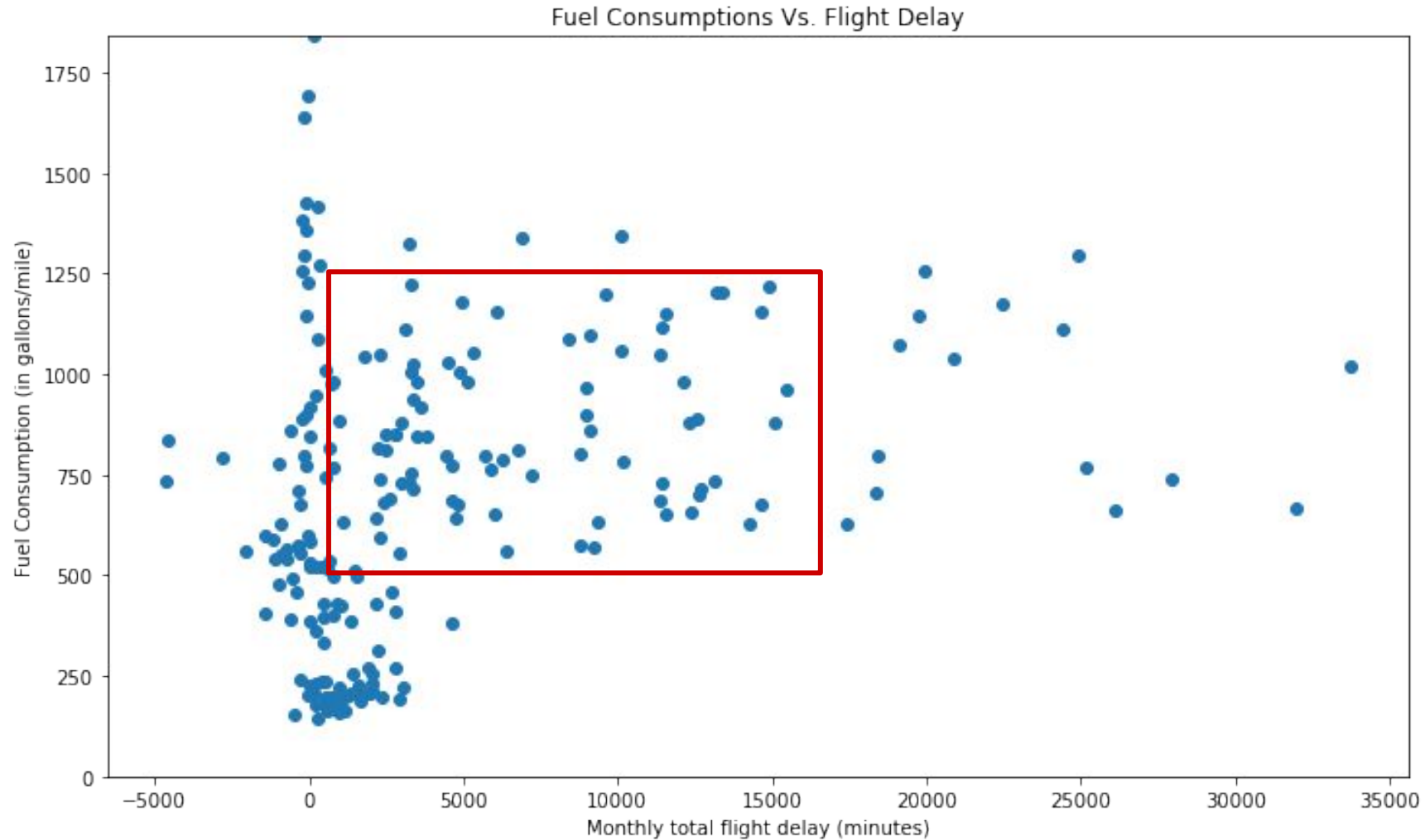
- **Airlines:** crew, fuel, maintenance, passenger accommodation
- **Passengers:** inconvenience, disrupted travel, higher overall flight prices
- **Lost Demand:** passengers opt for other carriers or reliable modes of transportation
- **Indirect:** business travelers, tourism sector
- **Fuel:** as shown in next slide

Total Cost of Delay in the U.S. (dollars, billion)

	2016	2017	2018	2019
Airlines	5.6	6.4	7.7	8.3
Passengers	13.3	14.8	16.4	18.1
Lost Demand	1.8	2.0	2.2	2.4
Indirect	3.0	3.4	3.9	4.2
Total	23.7	26.6	30.2	33.0

Source: US Federal Aviation Administration Report - 2019

Fuel Consumption in delayed flights

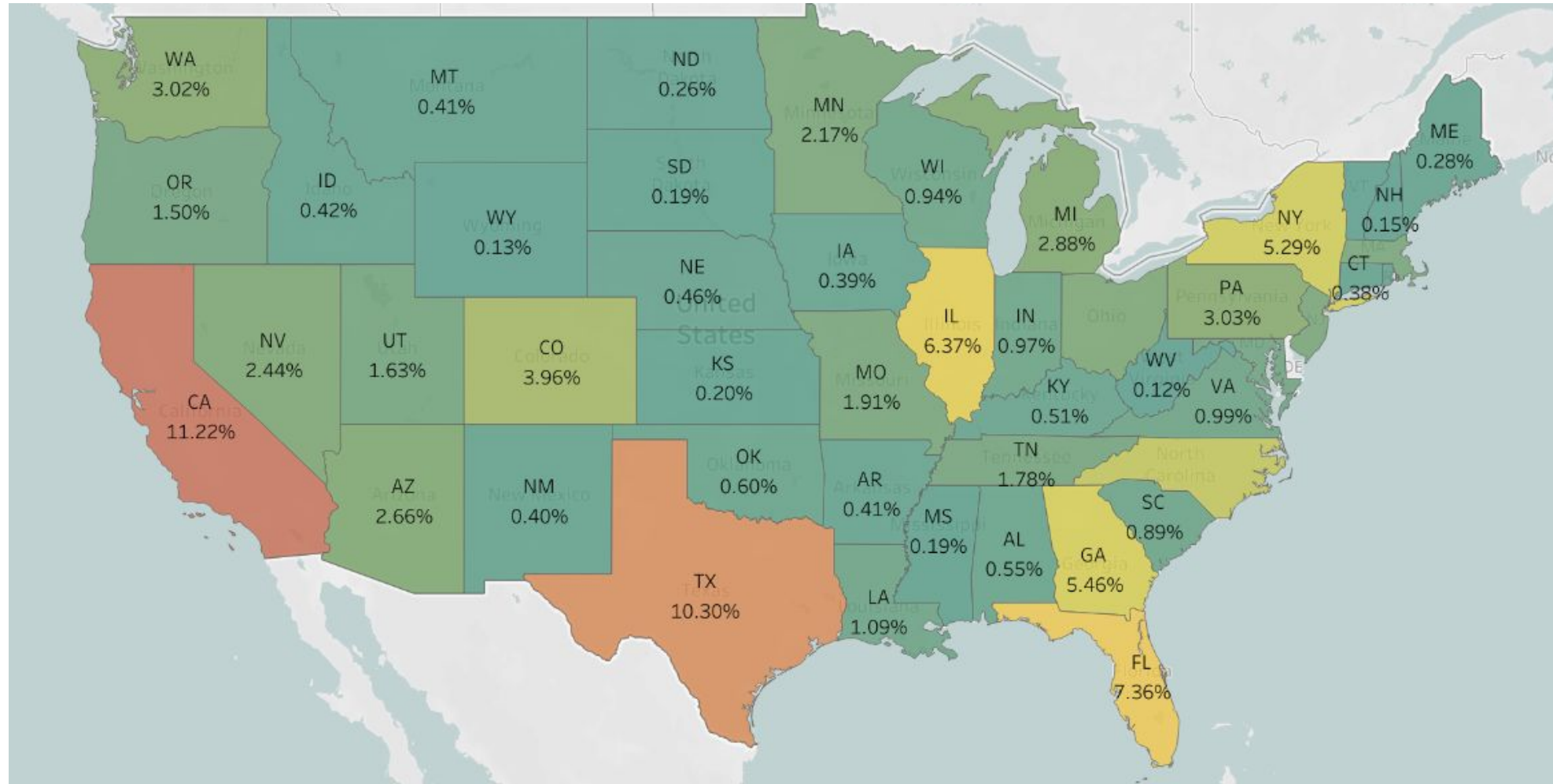


Data and Workflow

- Dataset provided by Lighthouse Labs consisting of US Domestic Flights
- Took a random sample of 200,000 from 2018 and 2019
- Tools used:



EDA: US Air Traffic



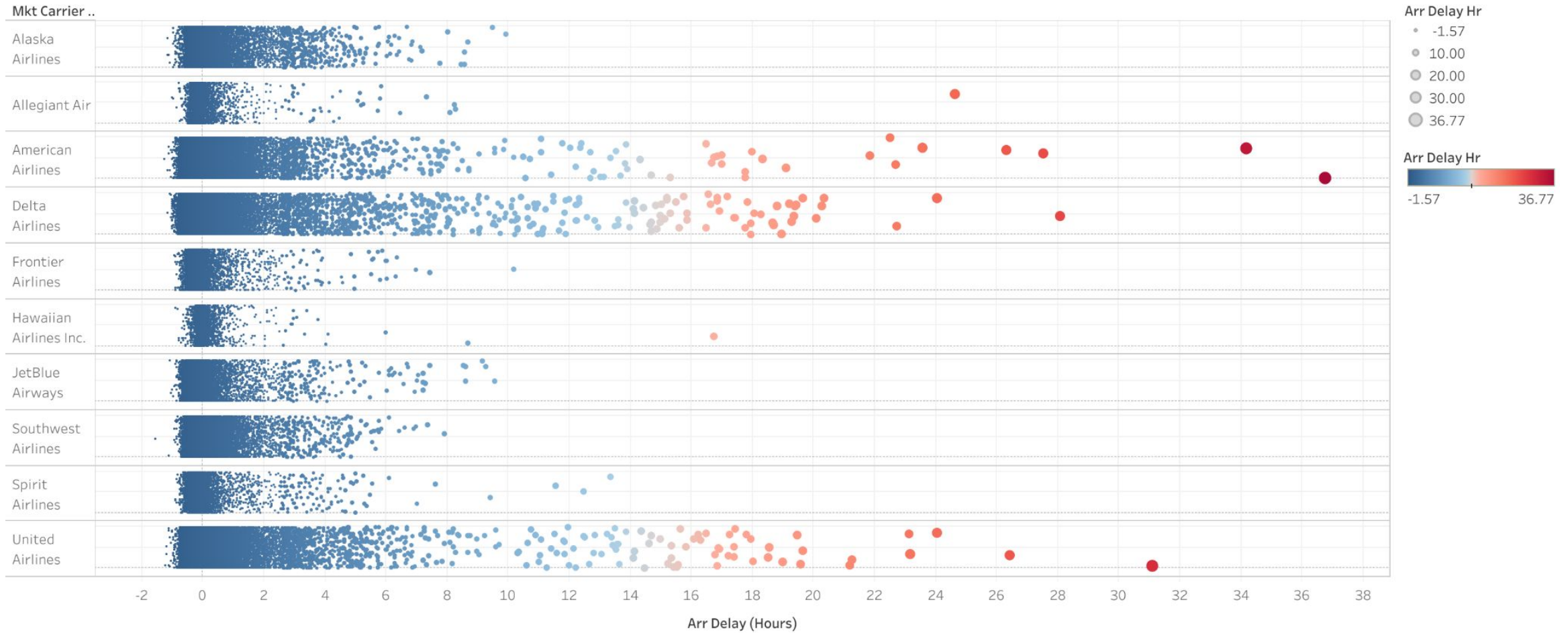
EDA: Monthly Delay Trends

Monthly Flight Delays

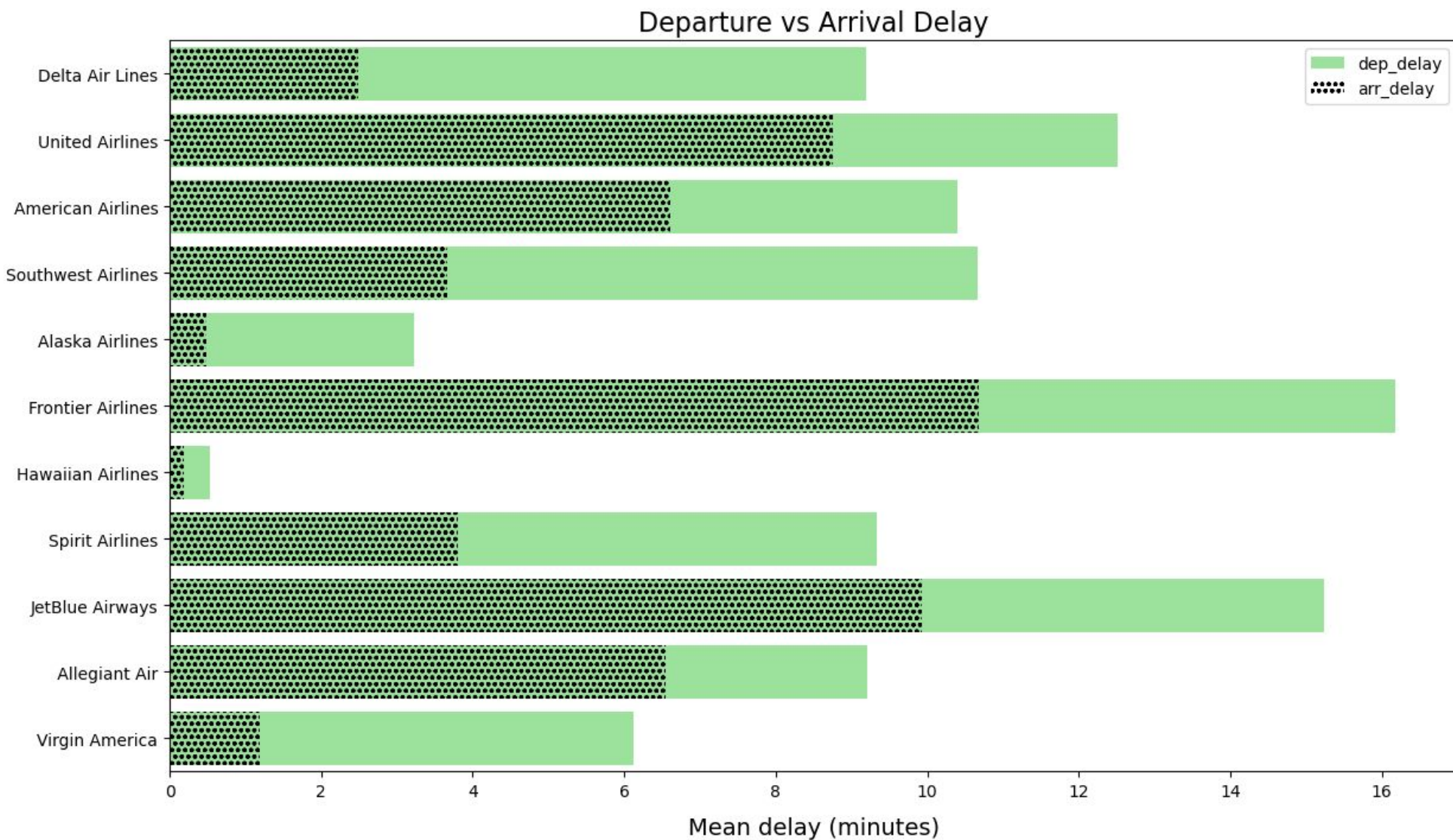


EDA: Carrier Delay Trends

Arrival Delay



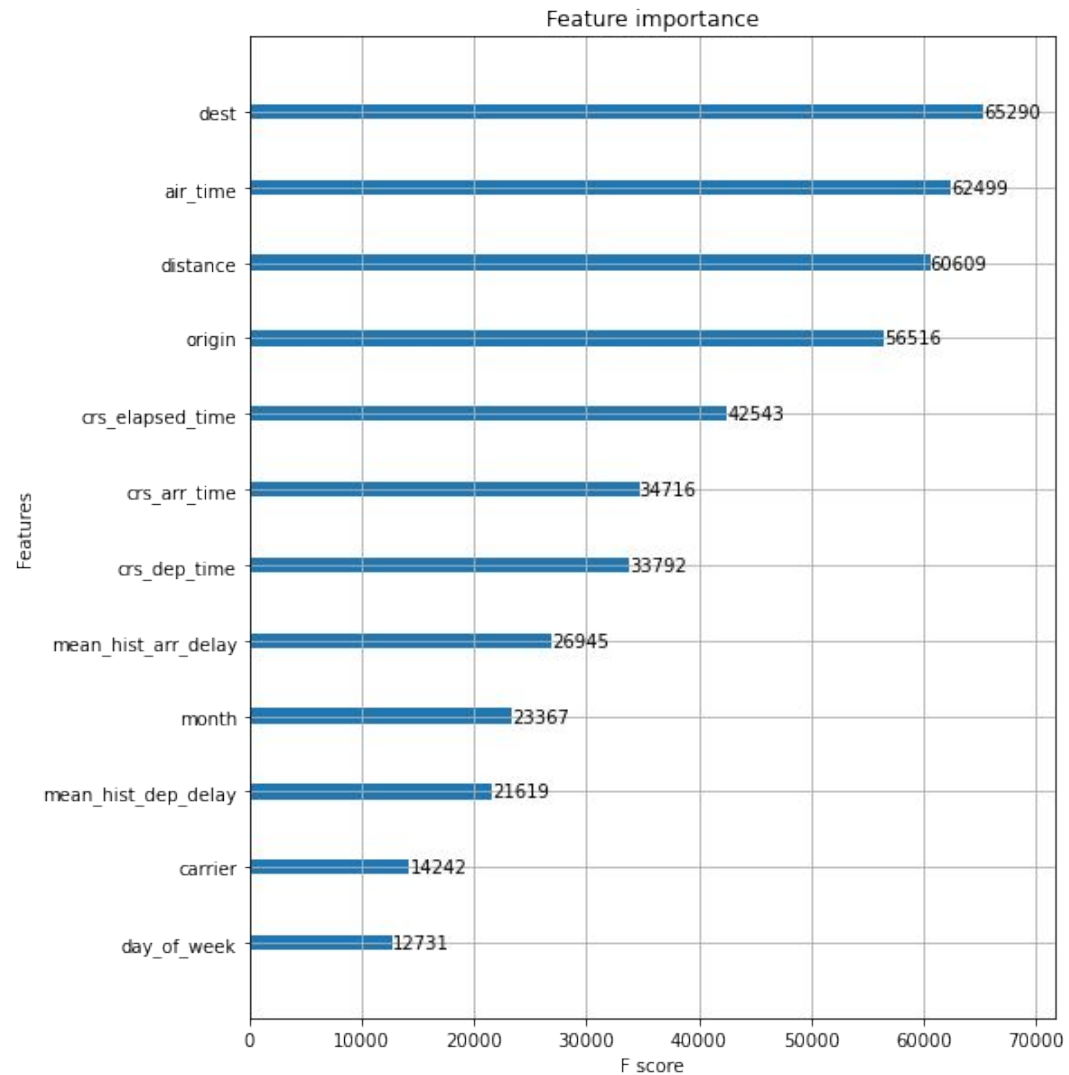
EDA: Delay Trends



Feature Selection

- Origin and Destination
- Departure Time
- Average historical dep & arr delay
- Carrier Encoding
- Month and weekday bins

Feature Importance



Results

		Feature Set 1	Feature Set 2	Feature Set 3
Linear Reg	R2 Score	0.068	0.0765	0.138
	RMSE	47.694	44.47	25.961
Random Forest	R2 score	0.011	0.023	0.129
	RMSE	49.116	50.001	26.106
XGBoost	R2 score	0.804	0.769	0.884
	RMSE	21.832	22.234	9.528

Challenges and Going Forward

- Coordination and planning
- No weather data
- Feature selection
- Limited server/processing power:
 - Use cloud