

PREDICTING FLIGHT DELAYS IN THE U.S.

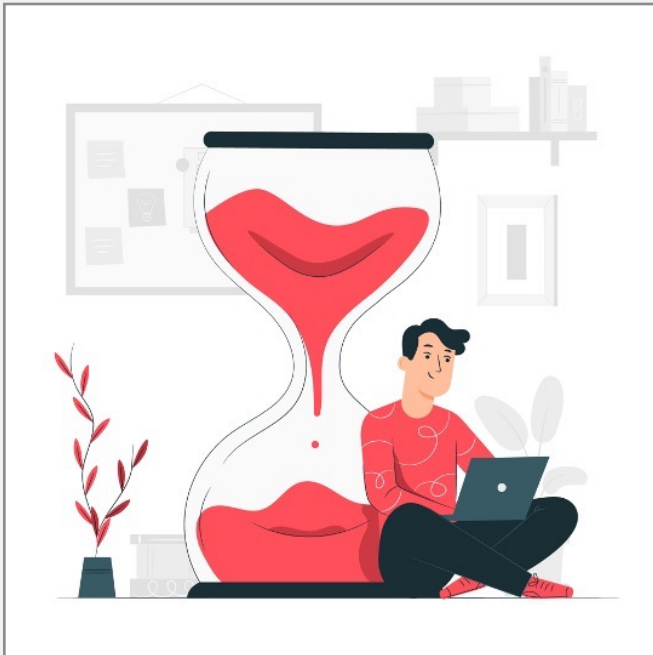


December 11th, 2023 | Final Presentation | 15.072 Advanced Analytics Edge

Fiona Aga, Guillaume Allègre, Jan Philipp Girgott, Theodoti Stoikou

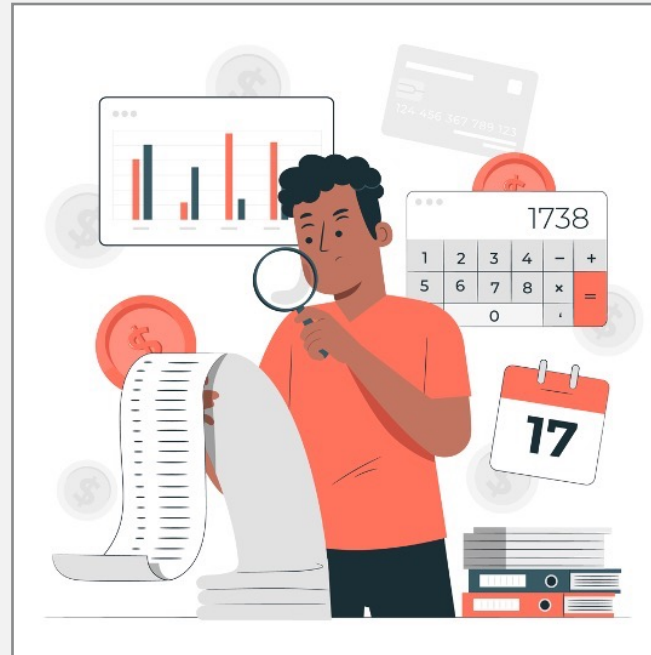
PROBLEM STATEMENT

AIRPLANE DELAYS ARE A PROBLEM FOR THREE MAIN REASONS: WAITING TIME, COST, RESCHEDULING



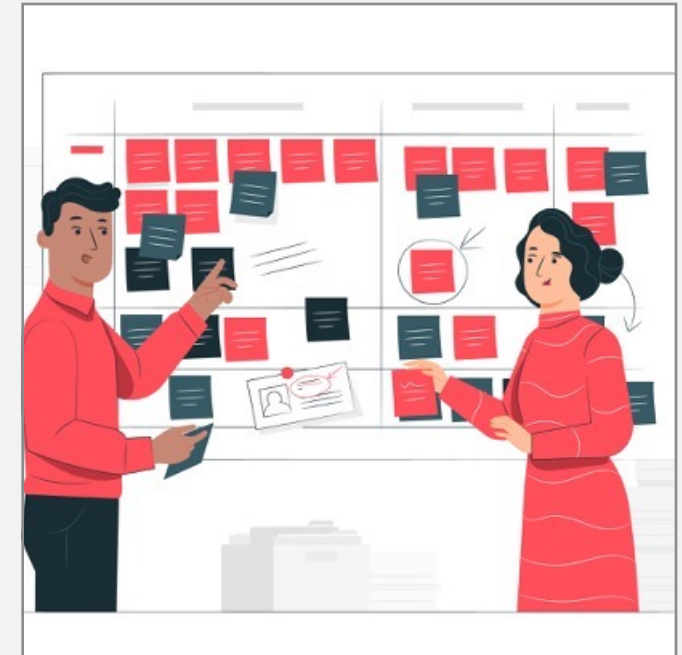
Waiting time

Frustration, late business travelers, cluttered airports



Costs for airlines

Damage payments, hotels, food and drinks



Schedule disruptions

New slots in busy landing/takeoff schedule

DATASET

WE ARE USING A DATASET FROM THE U.S. DOT WHICH WAS ENRICHED FROM VARIOUS OTHER DATA SOURCES



Data Source: U.S. Dep. of Transport

Data Categories



Flight

Date, departure/arrival, delay of previous flight



Airline

#flights per month, #crew members, aircraft details



Weather (Departure Airport)

Precipitation, temperature, maximum wind speeds



Airport

Previous airport, concurrent flights, passengers on this day

APPROACH (1/2)

WE ARE FOCUSING ON THE 3 MOST FREQUENT AIRPORTS AND CARRIERS IN OUR DATASET



Three Most Frequent Airports in Dataset

1. Minneapolis
2. Nashville
3. Houston



Three Most Frequent Carriers in Dataset

1. Southwest
2. Delta
3. SkyWest



1. Binary classification

Will a flight be delayed by more than 15 minutes or not?

2. Multi-class classification

What will be the reason for the delay if it is delayed? (Carrier, Weather, NAS, Security, Late Aircraft)

3. Regression

What will be the magnitude of the delay if it is delayed?

APPROACH (2/2)

WE ARE FOCUSING ON THE 3 MOST FREQUENT AIRPORTS AND CARRIERS IN OUR DATASET

1. Binary classification

Will a flight be delayed by more than 15 minutes or not?

2. Multi-class classification

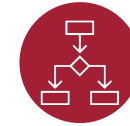
What will be the reason for the delay if it is delayed? (Carrier, Weather, NAS, Security, Late Aircraft)

3. Regression

What will be the magnitude of the delay if it is delayed?



Logistic Regression



CART/OCT



Random Forest

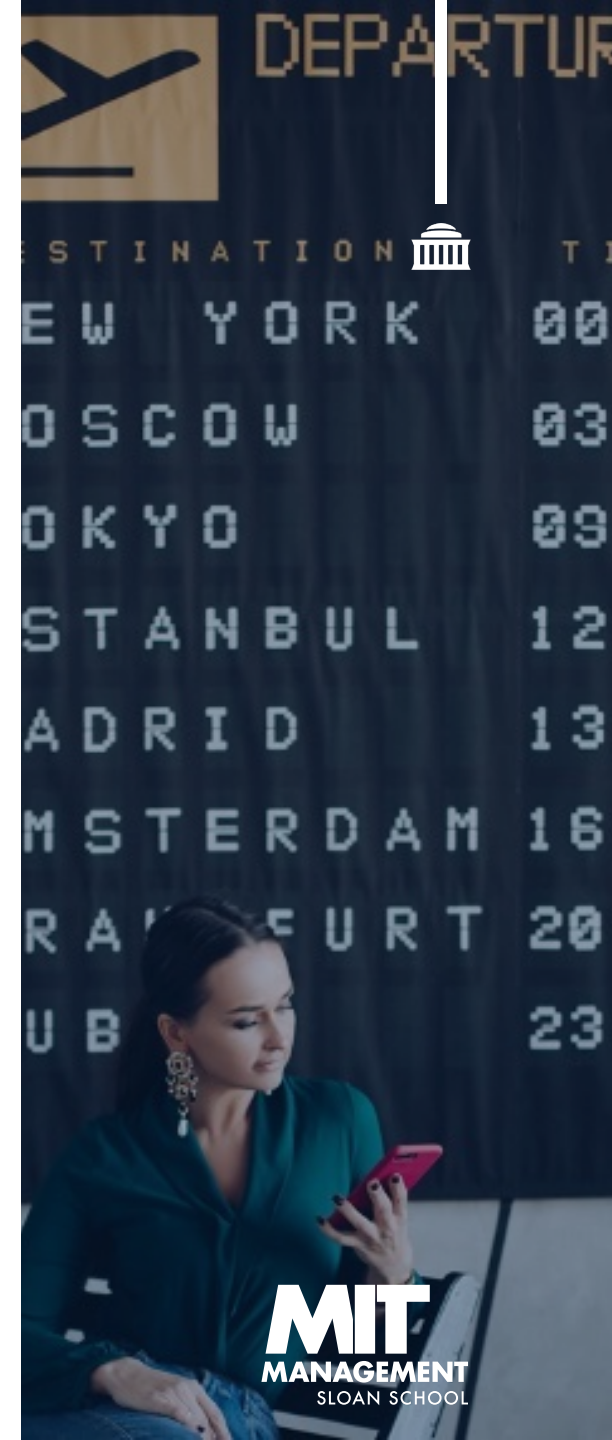


Linear/Partial/Polynomial Regression, Elastic Net, Regression Trees, XGBoost

BINARY CLASSIFICATION RESULTS

THE TREE-BASED METHODS ACHIEVE VERY HIGH AUCS OF UP TO 0.88

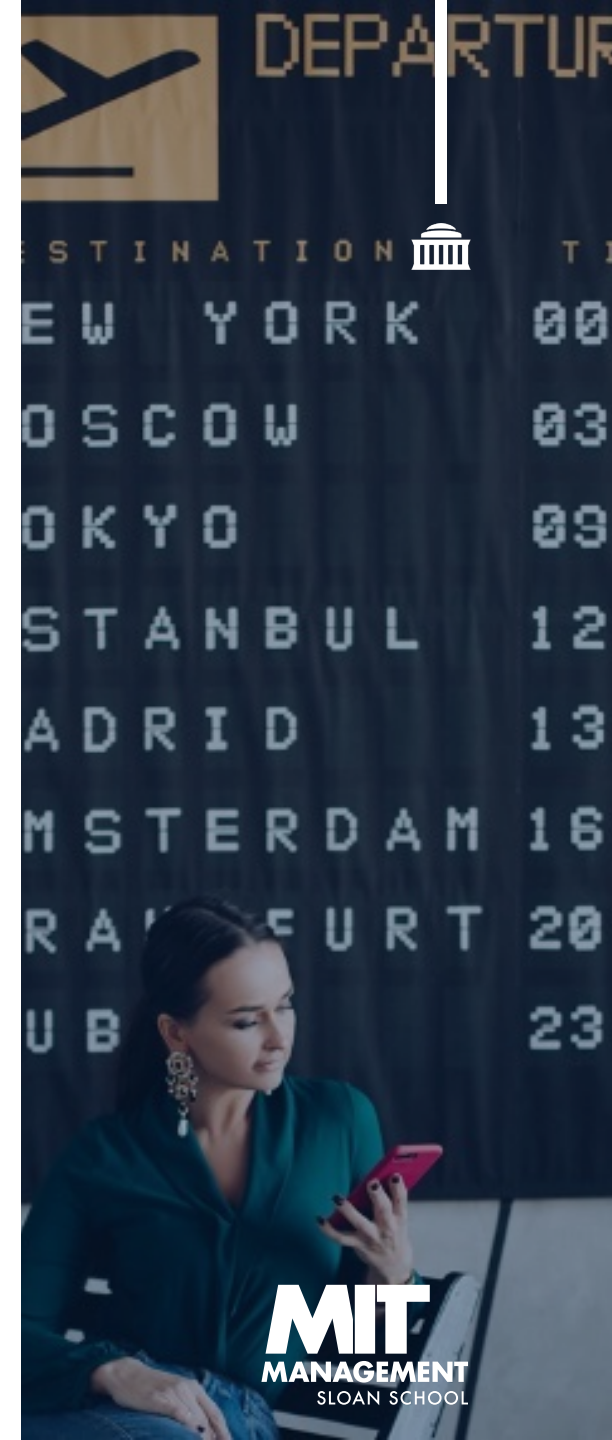
Criterion: AUC		Random Forest	CART	Logistic Regression
Airports	Minneapolis	0.83	0.83	0.80
	Nashville	0.85	0.83	0.83
	Houston	0.88	0.87	0.87
Carriers	Southwest	0.87	0.86	0.87
	Delta	0.81	0.83	0.81
	SkyWest	0.82	0.80	0.80



BINARY CLASSIFICATION RESULTS

THE MOST ACCURATE RESULTS ARE ACHIEVED FOR HOUSTON AND SOUTHWEST

Criterion: AUC		Random Forest	CART	Logistic Regression
Airports	Minneapolis	0.83	0.83	0.80
	Nashville	0.85	0.83	0.83
	Houston	0.88	0.87	0.87
Carriers	Southwest	0.87	0.86	0.87
	Delta	0.81	0.83	0.81
	SkyWest	0.82	0.80	0.80



MULTI-CLASS CLASSIFICATION RESULTS

RANDOM FOREST OUTPERFORMS OTHER
METHOS IN MULTICLASS CLASSIFICATION

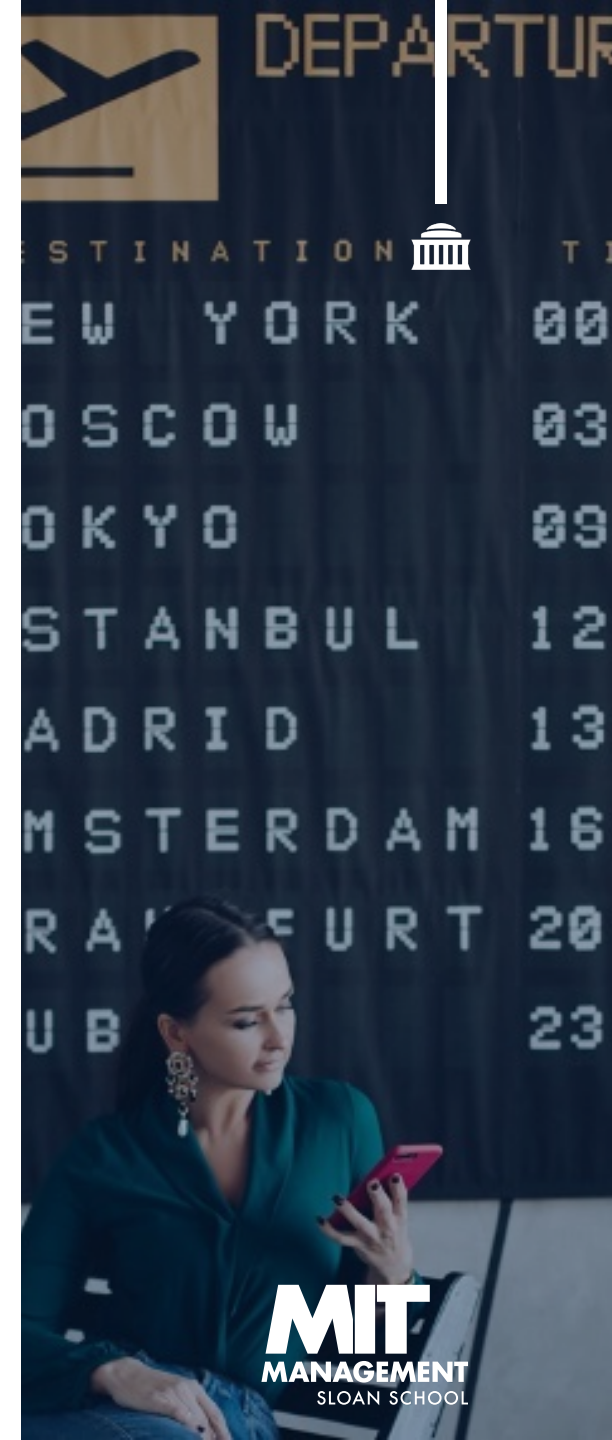
Criterion: Accuracy		Random Forest	CART	Logistic Regression
Airports	Minneapolis	72.0%	68.8%	69.3%
	Nashville	76.7%	72.6%	72.9%
	Houston	76.3%	73.5%	70.8%
Carriers	Southwest	77.1%	74.1%	70.3%
	Delta	74.7%	71.1%	72.1%
	SkyWest	74.9%	72.0%	71.2%



MULTI-CLASS CLASSIFICATION RESULTS

THE MOST ACCURATE RESULTS ARE ACHIEVED FOR NASHVILLE AND SOUTHWEST

Criterion: Accuracy		Random Forest	CART	Logistic Regression
Airports	Minneapolis	72.0%	68.8%	69.3%
	Nashville	76.7%	72.6%	72.9%
	Houston	76.3%	73.5%	70.8%
Carriers	Southwest	77.1%	74.1%	70.3%
	Delta	74.7%	71.1%	72.1%
	SkyWest	74.9%	72.0%	71.2%



FACTORS INFLUENCING DELAYS: AIRPORTS

WEATHER, TRAFFIC, PREVIOUS DELAYS AND SPECIFIC AIRLINES HAVE LARGEST EFFECTS



Common factors



Weather: maximum temperature, average wind speed, precipitation



Traffic: concurrent flights, airline flights during this month



Cascading effect: delays of previous flights on the plane

Airports-specific airline ranking

Minneapolis

Top Punctual: Endeavor Air, SkyWest Airlines

Medium Punctual: Delta, United Airlines

Late: American, Alaska, Midwest Airlines

Nashville

Top Punctual: Spirit, Delta, Midwest Airlines

Medium Punctual: Mesa Airlines, SkyWest Airlines

Late: JetBlue, Alaska, Atlantic Southeast

Houston

Top Punctual: Delta, Midwest Airlines

Medium Punctual: Mesa Airlines, JetBlue

Late: Southwest, SkyWest, Atlantic Southeast

FACTORS INFLUENCING DELAYS: AIRPORTS

SAME COMMON FACTORS FOR AIRLINES EVEN THOUGH SPECIFIC AIRPORTS CAUSE DELAYS



Common factors



Weather: maximum temperature, average wind speed, precipitation this month

SAME AS BEFORE



Cascading effect: delays of previous flights on the plane

Airline-specific airport ranking

Southwest

Top Punctual: Rhode Island, Boise

Medium Punctual: Charleston, Albuquerque

Late: Nashville, Houston

Delta

Top Punctual: Greensboro, Norfolk, Greenville

Medium Punctual: Penascola, Birmingham

Late: Minneapolis, Palm Springs

SkyWest

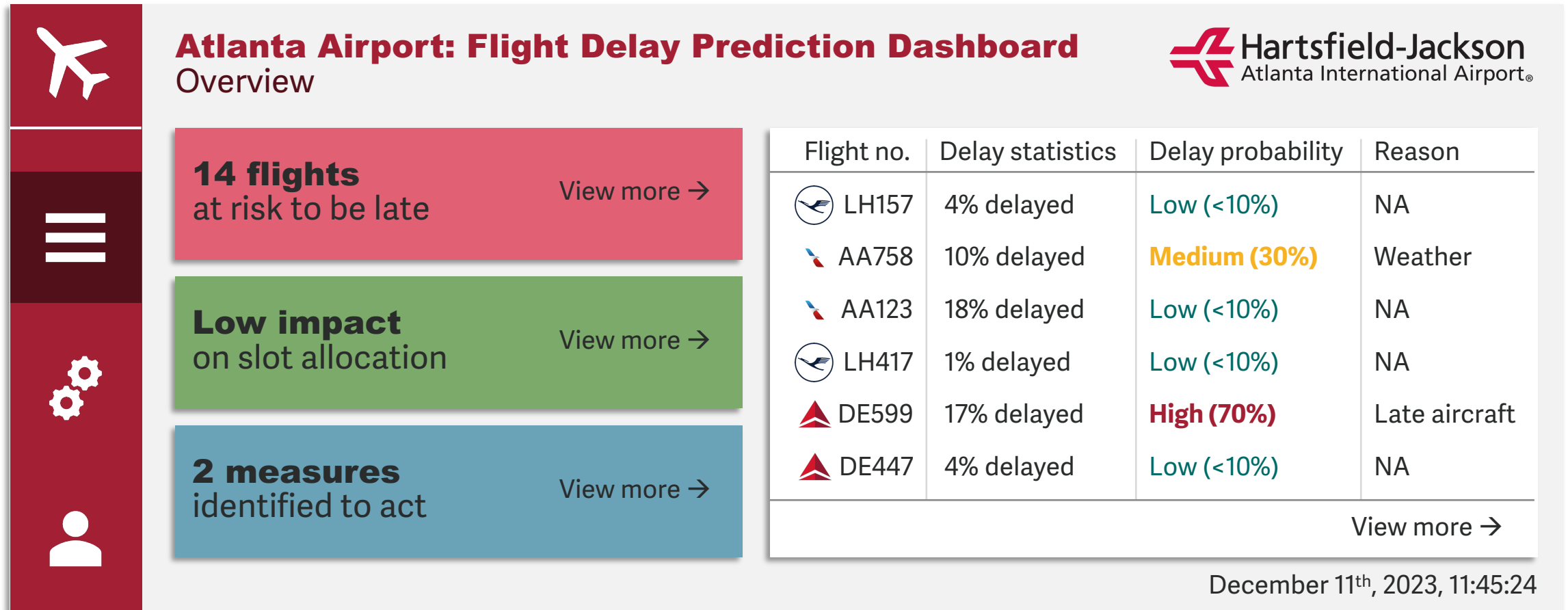
Top Punctual: Albuquerque, Boise, Minneapolis

Medium Punctual: Syracuse, Reno/Tahoe

Late: Birmingham, Houston, Richmond

ACTIONABLE INSIGHTS: DASHBOARD

AN INTERACTIVE DASHBOARD CAN PROVIDE VALUE TO AIR TRAFFIC CONTROL, CARRIERS, AND PASSENGERS



POTENTIAL COUNTERMEASURES

ACTIONS BY BOTH AIRPORTS AND CARRIERS ARE NECESSARY TO PREVENT DELAYS



ACTION ITEMS FOR AIRPORTS

- 1** **Adjust gate assignments** so that certain terminal parts do not become too crowded
- 2** **Reserve fall back departure slots** if there are many possibly delayed flights during a period
- 3** Take **preventative measures** in case of **bad weather events** (e.g., heat the runway to prevent ice)

ACTION ITEMS FOR CARRIERS

- 1** Staff **more experienced crews** on flights that are usually late to **avoid operational causes**
- 2** Use **new equipment on flights** which are usually late to **avoid equipment failures**
- 3** **Optimize padding time** between flights to **avoid cost of exceeding the take-off slot**

Launch joint improvement programs to find root causes of delays!

PREDICTING FLIGHT DELAYS IN THE U.S.



December 11th, 2023 | Final Presentation | 15.072 Advanced Analytics Edge

Fiona Aga, Guillaume Allègre, Jan Philipp Girgott, Theodoti Stoikou