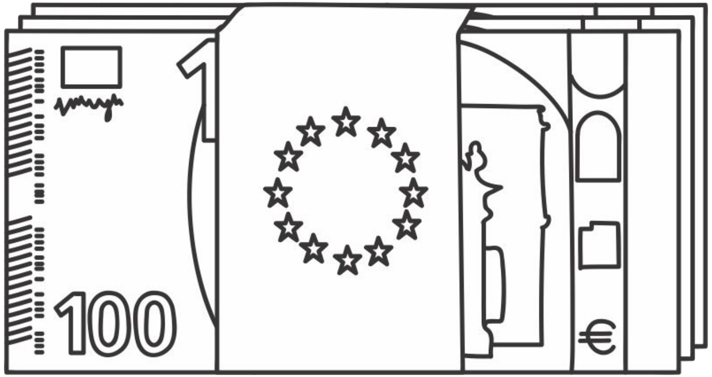
An Investigation into

Flight Delays

Rachael Alexandroff, Sofia Pignataro, Racquel Fygenson, Ruxin Shen
Group 13

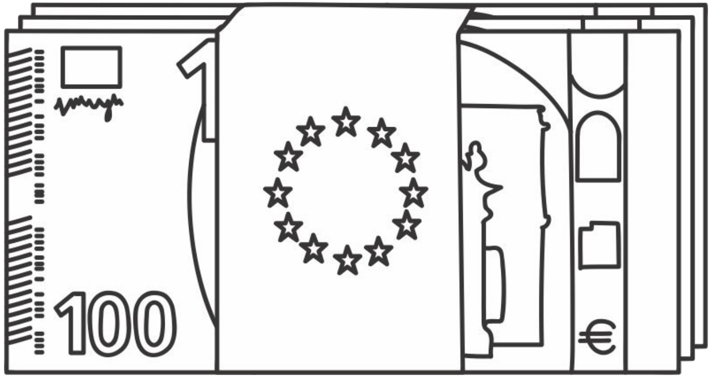
Flight delays are a hassle.

Flight delays are a hassle



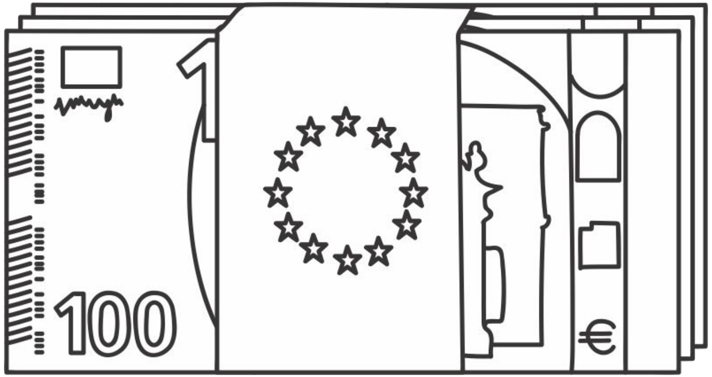
EU regulation 261/2004 requires **airlines to give you money** if your flight is delayed > 3 hours!

Flight delays are a hassle

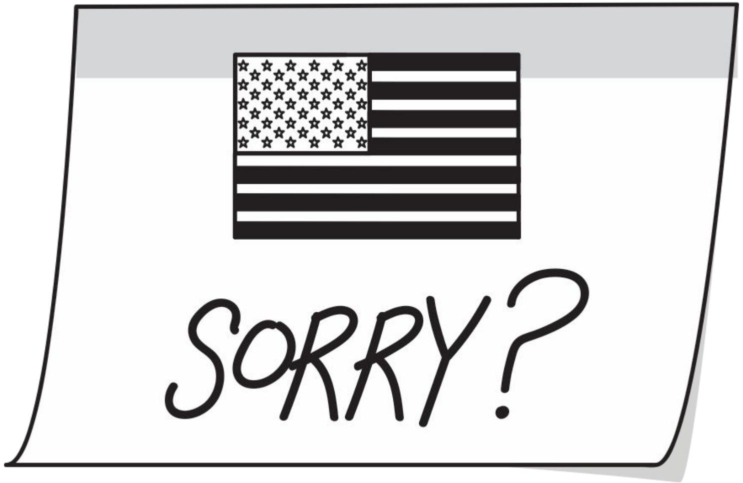


What about in the US?

Flight delays are a hassle



VS

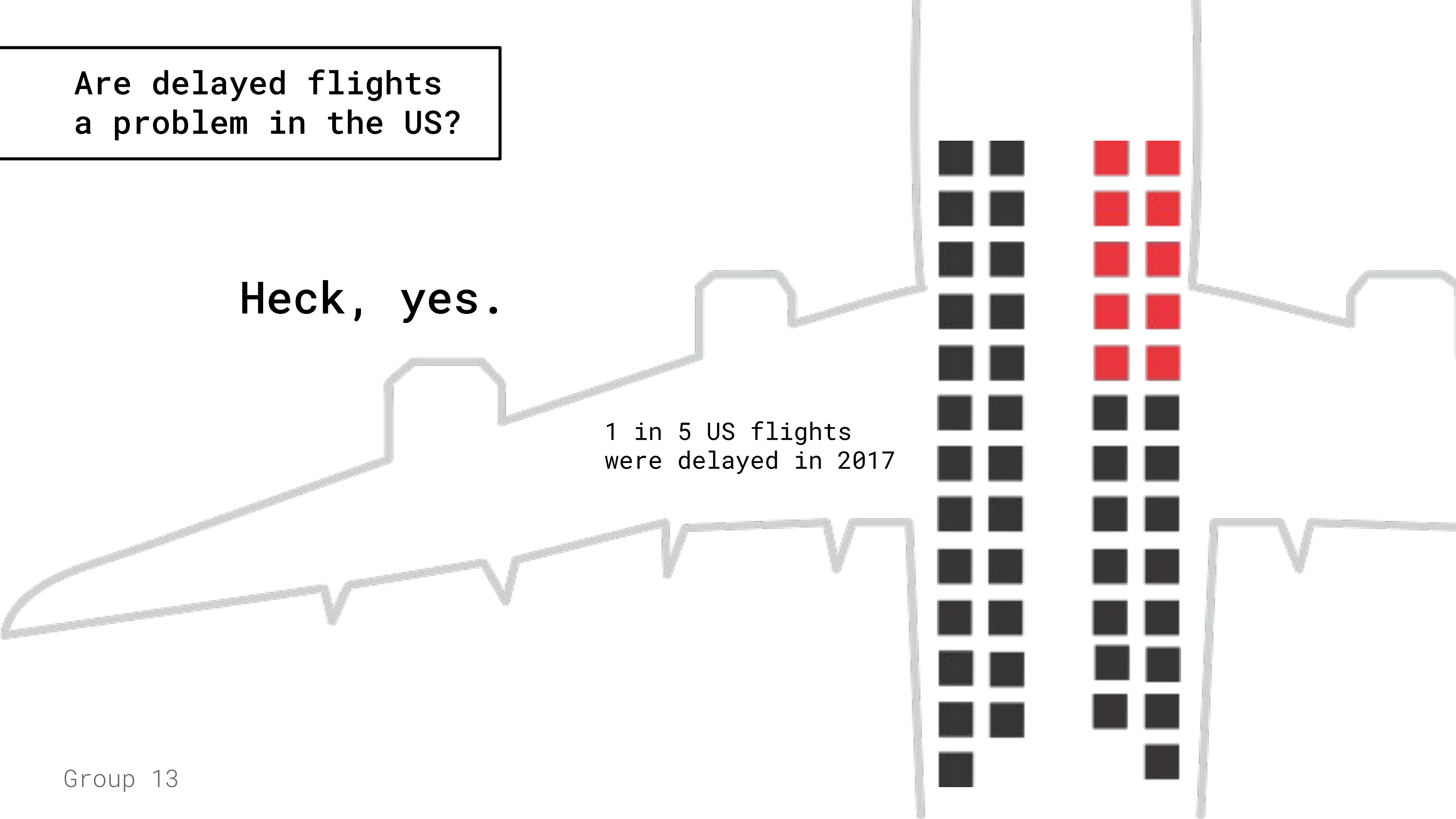


Flight delays are a hassle

Are delayed flights
a problem in the US?

Are delayed flights a problem in the US?

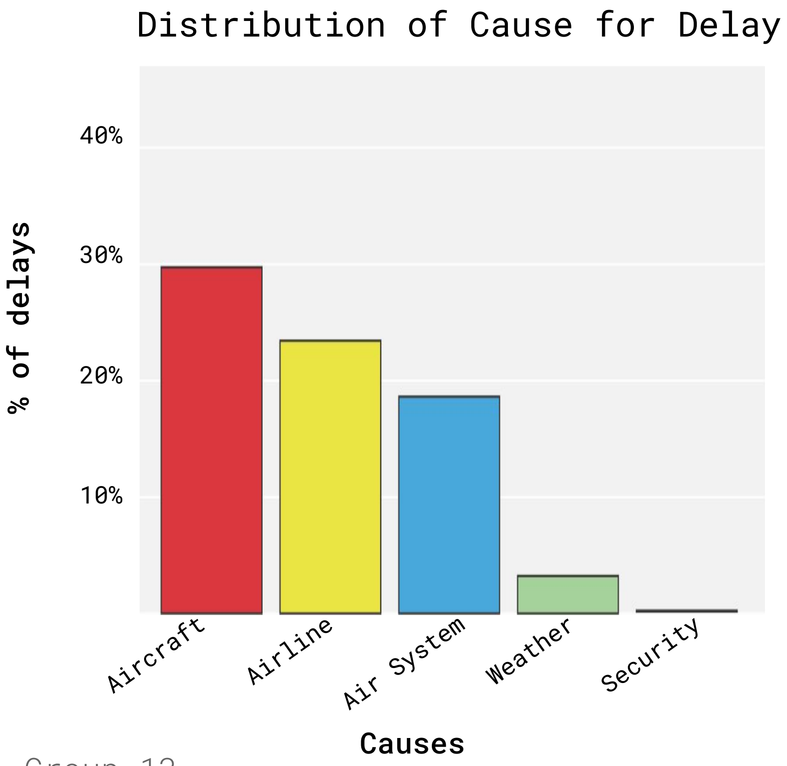
Heck, yes.



1 in 5 US flights
were delayed in 2017

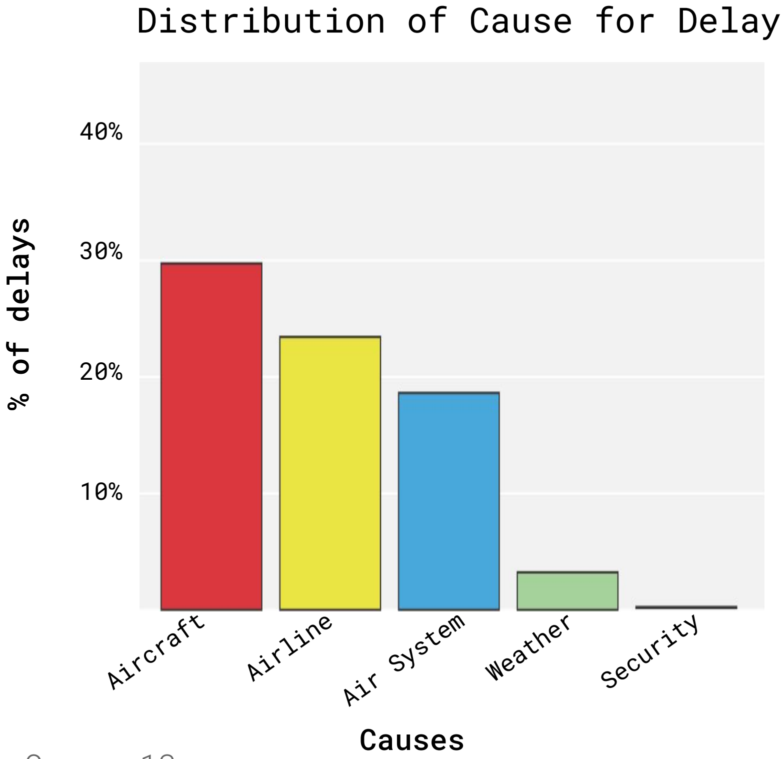
What factors correlate with delay?

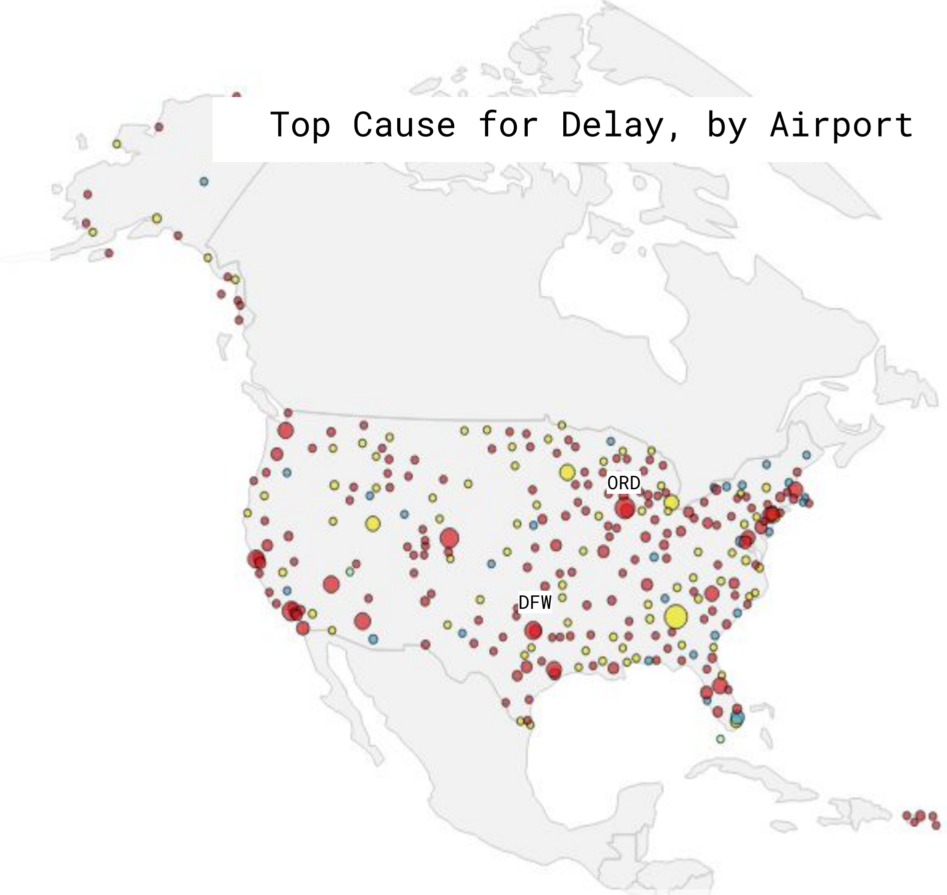
What factors affect delay?



Group 13

What factors affect delay?

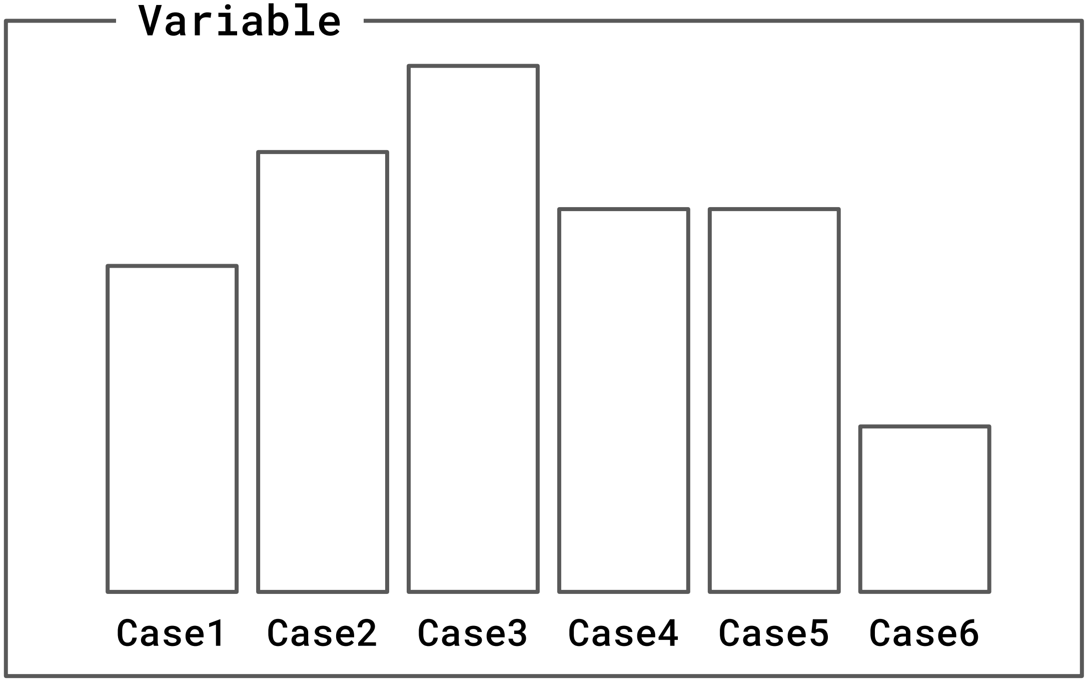




Weather database: missing airports for important cities, e.g. Chicago (ORD) and Dallas (DFW)

What factors affect delay?

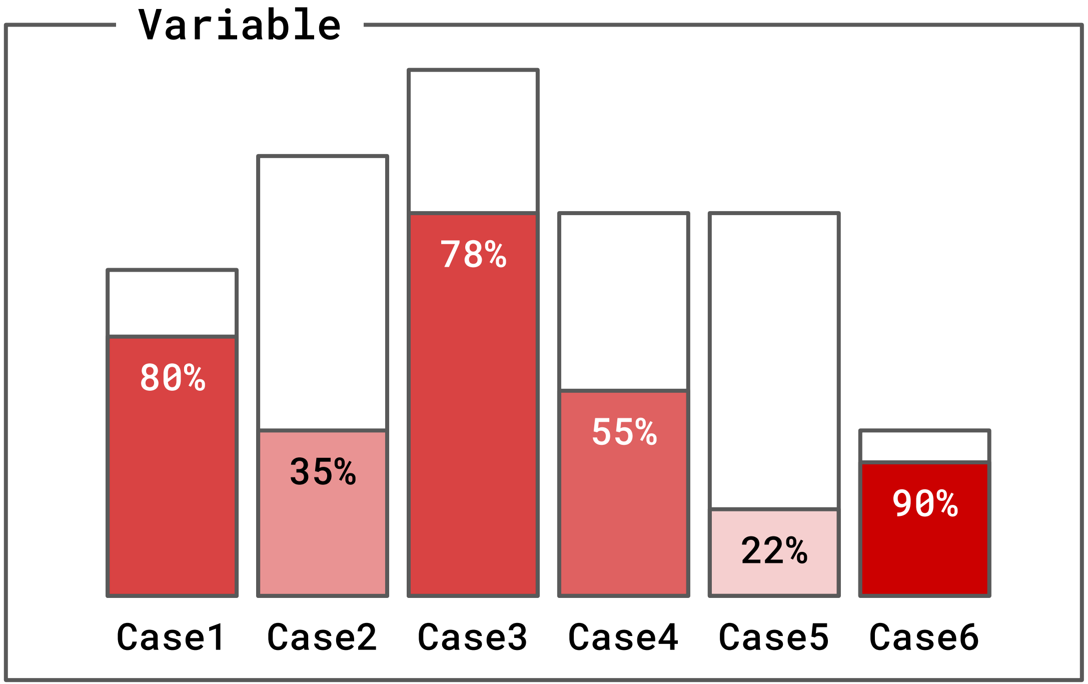
Proportional Summary*



Let's **separate by variables** we think might be relevant

What factors affect delay?

Proportional Summary*



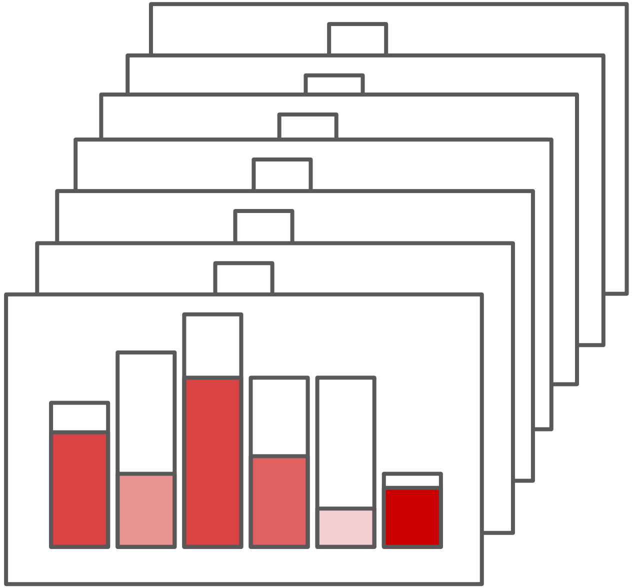
Let's **separate by variables** we think might be relevant

and look at the **proportion** of flights that are delayed

Group 13

What factors affect delay?

Proportional Summary*

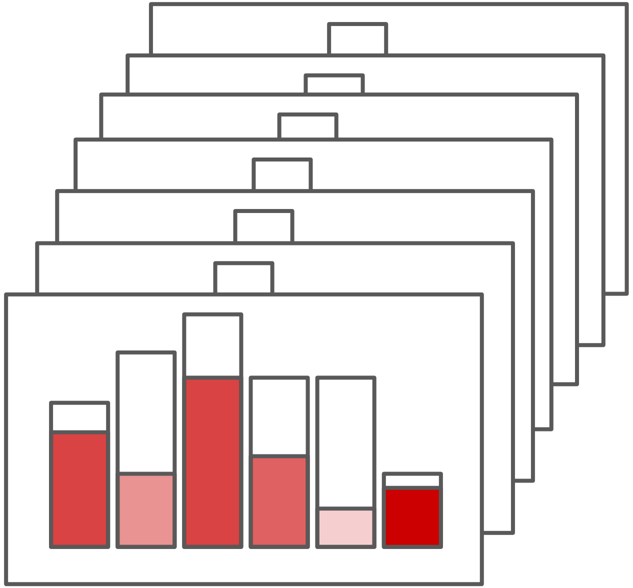


Variables we evaluated

- ➔ Day of the week
- ➔ Month of the year
- ➔ Time of day (4 buckets)
- ➔ Elapsed flight time
- ➔ Distance of flight
- ➔ Airlines
- ➔ Season

What factors affect delay?

Proportional Summary*



Variables we evaluated

- → [✓ Day of the week]
- → Month of the year
- → [✓ Time of day (4 buckets)]
- → Elapsed flight time
- → Distance of flight
- → [✓ Airlines (2 buckets)]
- → [✓ Season]

What factors affect delay?

Proportional Summary*

Data Used

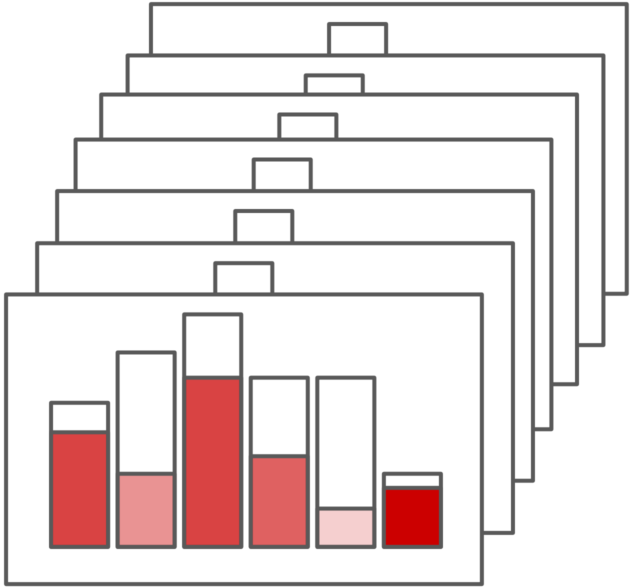
Flight Traffic

Data Not Used

Weather
Fare
Event

Outside scope of question → Fare

Retain important data points that were not represented in these dataframes → Weather, Event



Variables we evaluated

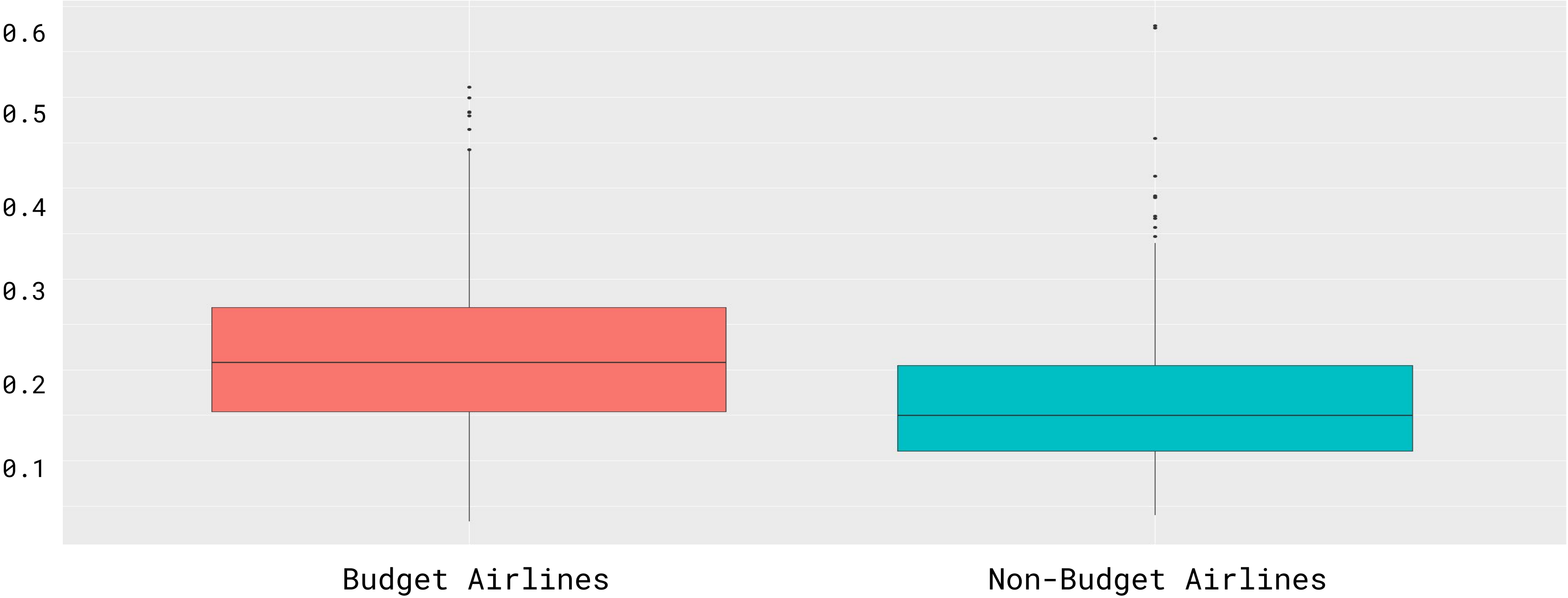
- ➔ Day of the week
- ➔ Month of the year
- ➔ Time of day (4 buckets)
- ➔ Elapsed flight time
- ➔ Distance of flight
- ➔ Airlines (2 buckets)
- ➔ Season

Group 13

What factors affect delay?

Hypothesis Testing

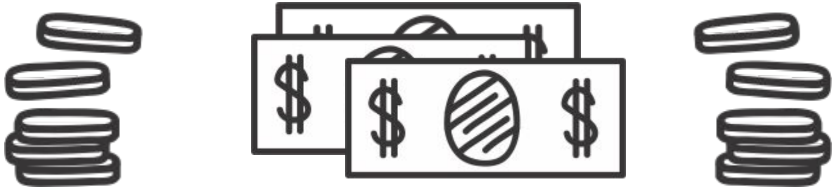
Airline Type



What factors affect delay?

Hypothesis Testing

Airline Type



	Budget
Not Budget	S

S	Significant
NS	Not Significant

$H_0 =$ Proportion of delayed flights are equal

$H_a =$ Proportion of delayed flights are not equal

“Budget” Airlines:

- ➔ Spirit
- ➔ JetBlue
- ➔ ExpressJet
- ➔ Frontier
- ➔ SkyWest
- ➔ Southwest
- ➔ Virgin

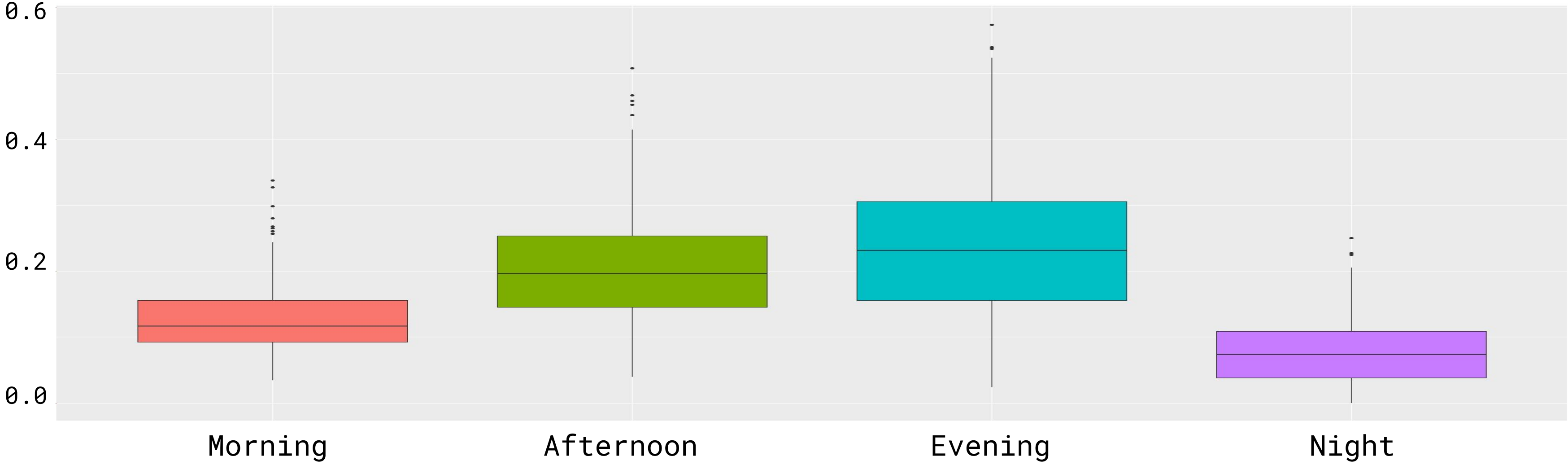
“Non-Budget” Airlines:

- ➔ American
- ➔ Delta
- ➔ Hawaiian Air
- ➔ United
- ➔ Alaska Air

What factors affect delay?

Hypothesis Testing

Time of Day



What factors affect delay?

Hypothesis Testing

Time of Day

	Morning	Afternoon	Evening
Afternoon	S		
Evening	S	S	
Night	S	S	S

H_0 = Proportion of delayed flights are equal

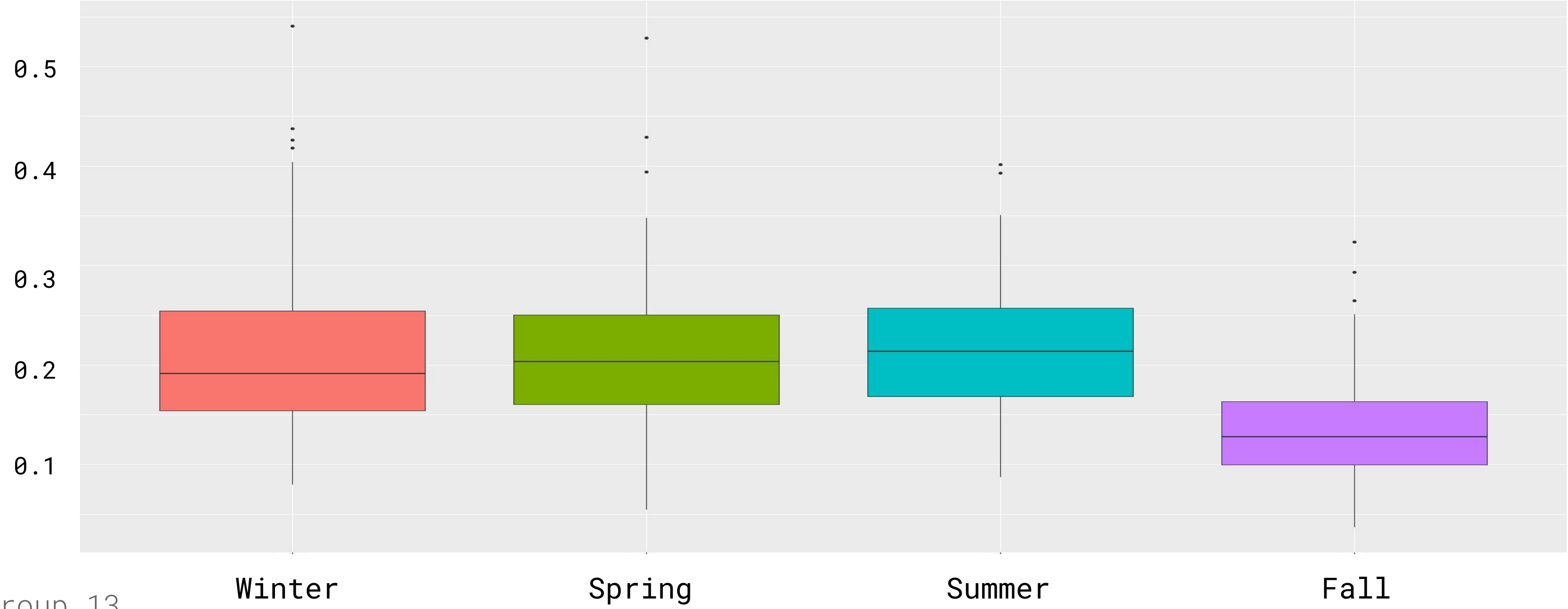
H_a = Proportion of delayed flights are not equal

S	Significant
NS	Not Significant

What factors affect delay?

Hypothesis Testing

Season



What factors affect delay?

Hypothesis Testing

Season

	Spring	Summer	Autumn
Summer	NS		
Autumn	S	S	
Winter	NS	NS	S

H_0 = Proportion of delayed flights are equal

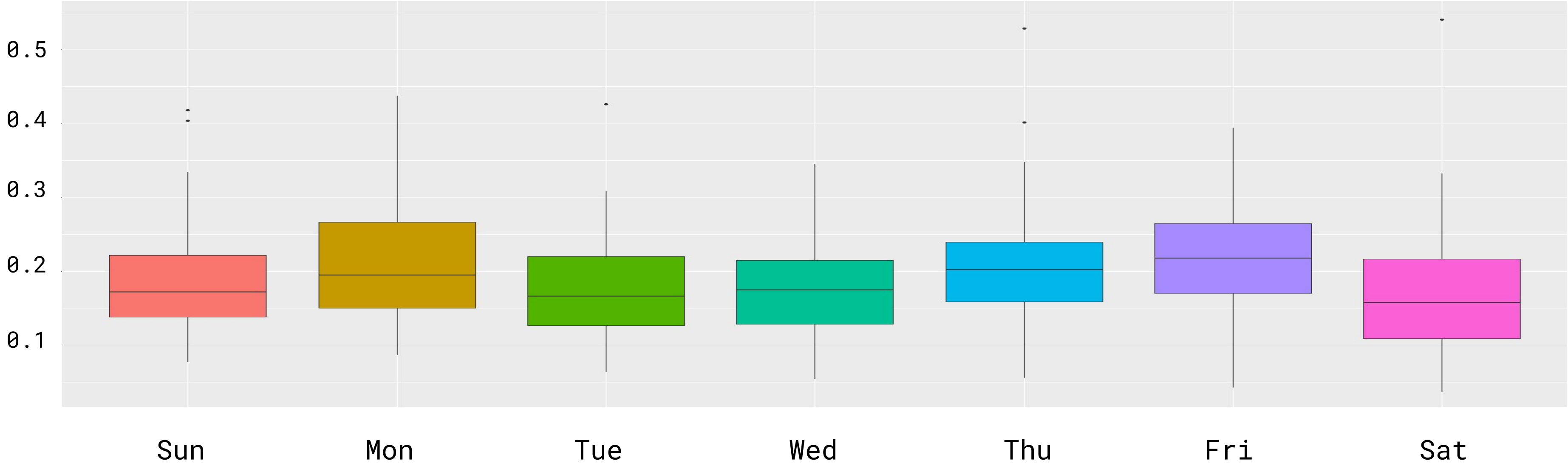
H_a = Proportion of delayed flights are not equal

S	Significant
NS	Not Significant

What factors affect delay?

Hypothesis Testing

Day of the Week



What factors affect delay?

Hypothesis Testing

Day of the Week

	Sun	Mon	Tue	Wed	Thu	Fri
Mon	S					
Tue	NS	S				
Wed	NS	S	NS			
Thu	S	NS	S	S		
Fri	S	NS	S	S	S	
Sat	S	S	S	S	S	S

H_0 = Proportion of delayed flights are equal

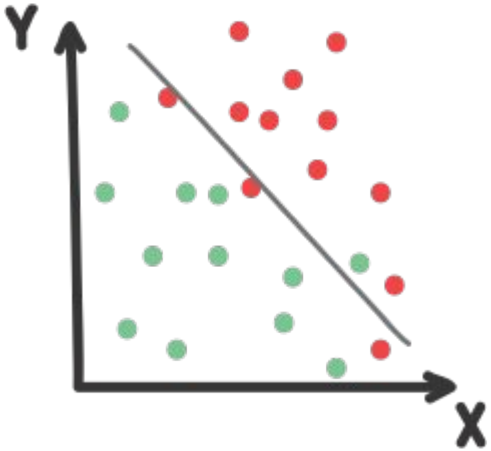
H_a = Proportion of delayed flights are not equal

S	Significant
NS	Not Significant

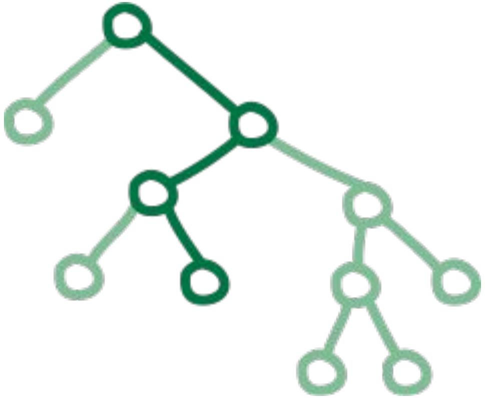
Can we predict delay?

Machine Learning

Benchmark:
Logistic Regression



Modeling:
Random Forest

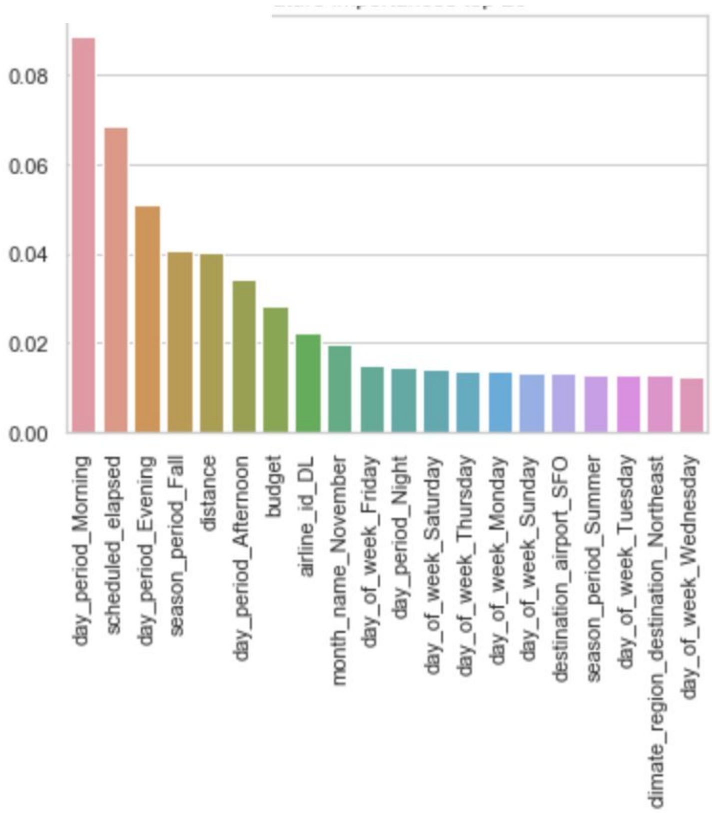


Can we predict instance of delay?

Random Forest

Delay: Y/N?

Features



Group 13

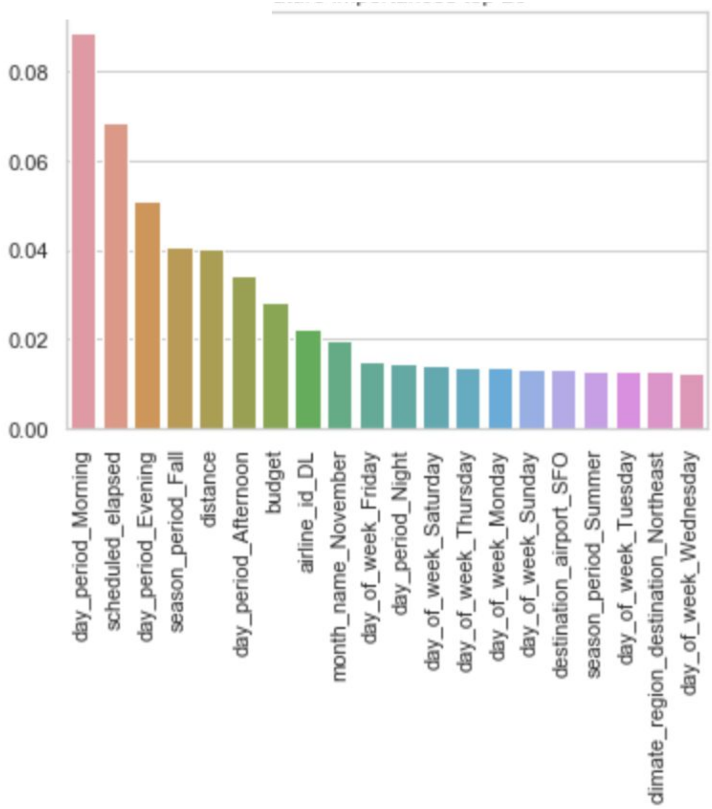
Can we predict instance of delay?

Random Forest

Delay: Y/N?

Time of day is an important feature

Features



Can we predict instance of delay?

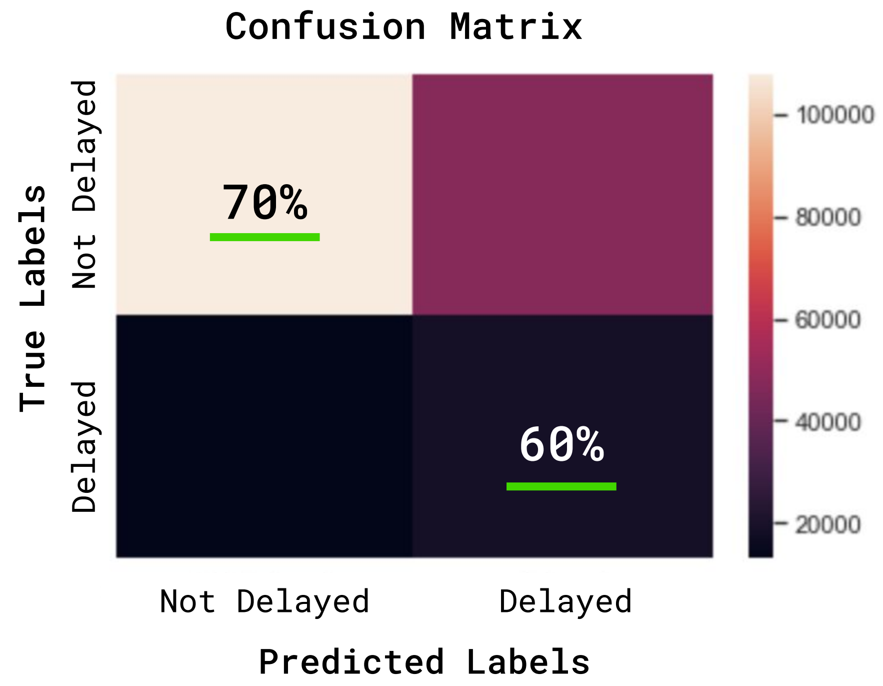
Random Forest

Delay: Y/N?

Right now, our model has a lot of false positives

Overall Accuracy: 68%

Recall



Can we predict instance of delay?

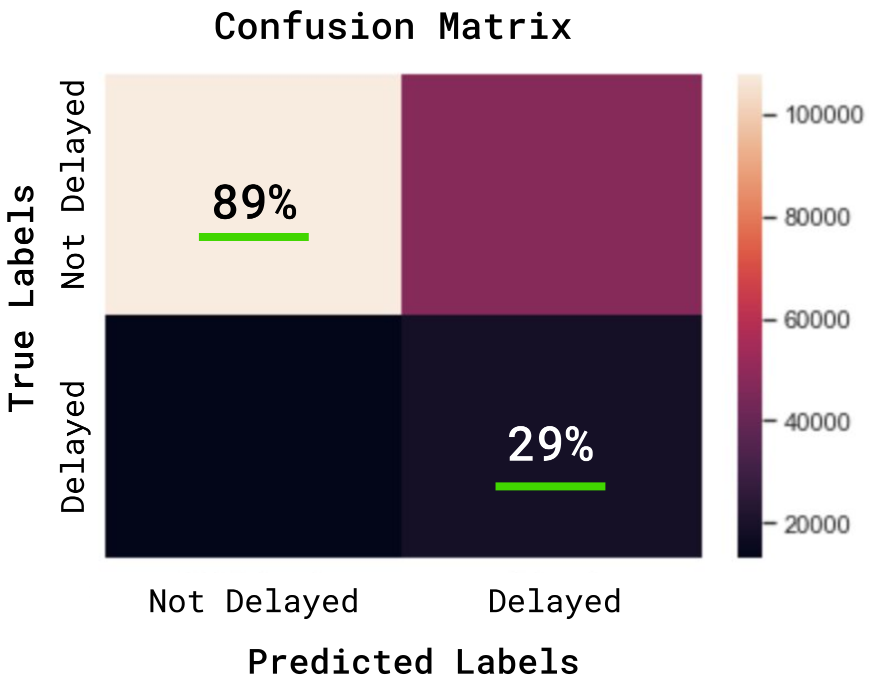
Random Forest

Delay: Y/N?

Right now, our model has a lot of false positives

Overall Accuracy: 68%

Precision



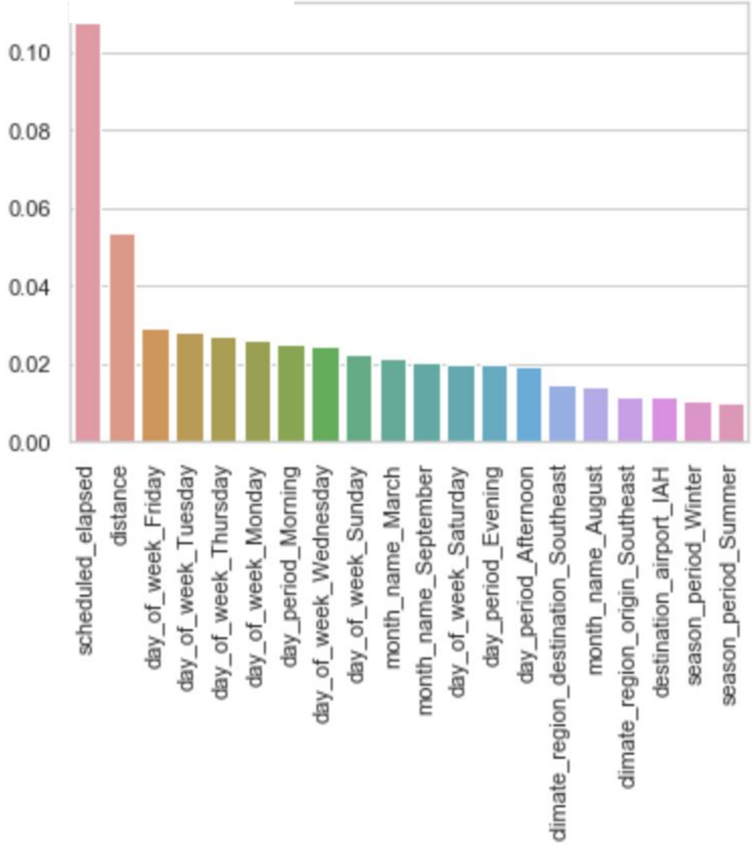
Can we predict length of delay?

Random Forest

Length of Delay?

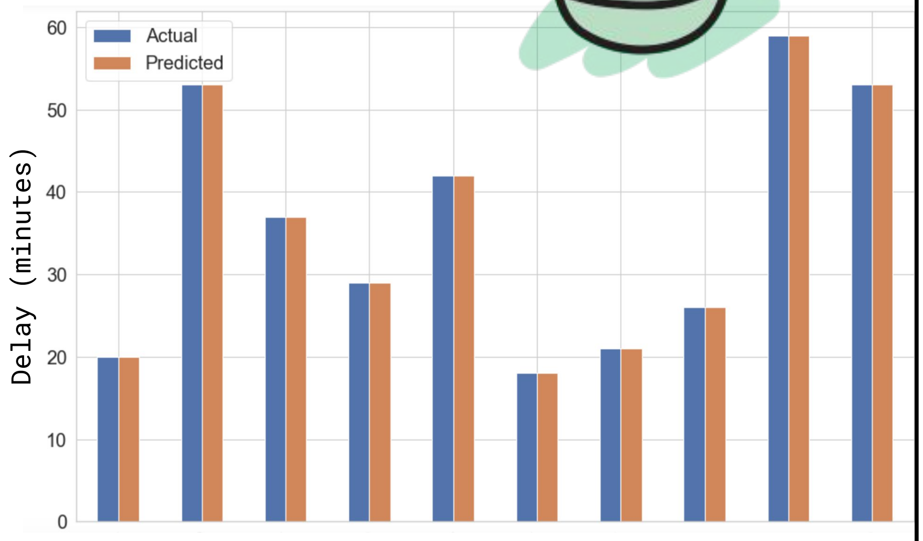
Flight **duration**,
flight **distance**,
day of week,
and **month** are
important features

Features



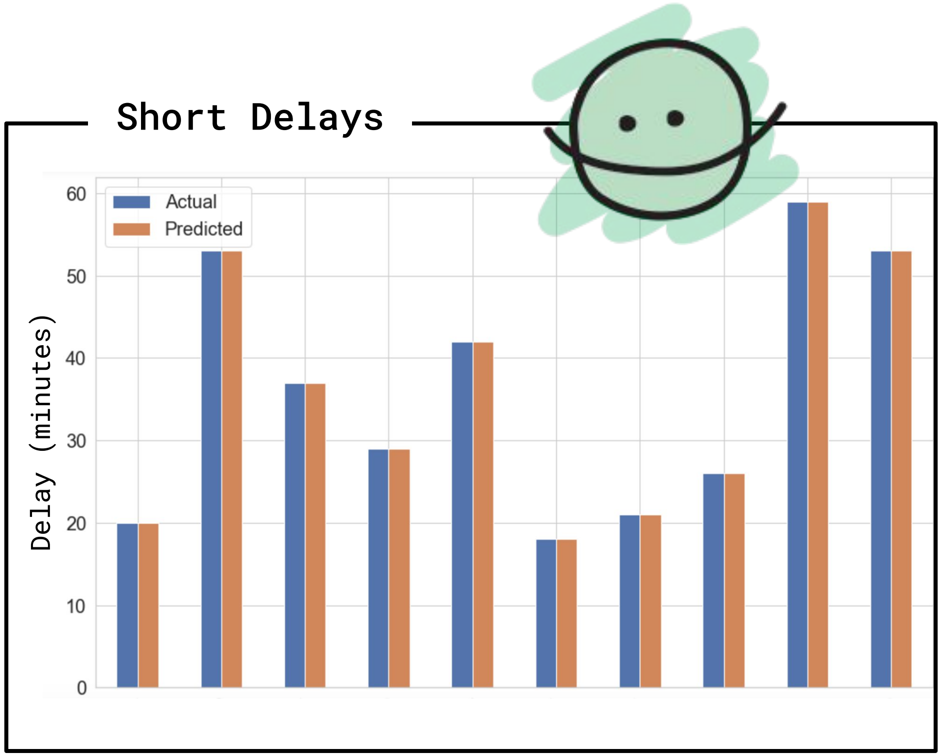
Can we predict length of delay?

Short Delays

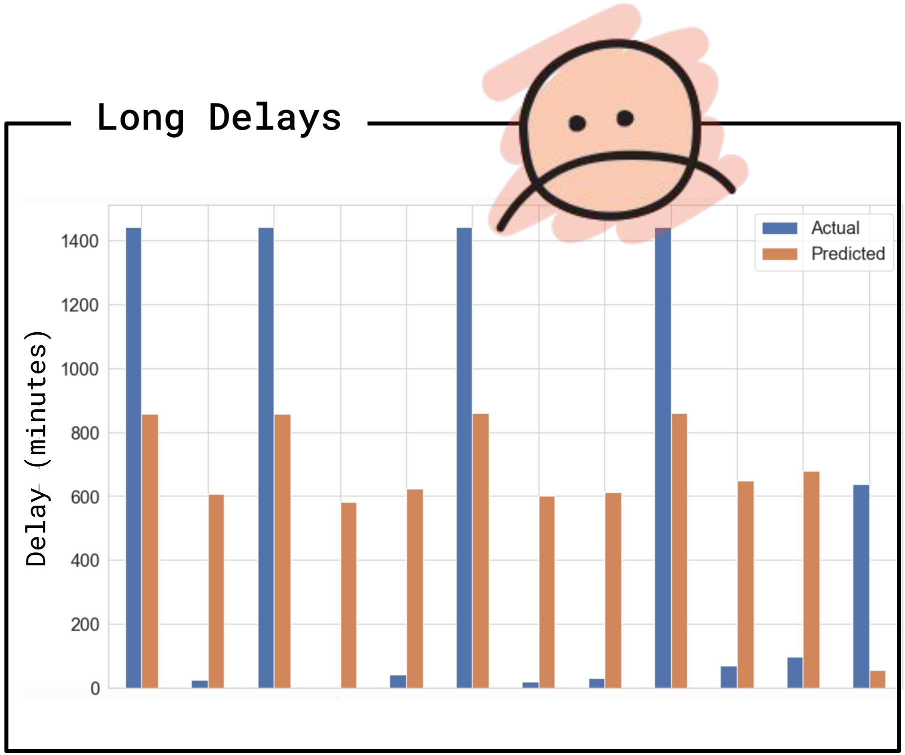


Can we predict length of delay?

Short Delays



Long Delays



Practical Application

(So What?)

Alert shoppers when a flight is at risk of being delayed.

Practical Application

(So What?)

Alert shoppers when a flight is at risk of being delayed.

KAYAK

Expedia®

Google

priceline®

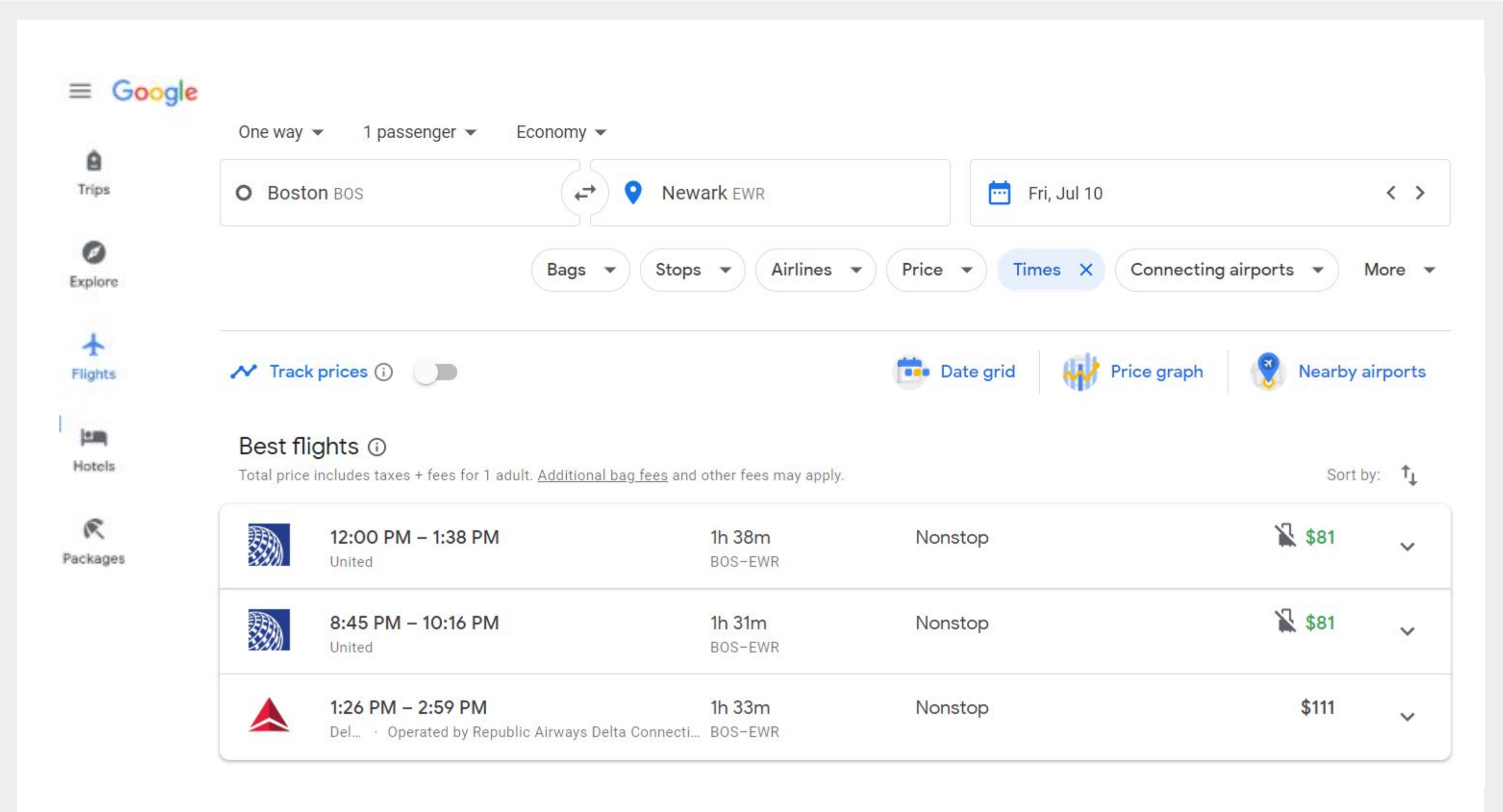
ORBITZ

Booking.com

Practical Application

(So What?)

Alert shoppers when a flight is at risk of being delayed.



Google

Trips
Explore
Flights
Hotels
Packages

One way ▾ 1 passenger ▾ Economy ▾

Boston BOS ⇄ Newark EWR

Fri, Jul 10

Bags ▾ Stops ▾ Airlines ▾ Price ▾ Times ✕ Connecting airports ▾ More ▾

Track prices ⓘ

Date grid Price graph Nearby airports

Best flights ⓘ

Total price includes taxes + fees for 1 adult. Additional bag fees and other fees may apply.

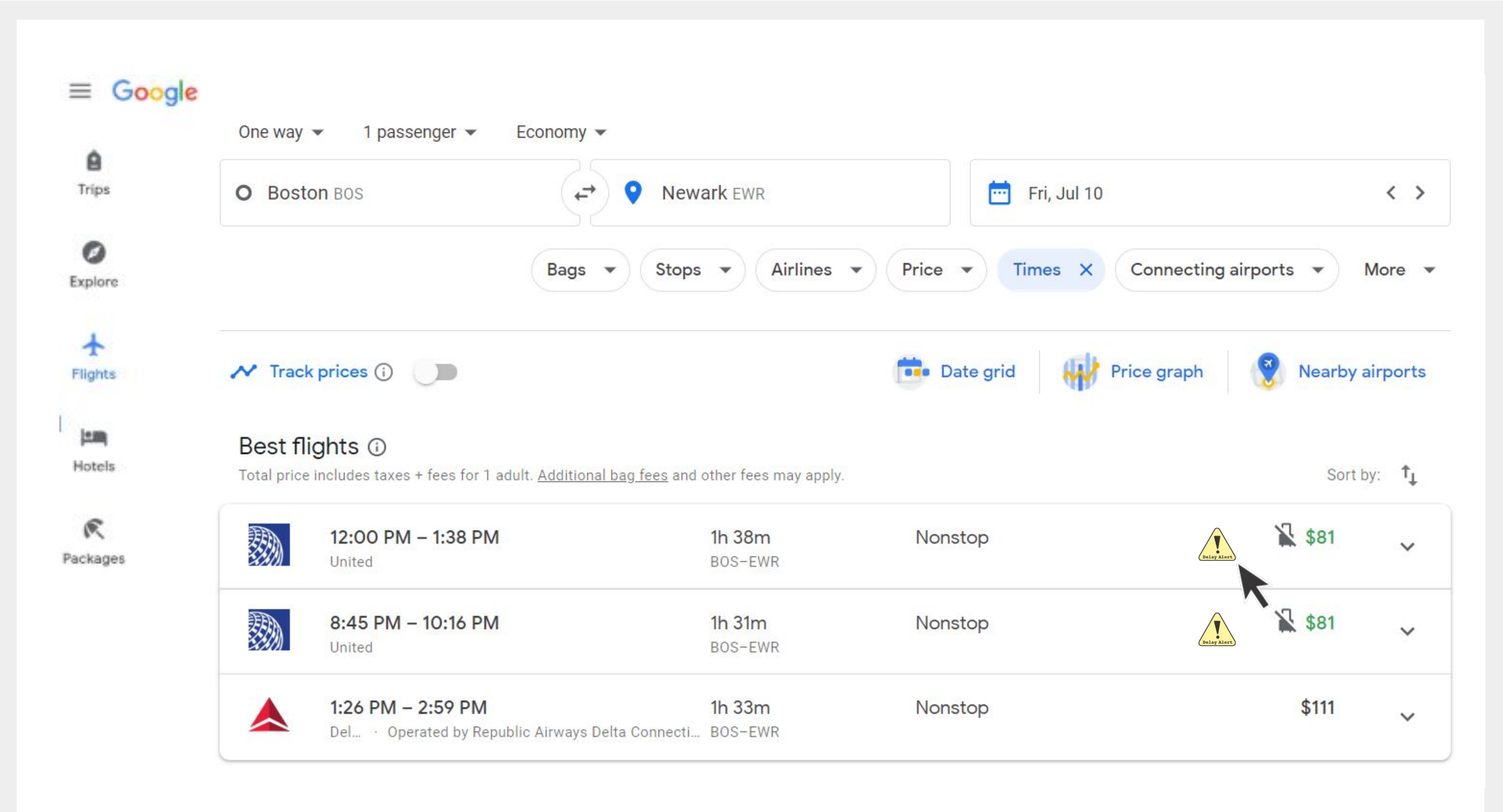
Sort by:

12:00 PM – 1:38 PM United	1h 38m BOS–EWR	Nonstop	$81	⌄
8:45 PM – 10:16 PM United	1h 31m BOS–EWR	Nonstop	$81	⌄
1:26 PM – 2:59 PM Del... · Operated by Republic Airways Delta Connecti...	1h 33m BOS–EWR	Nonstop	$111	⌄

Practical Application

(So What?)

Alert shoppers when a flight is at risk of being delayed.



One way · 1 passenger · Economy

Boston BOS → Newark EWR · Fri, Jul 10

Best flights

Total price includes taxes + fees for 1 adult. Additional bag fees and other fees may apply.

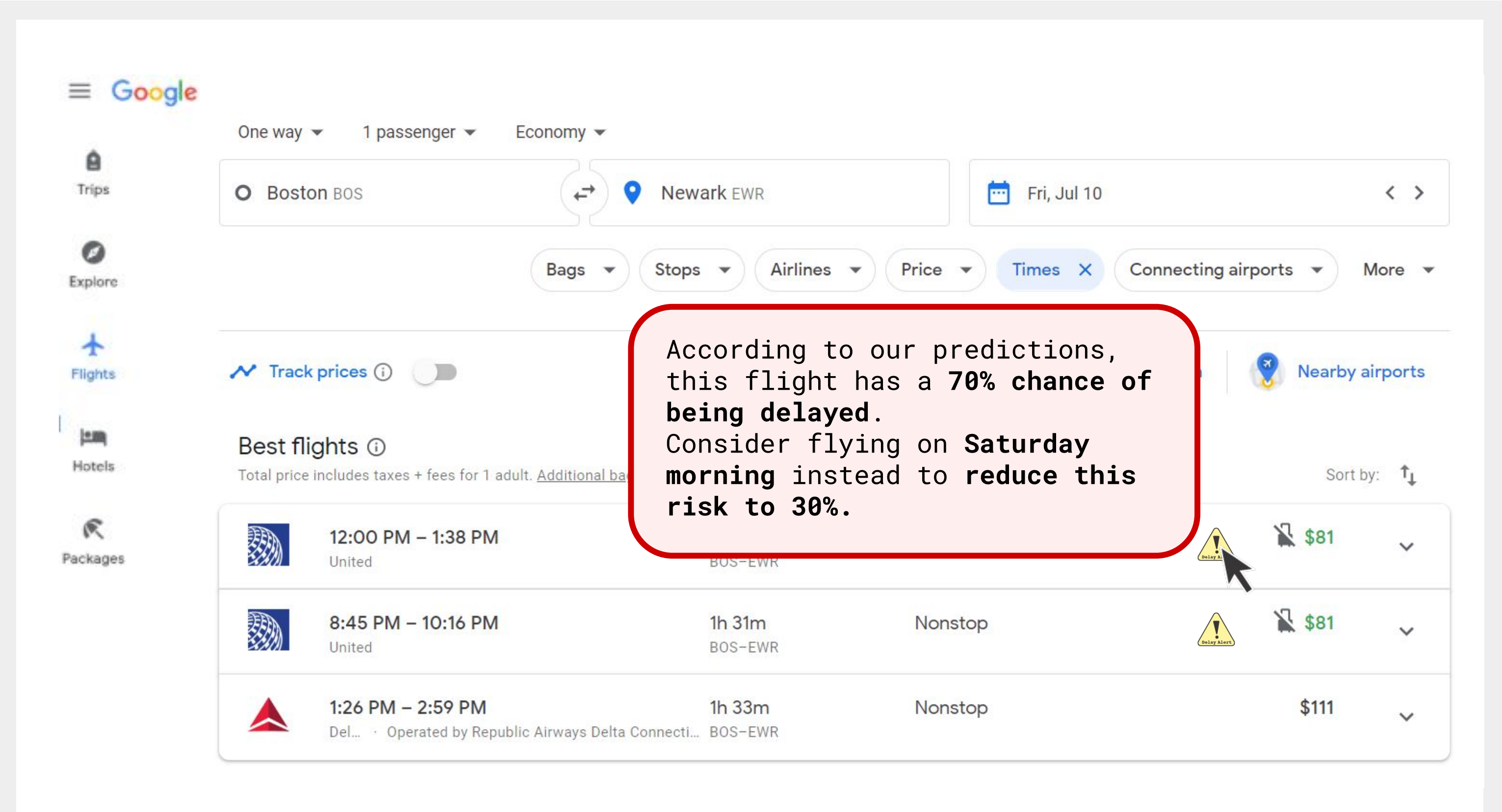
12:00 PM – 1:38 PM United	1h 38m BOS–EWR	Nonstop	⚠	$81
8:45 PM – 10:16 PM United	1h 31m BOS–EWR	Nonstop	⚠	$81
1:26 PM – 2:59 PM Del... · Operated by Republic Airways Delta Connecti...	1h 33m BOS–EWR	Nonstop		$111

Group 13

Practical Application

(So What?)

Alert shoppers when a flight is at risk of being delayed.



Boston BOS ⇄ Newark EWR

Fri, Jul 10

Best flights

Total price includes taxes + fees for 1 adult. Additional ba

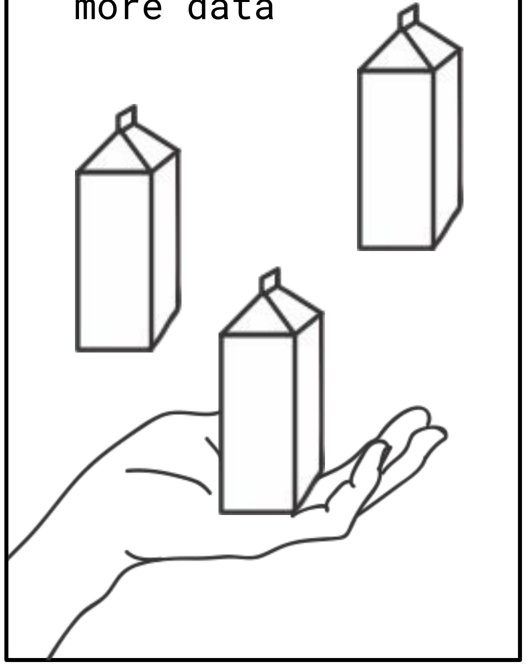
12:00 PM – 1:38 PM United	BOS–EWR			$81
8:45 PM – 10:16 PM United	1h 31m BOS–EWR	Nonstop	Delay Alert	$81
1:26 PM – 2:59 PM Del... · Operated by Republic Airways Delta Connecti...	1h 33m BOS–EWR	Nonstop		$111

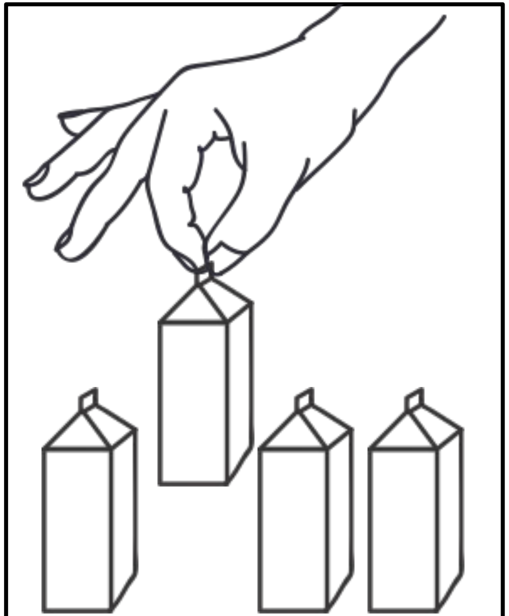
According to our predictions, this flight has a **70% chance of being delayed**.
Consider flying on **Saturday morning** instead to **reduce this risk to 30%.**

Group 13

Things we could do better

Get our hands on more data





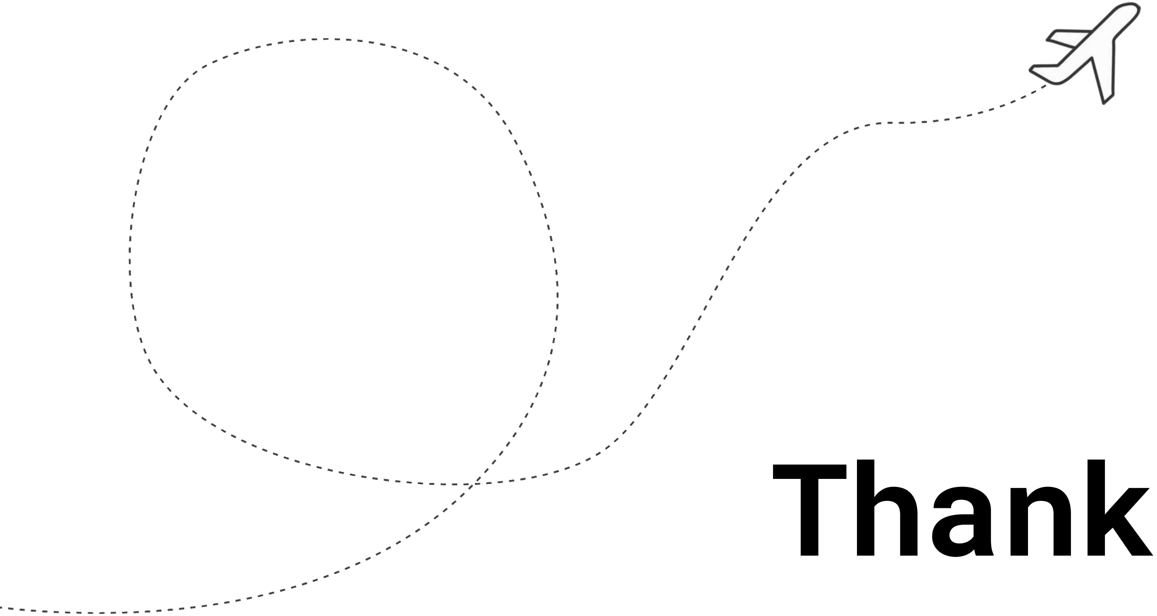
Optimize feature selection

Bucket delays by type:

- ➔ >3hrs (“Catastrophic”)
- ➔ <3hrs (“Not Catastrophic”)
- ➔ Cancelled

Bucket delays by cause:

- ➔ Airline
- ➔ Weather
- ➔ Air System
- ➔ Aircraft
- ➔ Security



Thank you

A quick overview...

Time of day and **length of flight** most affect whether a flight will be delayed.

Our model recall is **60% for delayed** flights and **70% for non-delayed** flights.

To improve we would use **more data**, optimize **feature selection** and investigate differences between **types of delays.**

... to open for questions :)