WILL THIS HAPPEN TO YOU?

Predicting delays on U.S. flights

MOTIVATION AND KEY QUESTION

- In 2015, 19% of flights were delayed 15 minutes or more.

What predicts flight delays?

Airline factors?

Airport factors?

Calendar factors?

Weather?

FOCUS ON RECALL, PRECISION, OR BOTH?

- If customer expects the flight to be on time and it's delayed, customer is frustrated. (recall)
- If customer expects flight to be delayed, leaves later for the airport, and misses the (actually on-time) flight, customer is frustrated. (precision)

  - Recall is important, but so is precision!

DATA

Flight data (source: Kaggle)

- All U.S. flights in 2015
- 5.9M flights out of and into ~300 airports

Weather data (source: NOAA)

- Roughly-hourly data from all airport and other observation stations
- 4.2M observations

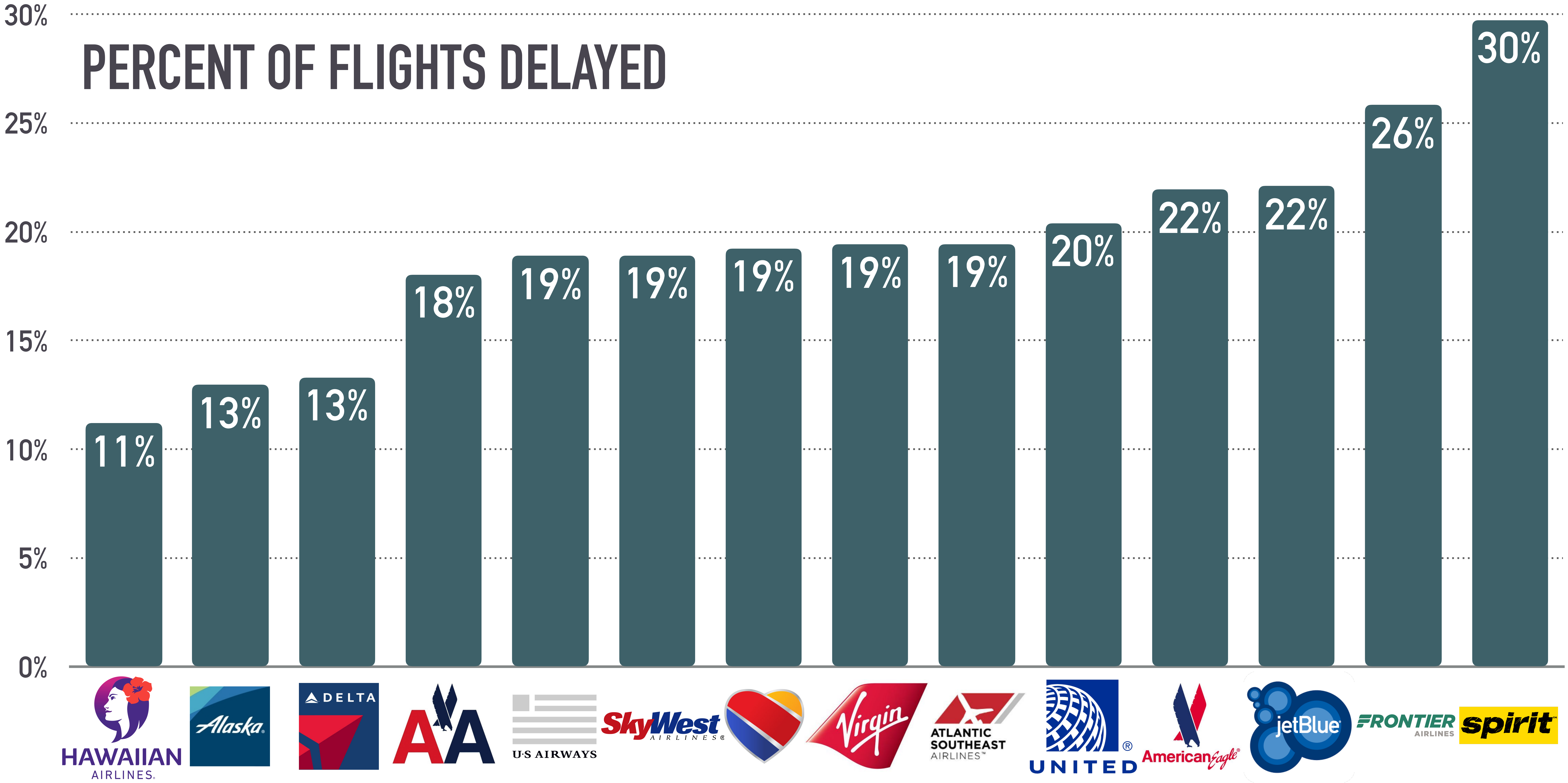
FEATURES

- Airline
- Origin airport
- Departure airport
- Month of year
- Day of week
- Hour of day
- Flight distance
- Weather data:
  - Precipitation
  - Temperature
  - Visibility
  - Cloud ceiling
  - Air pressure
  - Wind speed

ANALYTICAL APPROACH

- Test locally with 100K flights; then on AWS with 5.9M
- Bivariate exploratory analyses of features
- Model-building
  - Train/test split
  - Logistic regression with cross-validation
    - Class imbalance, so weight up the “delayed” class
  - Random forest classifier

PERCENT OF FLIGHTS DELAYED



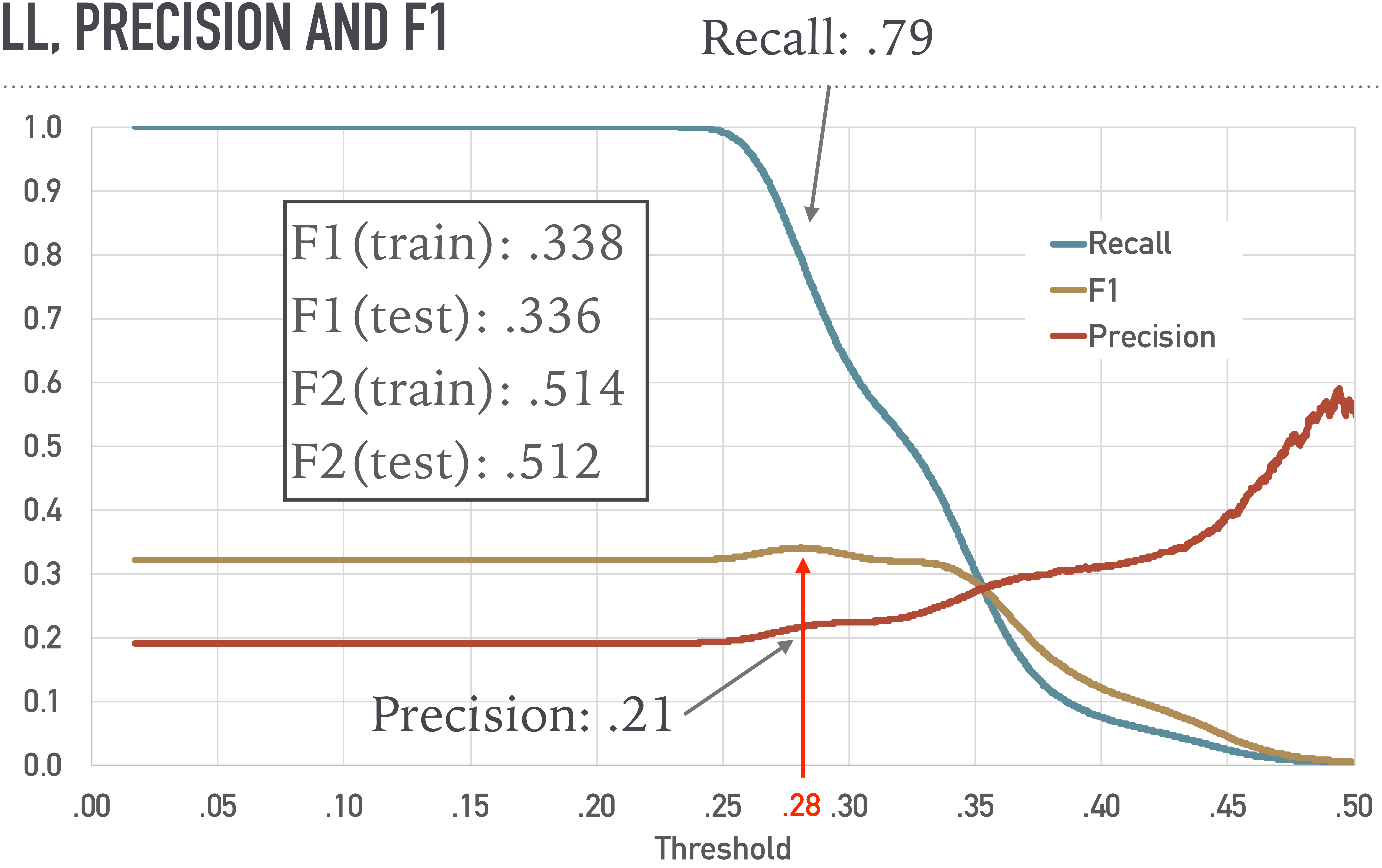
[Present Tableau visualizations]

ACCURACY AND F1/F2

- Logistic model accuracy is .81.
  - Great, right?
  - No. See confusion matrix:
- F1 (and F2) better metrics for me because they balance recall and precision

		Prediction	
		Not delayed	Delayed
Actual	Not delayed	2,676,738 (81.0%)	6,675 (0.2%)
	Delayed	617,502 (18.7%)	5,494 (0.2%)

RECALL, PRECISION AND F1



F1(train): .338

F1(test): .336

F2(train): .514

F2(test): .512

SO, IF YOU WANT TO BE ON TIME:

DO fly…

- on Delta
- in Apr, Sep or Nov
- between 7a and noon
- on Saturday
- into or out of ATL

DON'T fly…

- on wet or windy days
- between 6p and midnight
- in Feb, Jun or Jul
- on Monday or Thursday
- on Spirit, Southwest or United
- out of ORD or into LAX

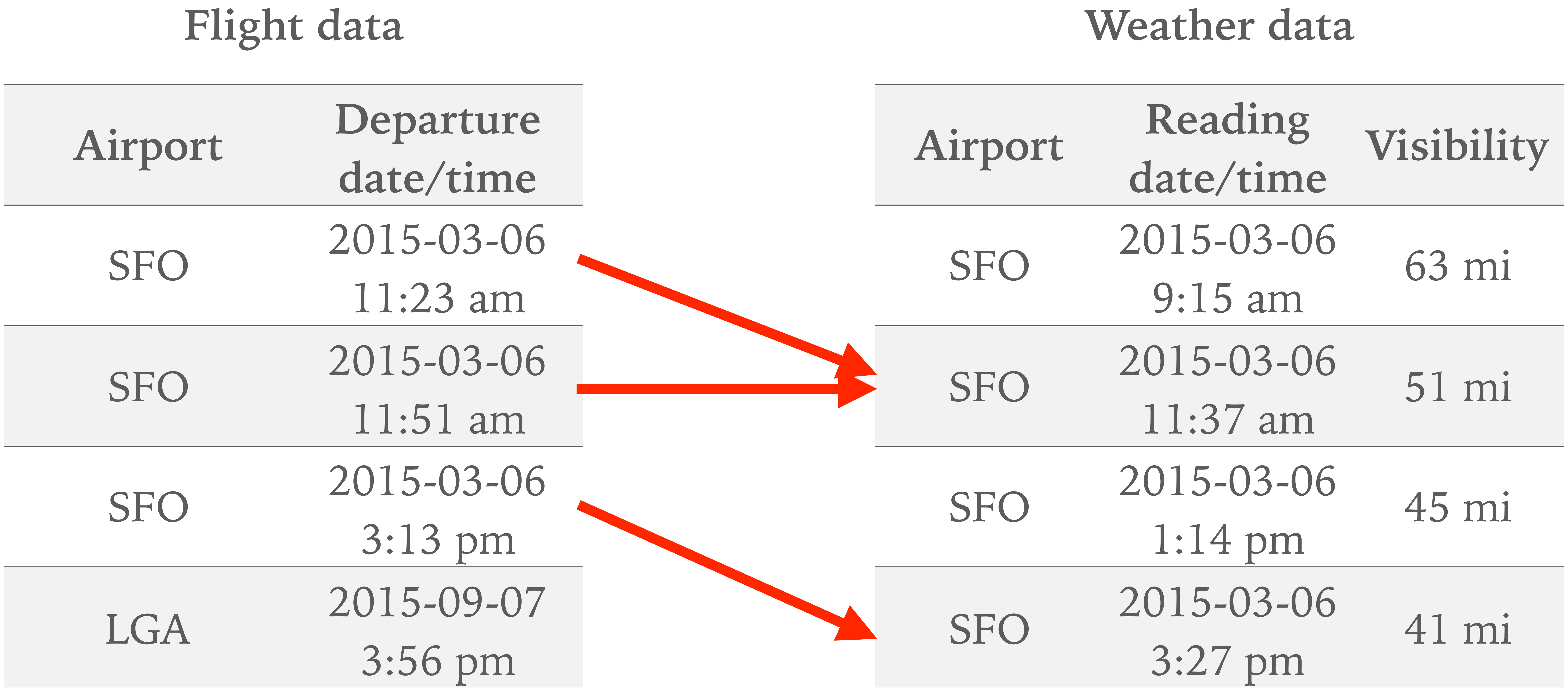
BON VOYAGE!

SHOUT-OUTS

Vaughn, *F1 master*
Patrick, *chart interpreter*
TJ, *regex whiz*

AWESOME PANDAS MERGE COMMAND

- pd.**merge_asof**() with parameter: (direction = 'nearest')



Flight data

Airport	Departure date/time
SFO	2015-03-06 11:23 am
SFO	2015-03-06 11:51 am
SFO	2015-03-06 3:13 pm
LGA	2015-09-07 3:56 pm

Weather data

Airport	Reading date/time	Visibility
SFO	2015-03-06 9:15 am	63 mi
SFO	2015-03-06 11:37 am	51 mi
SFO	2015-03-06 1:14 pm	45 mi
SFO	2015-03-06 3:27 pm	41 mi