

# ONLINE DATA SCIENCE

## Module 3 Final Project

1

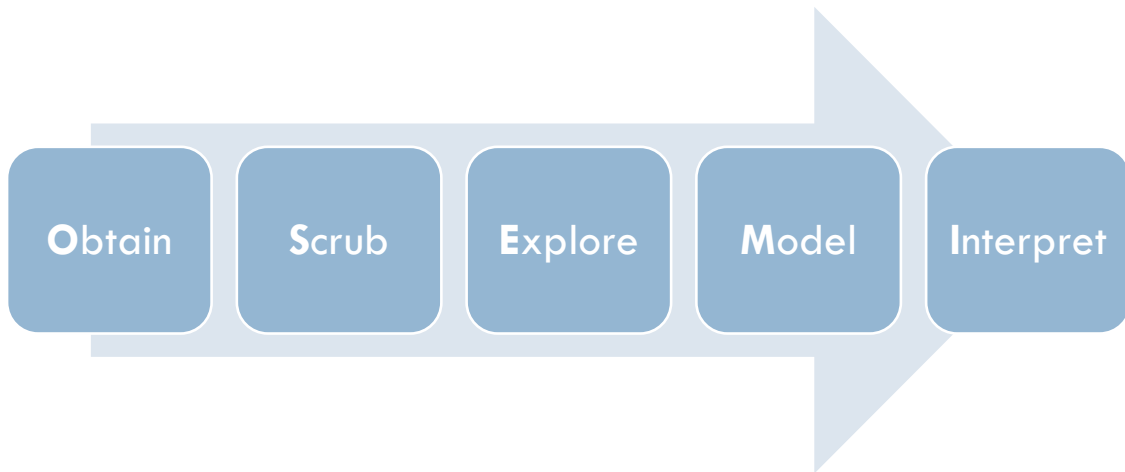
## A Mobile App to Predict Flight Delays

Help users to avoid those lengthy waits—often before delays and cancelations are announced by the airline.



2

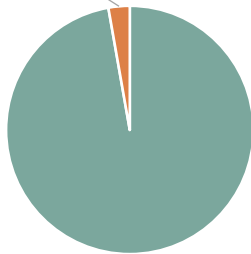
## Process—OSEMN Framework



3

## 1. Obtain

50000,  
3%  
2015 Flights by Delta,  
American, US Airways



■ Subset used for modeling

Airports

Airline

Distance

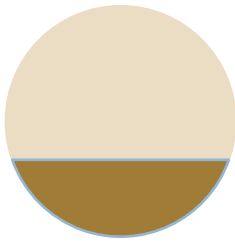
Delay  
Time

Canceled

Flight  
Number

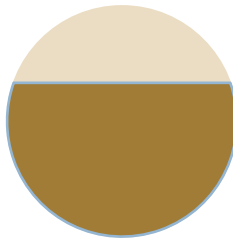
4

## 2. Scrub



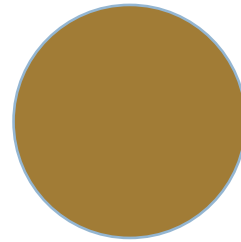
### Unclean

Remove/replace  
missing data



### Cleaner

Remove outliers,  
extraneous  
values



### Clean

Ready for  
modeling

5

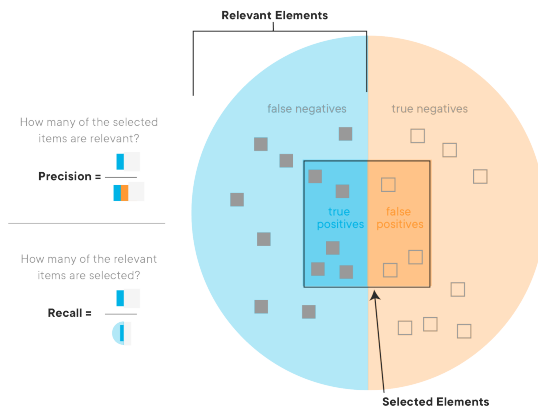
## 3. Explore

Classifying Flight Departures



6

## 4. Model



**Accuracy—91%**

- 91% of all the model's predictions were correct.

**Precision—88%**

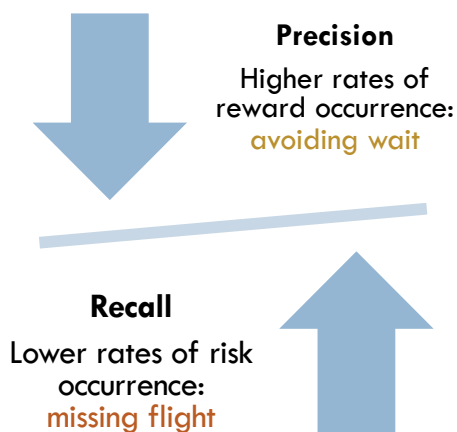
- For every 10 flights that the model classified as *delayed/canceled*, 8.8 truly were.

**Recall—75%**

- The model caught 7.5 out of every 10 delayed/canceled flights.

7

## 5. Interpret



8

## Future work

Increase resources to study larger sample, more airlines

Correct for possible “class imbalance” issue

Continue tweaking models

9

Thank You!

10