

PREDICTING FLIGHT DELAY

SARA ZONG
FEB 10TH, 2021

MOTIVATION:

- ▶ Being stuck at the airport is horrible.



- ▶ Moreover, if we are traveling on a tight business schedule, we want more control over our schedule.
- ▶ The goal is to help travelers avoid flight delay.

TOOLS:



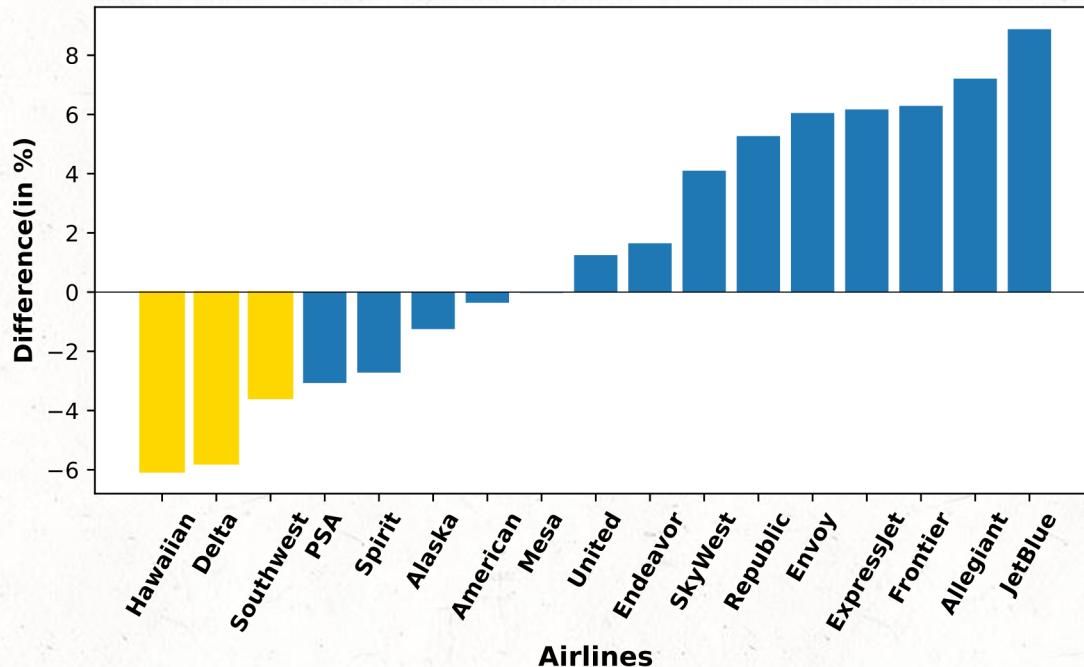
DATA COMES FROM THE BUREAU OF TRANSPORTATION

- ▶ Time: January 2019
- ▶ Size: 565,963 records for one month
- ▶ Features: Date, Airlines, Origin, Destination, Scheduled departure, Scheduled arrival, Airtime, Distance, **incoming flights, interaction of scheduled arrival and incoming flights**
- ▶ Target: Arrival delay > 15min

RANKING OF AIRLINES WITH THE LEAST TO THE MOST DELAY

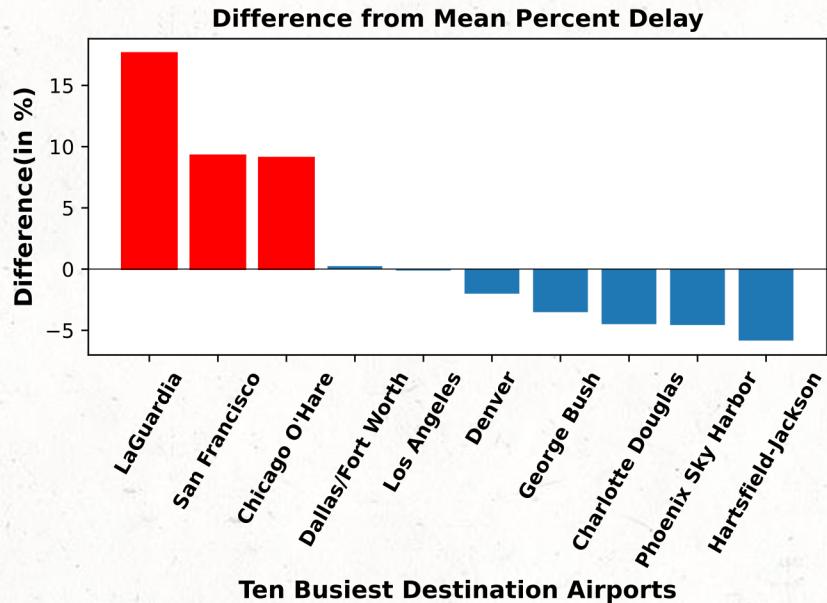
Hawaiian, Delta, and Southwest have the least flight delay.

Difference from Mean Percent Delay by Airlines

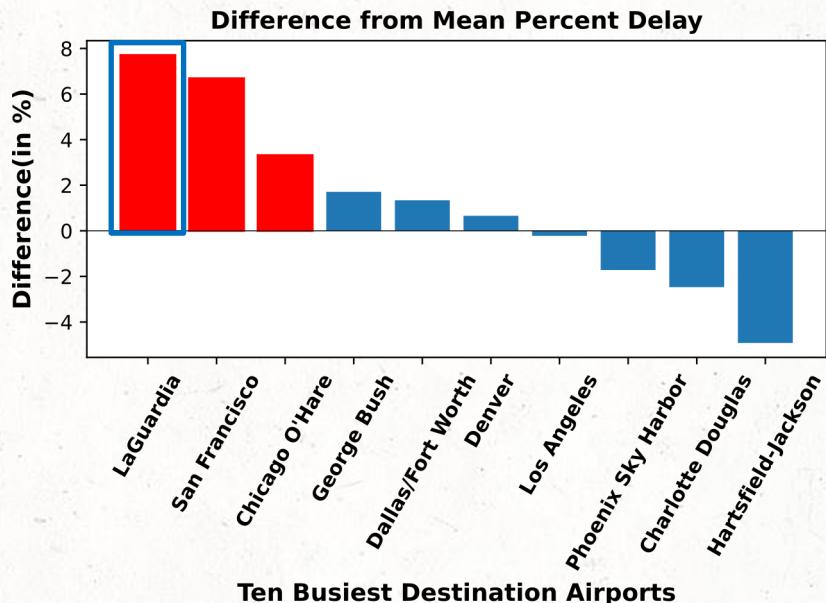


RANKING OF AIRPORTS WITH THE MOST TO THE LEAST DELAY

LaGuardia, San Francisco, and Chicago O'Hare have the most flight delay.



In 2019, LaGuardia, San Francisco, and Chicago O'Hare have the most flight delay.



RANDOM FORESTS WORKS THE BEST WITH THE DATASET.

RF	PRECISION	RECALL	F1 SCORE
Delay(TH 0.5)	0.51	0.23	0.32
Delay(TH 0.3)	0.37	0.45	0.41
Delay(TH 0.1)	0.25	0.79	0.38

Of about 20,000 delays, about 5,000 would be classified delay (at threshold of 0.5).

CONCLUSIONS:

- ▶ Some airlines tend to be better than the others on having less flight delay.
- ▶ Some airports tend to be better than the others on having less flight delay, but we will also need to consider when we are traveling.
- ▶ Random Forests is the best model so far, but for sure more work could be done to improve the recall rate.

FUTURE WORK:

- ▶ Focus on a small area (for example, SF bay area) and bring in weather information.
- ▶ Feature engineering: count the number of flight delay per airport per airline a couple days in advance

THANKS!



any
questions
?

CREDITS

- ▶ https://www.transtats.bts.gov/DL_SelectFields.asp?Table_ID=236&DB_Short_Name=On-Time
- ▶ Presentation template by [SlidesCarnival](#)
- ▶ <https://timesofmalta.com/articles/view/air-passenger-rights-when-flights-are-delayed-or-cancelled.709080>

APPENDIX:

Random Forests Confusion Matrix

Actual			
On Time		87606	4760
Delay	On Time	15941	4886
	Delay		
Prediction			

Logistic Regression Confusion Matrix

Actual			
On Time		89673	2693
Delay	On Time	20205	622
	Delay		
Prediction			

APPENDIX:

LR	Precision	recall	f1-score
On Time	0.82	1.00	0.90
Delay	0.76	0.02	0.04

Ensemble	Precision	recall	f1-score
On Time	0.82	1.00	0.90
Delay	0.87	0.01	0.03