

# 2018 Flights Delay



The background image shows a flight information display board with multiple panels. The panels display flight details such as time, flight number, origin, and status. Many flights are marked as 'Delayed' in red text, while others are 'At gate' in yellow. The board is angled, showing several panels simultaneously.

Time	Flight	From	Hall	Status
19:20	QR 5849	Nagoya	A	Delayed
19:30	HX 629	Seoul/ICN	A	Delayed
19:30	KA 745	Zhengzhou	A	Delayed
19:30	WE 608	Phuket	A	Delayed
19:30	CX 6249	Quanzhou	A	Delayed
19:35	HX 6683	Okayama	A	Delayed
19:35	PX 008	Port Moresby	A	At gate
19:22				
20:05	CI 923	Takamatsu	A	Delayed
20:05	UO 523	Singapore	A	Delayed
20:10	CX 734	Manila	A	Delayed
20:15	CX 918	Chicago	A	Est at 20:50
20:20	CX 807	Chicago		
Time	Flight	From	Hall	Status
20:25	CX 776	Jakarta	A	Delayed
20:30	CX 369	Shanghai/PVG	A	Delayed
20:30	CX 656	Bangkok		
20:35	FD 524	Phuket		
20:55	FJ 5395	Osaka/Kansai		
20:55	HX 693	Sapporo		
Time	Flight	From		
20:55	UO-639	Fukuoka		
21:00	CX 581	Sapporo		
21:00	TR 974	Singapore		
21:05	CX 784	Denpasar		

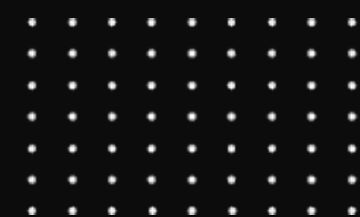
Classification Analysis Summary  
Daniel M. Smith July 3, 2021



# Summary

*2018 Domestic Flights Analysis  
Using Machine Learning Classification Models  
To determine if flights are delayed by looking at  
Distance, Airline, Origin, Destination Day of Week, Month, Time of Flight*

# Outline



1 Data

2 Methods

3 Results

3 Conclusions





# Business Problem

Domestic Flights Analysis for Southwest Airlines

Looking at industry delays and routes

For improvement opportunities

Which Airlines usually late/early?

Which routes are late/early?



# Data

Airline and Cancellation [Dataset](#) on Kaggle by Yuanyu 'Wendy' Mu

All Records from United States Department of Transportation

2018 data containing 7.21M records 851MB

Initial Features included:

FL\_DATE - Date of Flight  
OP\_CARRIER - Flight Carrier  
OP\_CARRIER\_FL\_NUM - Flight Carrier Identifier  
ORIGIN- Start Airport  
DEST- Destination Airport  
CRS\_DEP\_TIME - Computer Reservation System (CRS) Departure Time  
DEP\_TIME - Actual Departure Time  
DEP\_DELAY - Dep Time minus CRS Dep Time in Min  
TAXI\_OUT - Time To taxi  
WHEELS\_OFF - Time Wheels in Air  
WHEELS\_ON - Time Wheels on Ground  
TAXI\_IN - Time To taxi  
CRS\_ARR\_TIME - Computer Reservation System (CRS) Arrival Time

ARR\_TIME - Actual Arrival Time  
ARR\_DELAY - ARR\_Time minus CRS\_ARR\_TIME in Min  
CANCELLED - Flight Cancelled or not  
CANCELLATION\_CODE - Cancel Code  
DIVERTED - Flight Was diverted or Not  
CRS\_ELAPSED\_TIME -CRS scheduled Flight Time  
ACTUAL\_ELAPSED\_TIME - Actual Flight Time  
AIR\_TIME - Time in the Air  
DISTANCE - Distance of Flight  
CARRIER\_DELAY - Carrier Delay in Min  
WEATHER\_DELAY - Weather Delay in Min  
CANCELLATION\_CODE - Cancelled Code  
NAS\_DELAY - National Air Service Delay in Min  
SECURITY\_DELAY - Sec Delay in Min  
LATE\_AIRCRAFT\_DELAY - Delay due to late Aircraft in Min



# Methods

Exploratory Data Analysis with Descriptive Statistics

Classification Analysis with Machine Learning Algorithms

Logistic Regression, Decision Trees, Random Forests, XGBoost

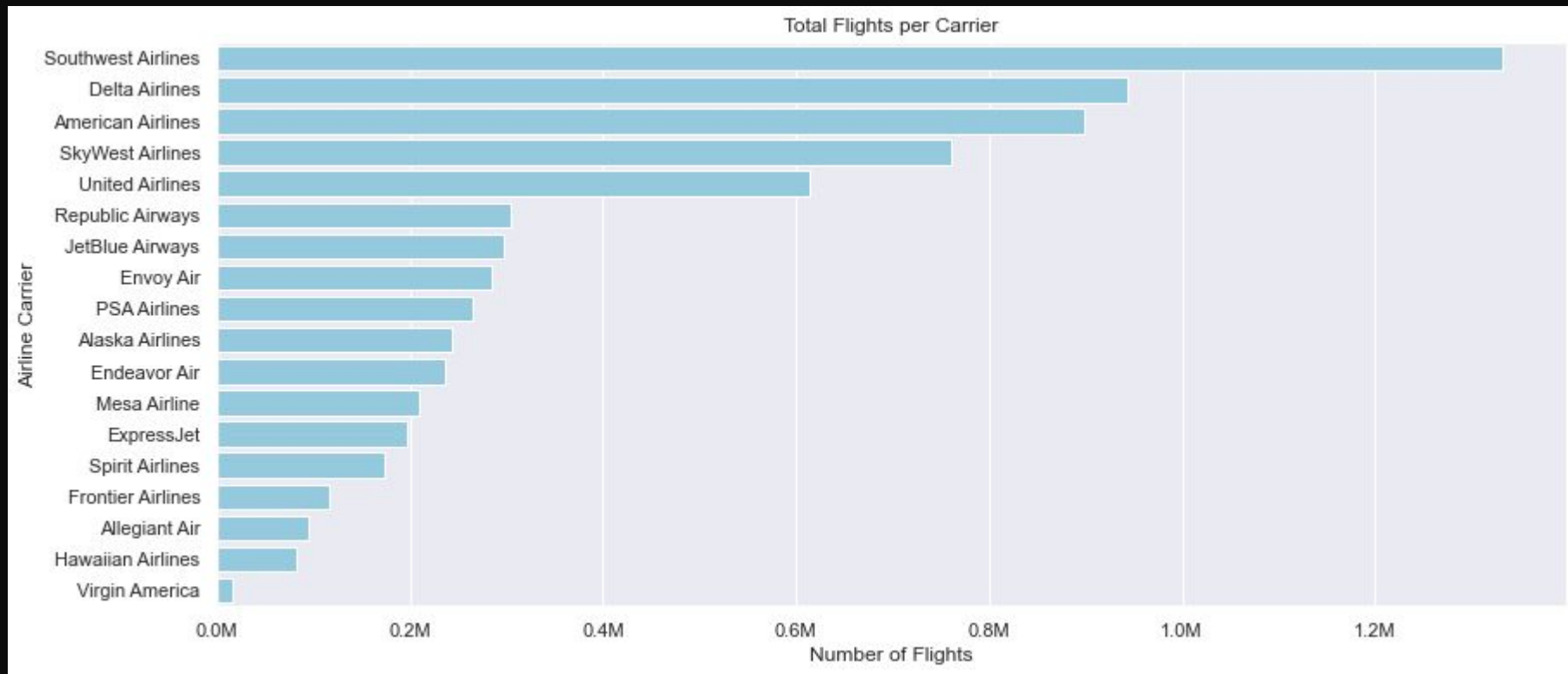
■ With GridSearch narrowing down the most optimal Hyperparameters





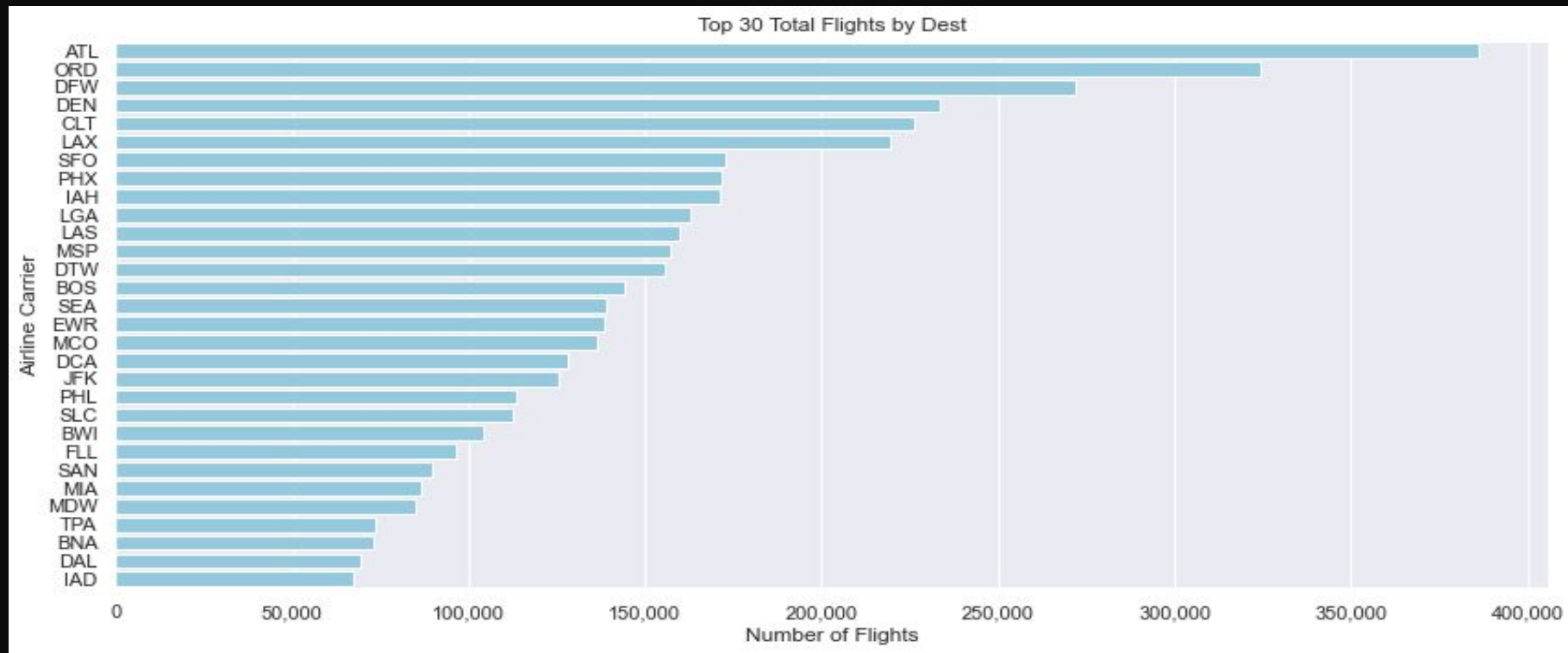


# Results-by Carrier





# Results-By Destination





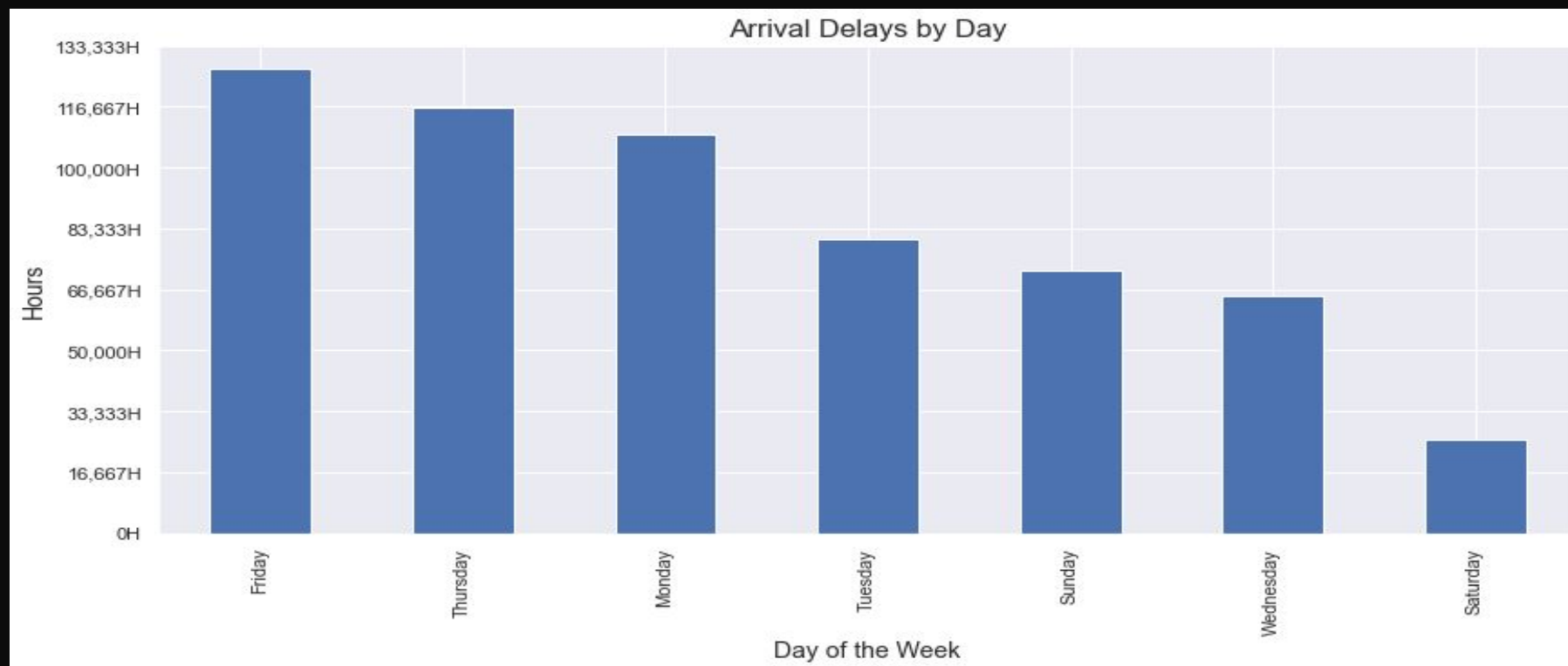


# Results-Delay by Dest



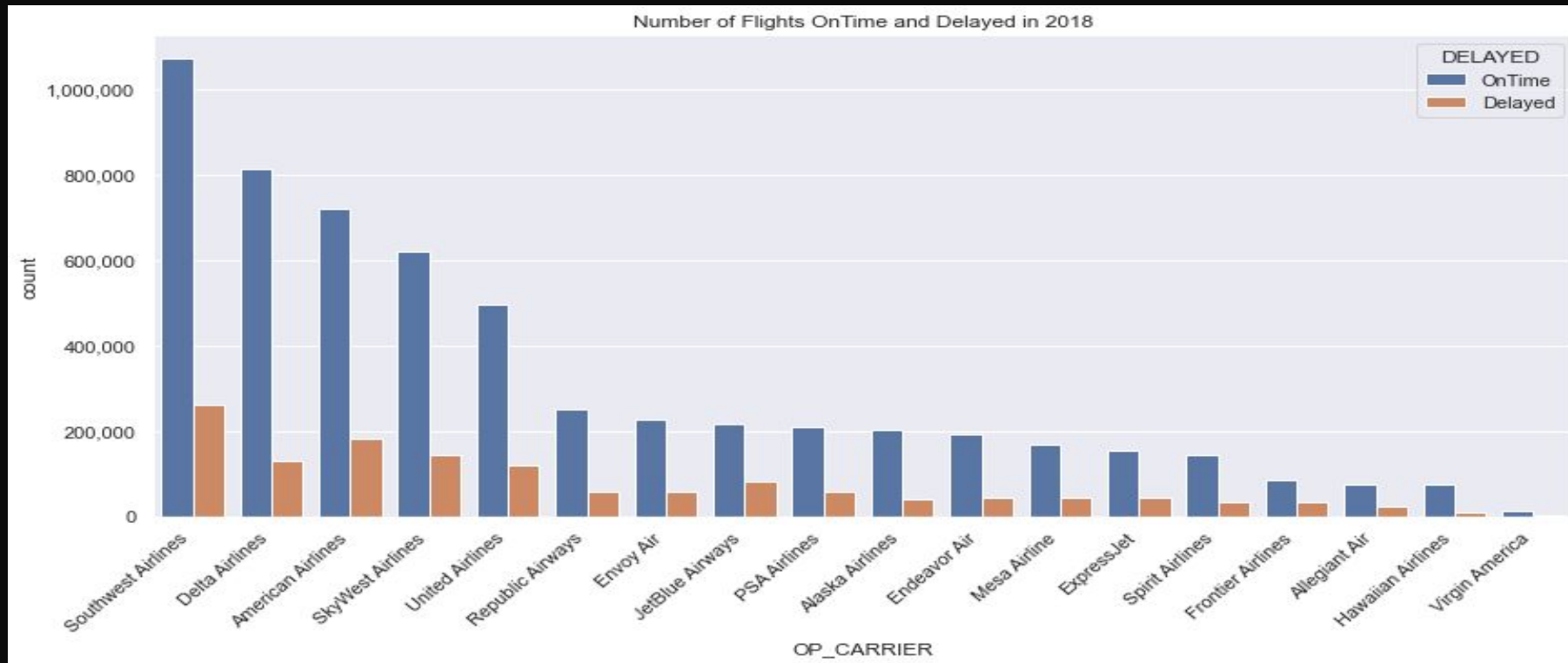


# Results-Delays by Day



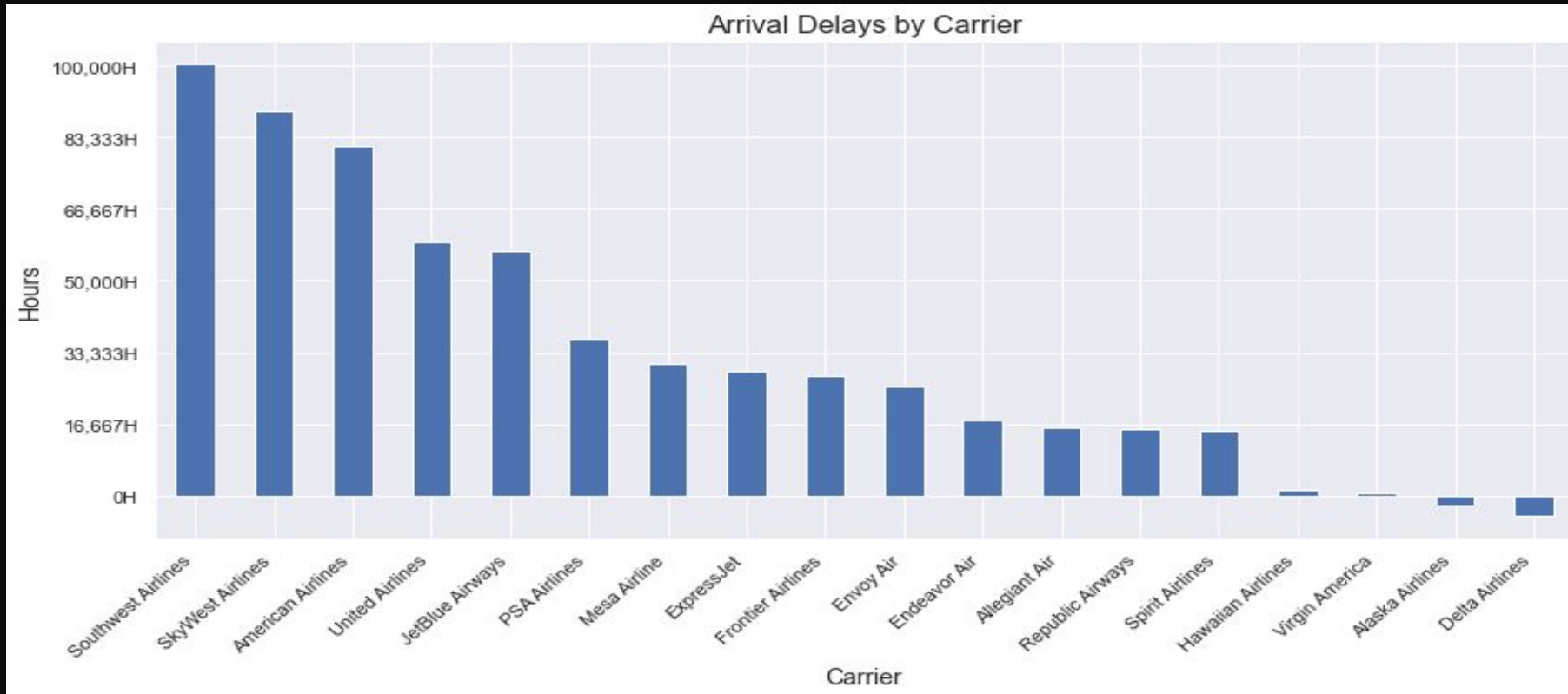


# Results-OnTime Delayed





# Results-Delay by Carrier





# Results-A A Late

## Aircraft Arriving Late: Causes of the Original Delay

Most Recent Month    Year To Date

Note: Data are available from June 2003 through April 2021.

		Number of Operations	Delayed Minutes	% of Total Delayed Minutes
<b>Air Carrier Delay</b>		244,877	17,265,654	48.07%
<b>Security Delay</b>		1,168	82,023	0.23%
<b>National Aviation System Delay</b>	<b>Weather</b>	146,724	10,447,002	29.09%
	<b>Volume</b>	54,163	3,803,737	10.59%
	<b>Equipment</b>	725	50,369	0.14%
	<b>Closed Runway</b>	10,841	752,097	2.09%
	<b>Other</b>	3,952	279,536	0.78%
<b>Extreme Weather Delay</b>		45,424	3,235,932	9.01%
<b>Total Aircraft Arriving Late</b>		<b>507,874</b>	<b>35,916,350</b>	<b>100.00%</b>



# Results-Delta Vs SW

## Delta Airlines Inc. (DL) (January - December, 2018)

[Most Recent Month](#)[Year to Date](#)[View Pie Chart](#)[Print Table](#)[Download Raw Data](#)

	Number of Operations	% of Total Operations	Delayed Minutes	% of Total Delayed Minutes
On Time	813,648	85.71%	N/A	N/A
Air Carrier Delay	39,509	4.16%	2,913,069	34.25%
Weather Delay	6,904	0.73%	730,316	8.59%
National Aviation System Delay	47,306	4.98%	2,332,362	27.42%
Security Delay	104	0.01%	5,965	0.07%
Aircraft Arriving Late	36,367	3.83%	2,524,793	29.68%
Cancelled	3,528	0.37%	N/A	N/A
Diverted	1,918	0.20%	N/A	N/A
Total Operations	949,283	100.00%	8,506,505	100.00%

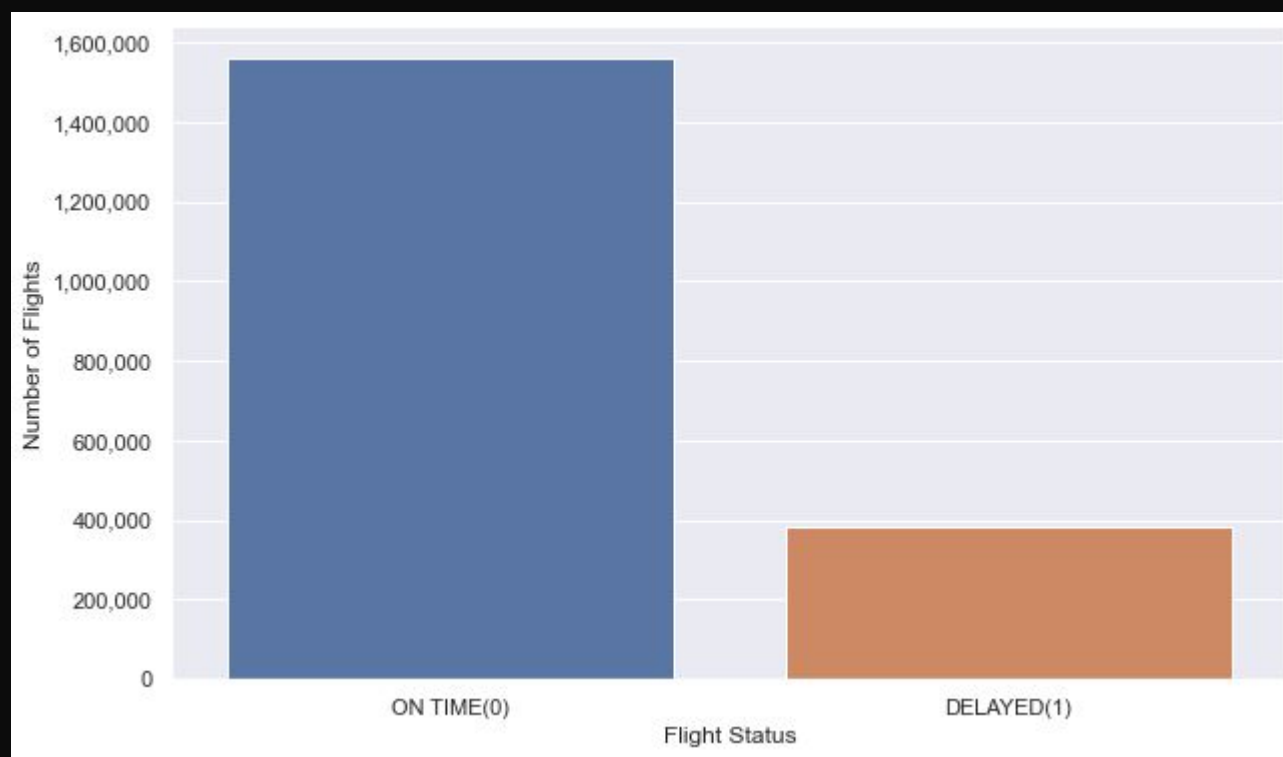
## Southwest Airlines Co. (WN) (January - December, 2018)

[Most Recent Month](#)[Year to Date](#)[View Pie Chart](#)[Print Table](#)[Download Raw Data](#)

	Number of Operations	% of Total Operations	Delayed Minutes	% of Total Delayed Minutes
On Time	1,071,259	79.20%	N/A	N/A
Air Carrier Delay	82,371	6.09%	4,155,843	31.51%
Weather Delay	4,553	0.34%	353,248	2.68%
National Aviation System Delay	52,246	3.86%	2,178,014	16.51%
Security Delay	841	0.06%	40,498	0.31%
Aircraft Arriving Late	120,339	8.90%	6,463,171	49.00%
Cancelled	18,275	1.35%	N/A	N/A
Diverted	2,668	0.20%	N/A	N/A
Total Operations	1,352,552	100.00%	13,190,774	100.00%



# Results-Delay Split







# Results-Modeling

Modeling with continuous features Distance, Flight Time,  
Categoricals Weekdays, Months, Top 5 Airlines, Top 30 Origins and Destinations  
To Classify if Delayed or not. Delays are on Arrival Delays and gt=15 mins

Best Predictive Results were found with the XGBoost algorithm  
With a Recall of 59%, Accuracy 66% , F1 value of .59

MODEL	RECALL	ACCURACY	F1
XGBoost	59%	66%	59%
Random Forest	59%	65%	57%
Decision Tree	39%	68%	55%
Logistic Regression	DNF		



# Conclusions

Predicted 59% of delayed flights

SW Focus on Reducing Backup Delays as that is 50% of All Delays

## Challenges

Large Dataset and finding an appropriate model for the complexity of the data was a challenge.

## Future Improvements

Reduction of origins and destinations: Top 20, Top 10

Try PCA analysis



Thank You!

Email: danielmsmith1@gmail.com



GitHub: @SunTzuLombardi

LinkedIn: [linkedin.com/in/daniel-m-s-a8587a1/](https://www.linkedin.com/in/daniel-m-s-a8587a1/)



# CREDITS



Template Created By : MyFreeSlides.com



**Photo/Graphics/Icons/Designs Credits:**

unsplash.com, pixabay.com, pexels.com

&

Microsoft Inc., TrueMitra Designs

Title Slide/Main Image Credit : <https://www.pexels.com/@oscar-chan-1470482>

## Get More Resources :



*Templates & themes*



*Design Elements*



*Backgrounds*