



# Flight Delays

**Presented by:**

**•Gilles Bouyer**

**August 8, 2018**

# Waiting in the airport ...



# Waiting in the airport ...



# Waiting in the airport ...



# Hard Case



# Welcome to the delay predictor app

**Enter the flight info and I will tell you if you will arrive late.**

Flight Number:	<input type="text"/>
Departing From:	<input type="text"/>
Current Delay at Departure(in minutes):	<input type="text"/>
Destination Airport:	<input type="text"/>
Month:	<input type="text"/>
Day of the Week:	<input type="text"/>
Distance (miles)	<input type="text"/>

**Click Here For Prediction**

[Predictor](#)

# Example Delayed

Enter the flight info and I will tell you if you will arrive late.

Flight Number:	AA3112
Departing From:	Chicago
Current Delay at Departure(in minutes):	10
Destination Airport:	Milwaukee
Month:	August
Day of the Week:	3
Distance (miles)	100

Click Here For Prediction

You will be delayed

# Example Delayed

Enter the flight info and I will tell you if you will arrive late.

Flight Number:	<input type="text" value="AA447"/>
Departing From:	<input type="text" value="Chicago"/>
Current Delay at Departure(in minutes):	<input type="text" value="12"/>
Destination Airport:	<input type="text" value="Phoenix"/>
Month:	<input type="text" value="August"/>
Day of the Week:	<input type="text" value="3"/>
Distance (miles)	<input type="text" value="1140"/>

Click Here For Prediction

You will be on-time

<https://fr.flightradar24.com/live/flight/AAL447>

# Behind the scene

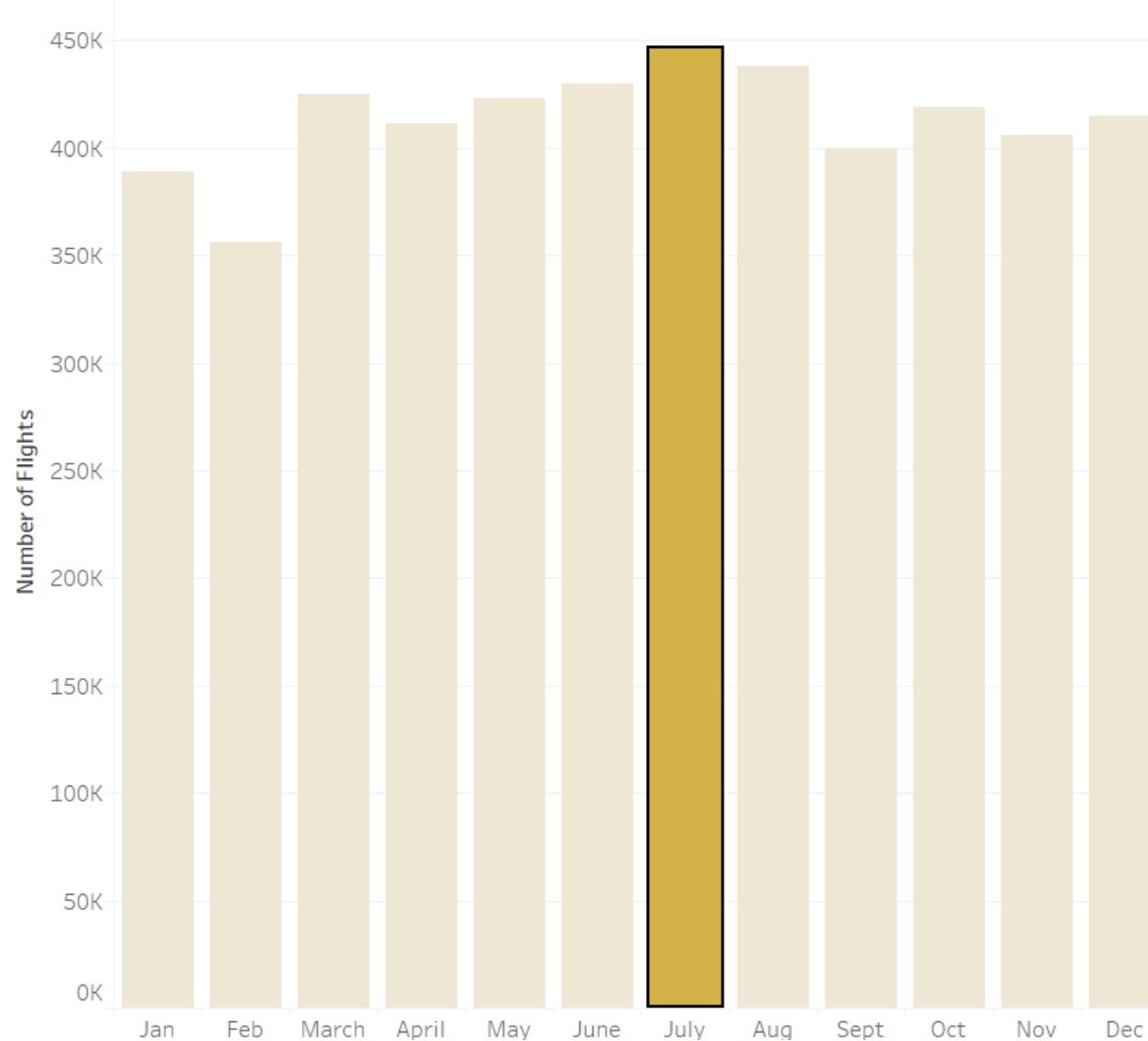


# Collect Data – My thanks to Russell Jurney

- 2015 Flights delays from:
- [http://s3.amazonaws.com/agile\\_data\\_science/On\\_Time\\_On\\_Time\\_Performance\\_2015.csv.bz2](http://s3.amazonaws.com/agile_data_science/On_Time_On_Time_Performance_2015.csv.bz2)
- 5 Million records – 109 features. → 27 predicting features

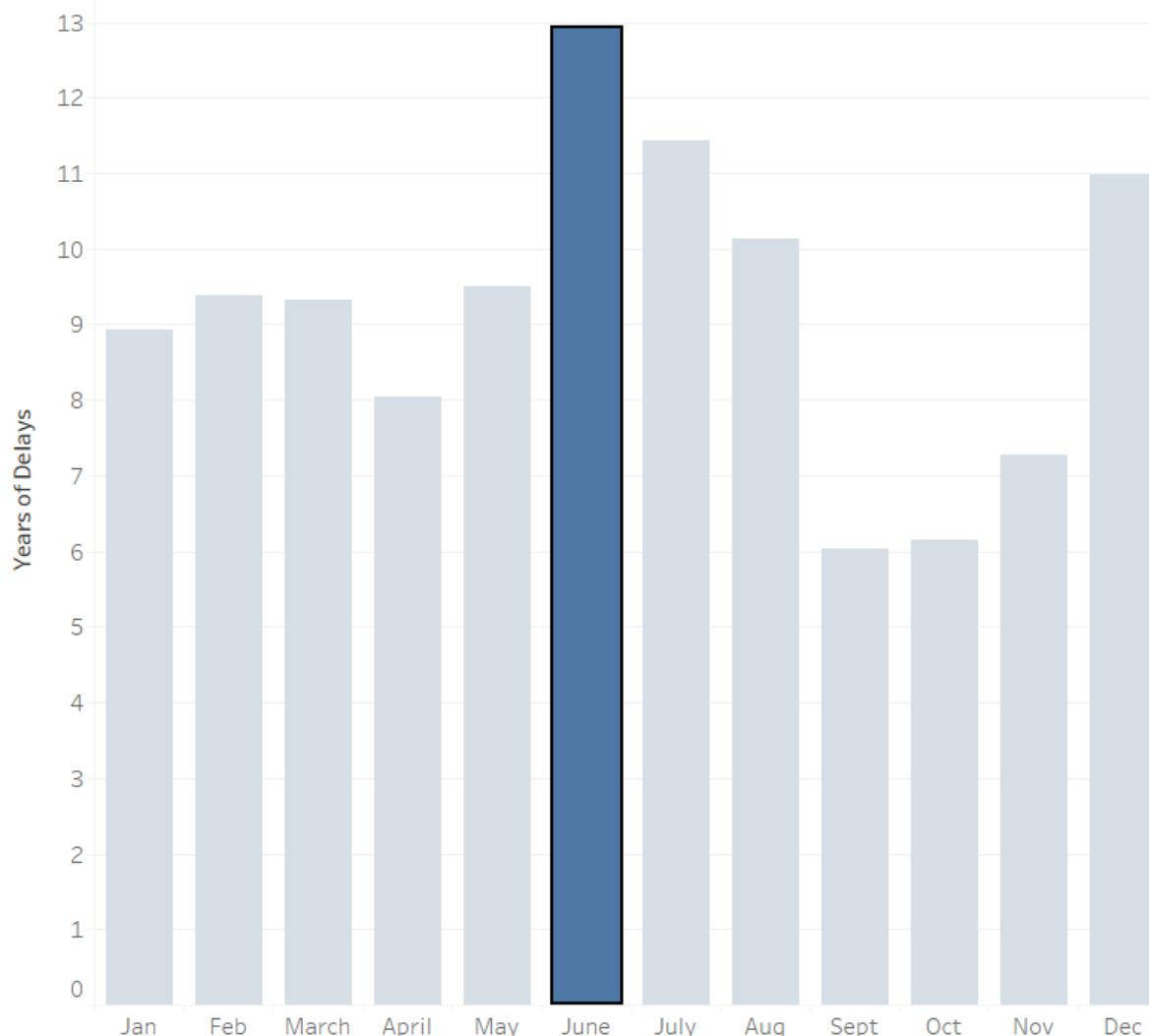
	Year	Quarter	Month	DayofMonth	DayOfWeek	FlightDate	UniqueCarrier	AirlineID	Carrier	TailNum	FlightNum	OriginAirportID	OriginAirportSeqID
0	2015	1	1	1	4	2015-01-01	AA	19805	AA	N001AA	1519	11298	1129803
1	2015	1	1	1	4	2015-01-01	AA	19805	AA	N001AA	1519	13244	1324402
2	2015	1	1	1	4	2015-01-01	AA	19805	AA	N002AA	2349	13930	1393003
3	2015	1	1	1	4	2015-01-01	AA	19805	AA	N003AA	1298	11298	1129803
4	2015	1	1	1	4	2015-01-01	AA	19805	AA	N003AA	1422	11298	1129803
	DelayMinutes	DepDel15	DepartureDelayGroups	DepTimeBlk	TaxiOut	WheelsOff	WheelsOn	TaxiIn	CRSArrTime	ArrTime	ArrDelay	ArrDelayMinutes	ArrDel15
	0.0	0.0	-1.0	1300-1359	16.0	1358.0	1457.0	7.0	1510	1504.0	-6.0	0.0	0.0
	0.0	0.0	-1.0	1500-1559	9.0	1555.0	1712.0	9.0	1730	1721.0	-9.0	0.0	0.0
	0.0	0.0	0.0	1800-1859	31.0	1916.0	2125.0	16.0	2115	2141.0	26.0	26.0	1.0
	100.0	1.0	6.0	1800-1859	33.0	2033.0	2306.0	6.0	2120	2312.0	112.0	112.0	1.0
	78.0	1.0	5.0	0800-0859	30.0	948.0	1039.0	4.0	925	1043.0	78.0	78.0	1.0

# Visualizing the Delays



- Highest number of flight delayed occurred was in July

# Total Delays, in Years



- June 2015 had the most delays equivalent to 13 years.

# Delay per flight In the US 2015

June is most delayed month with 16 minutes average delay per flight.



# Enrich the Data - My thanks to Russell Jurney

## Plane Information (Tails File) [https://github.com/rjurney/Agile\\_Data\\_Code\\_2/tree/master/data](https://github.com/rjurney/Agile_Data_Code_2/tree/master/data)

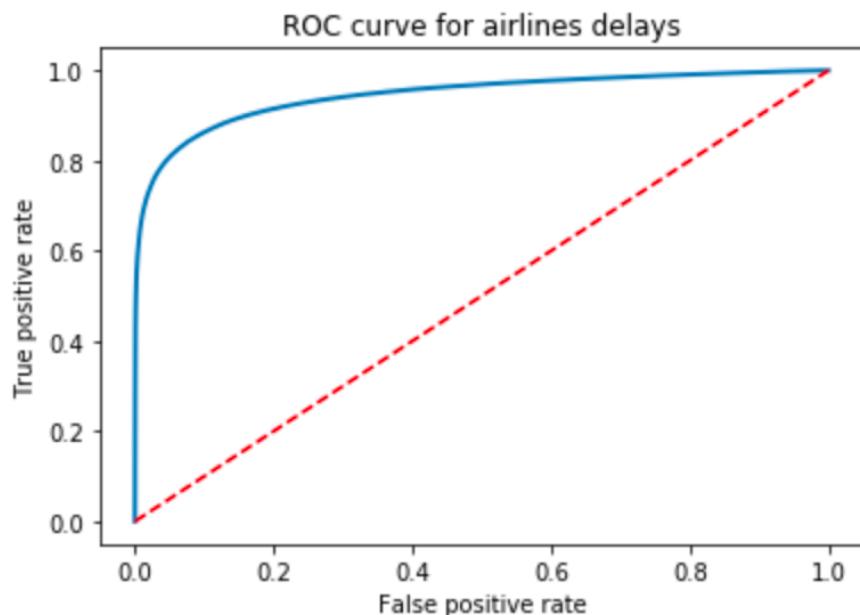
TailNum	engine_manufacturer	engine_model	manufacturer	mfr_year	model	owner	owner_state	serial_number	
0	N933EV	GE	CF34 SERIES	BOMBARDIER INC	2005	CL-600-2B19	DELTA AIR LINES INC	GEORGIA	8022
1	N917WN	CFM INTL	CFM56-7B24	BOEING	2008	737-7H4	SOUTHWEST AIRLINES CO	TEXAS	36624
2	N438WN	CFM INTL.	CFM56 SERIES	BOEING	2003	737-7H4	SOUTHWEST AIRLINES CO	TEXAS	29833
3	N283VA	CFM INTL	CFM56-5B4/3	AIRBUS	2015	A320-214	VIRGIN AMERICA INC	CALIFORNIA	6787
4	N473UA	IVCHENKO	AL-25SERIES	AIRBUS INDUSTRIE	2001	A320-232	UNITED AIRLINES INC	ILLINOIS	1469

Idea: Join Delays and plane info on TailNum attributes



# Predict: Logistic regression

ROC AUC score = 0.9449876496456802



Confusion Matrix

	Predicted On-time	Predicted Late
Actually On-Time	1,200,455	18,127
Actually Late	79,031	189,930

Scores

	Precision	Recall	F1-Score	No Test Records
On-Time	94%	99%	96%	1,218,582
Delayed	91%	71%	80%	268,961
Overall	93%	93%	93%	1,487543

# Used in the project

- Jupyter Notebook -> Collection, Enrichment, Prediction
- Sklearn -> Prediction
- PostgreSQL → Enrichment
- AWS → Enrichment
- Tableau --> Visualization
- Python / Html / Flask → App
- PowerPoint → Presentation



- Thank you