



Flight Arrival Delay Prediction using Machine Learning

Ben Jackany
Data Scientist

June 14, 2021

AGENDA

- ❏ Introduction and problem statement
- ❏ Choice of 2019 data over 2020 for ML Models
- ❏ Data Gathering and Preprocessing
- ❏ Exploratory Data Analysis
- ❏ Machine Learning Models
- ❏ Model Evaluation and Selection
- ❏ Conclusion and Recommendations

Introduction and problem statement

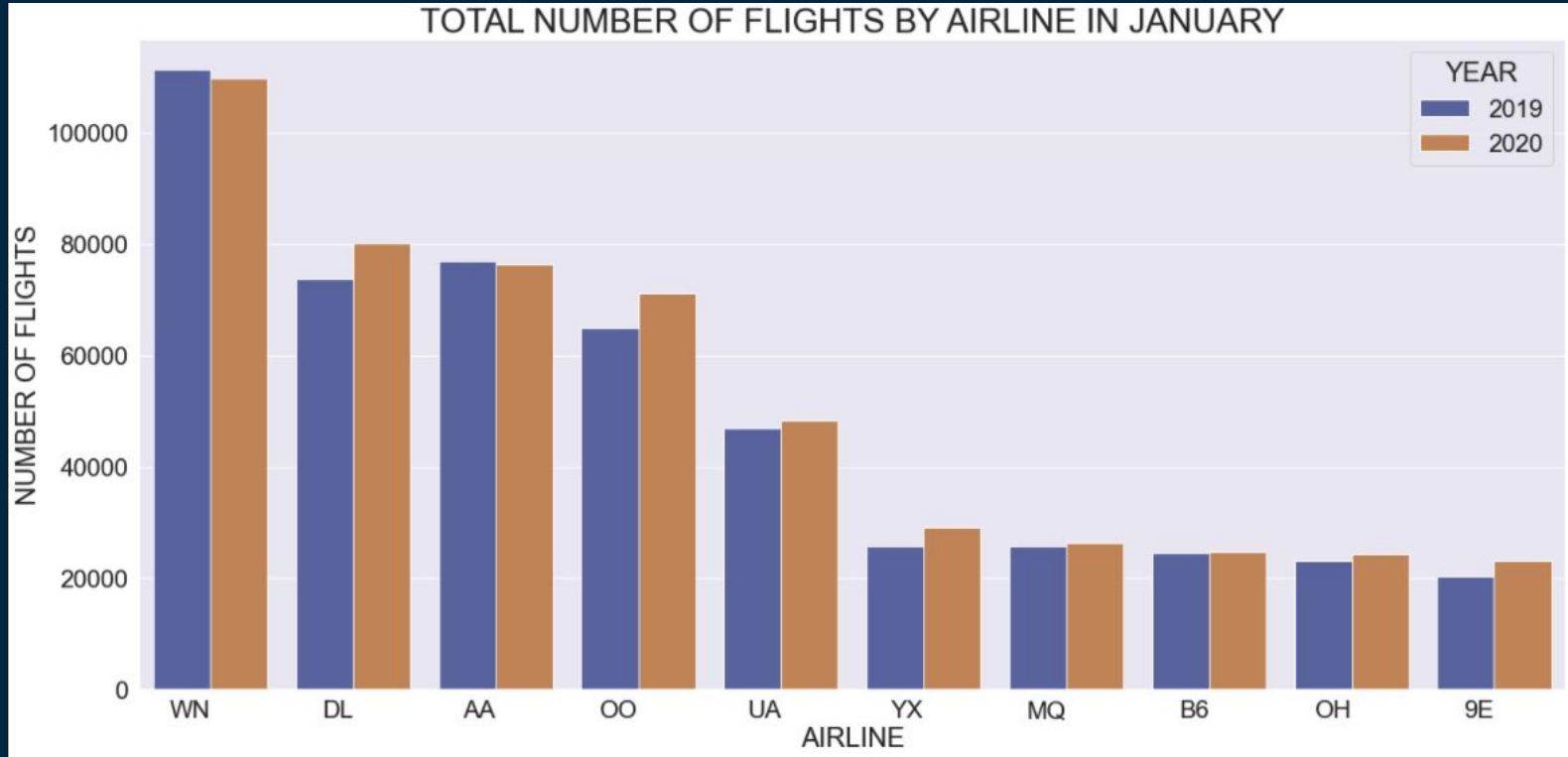
Flight Delays

One in Five flights arrive late at destination

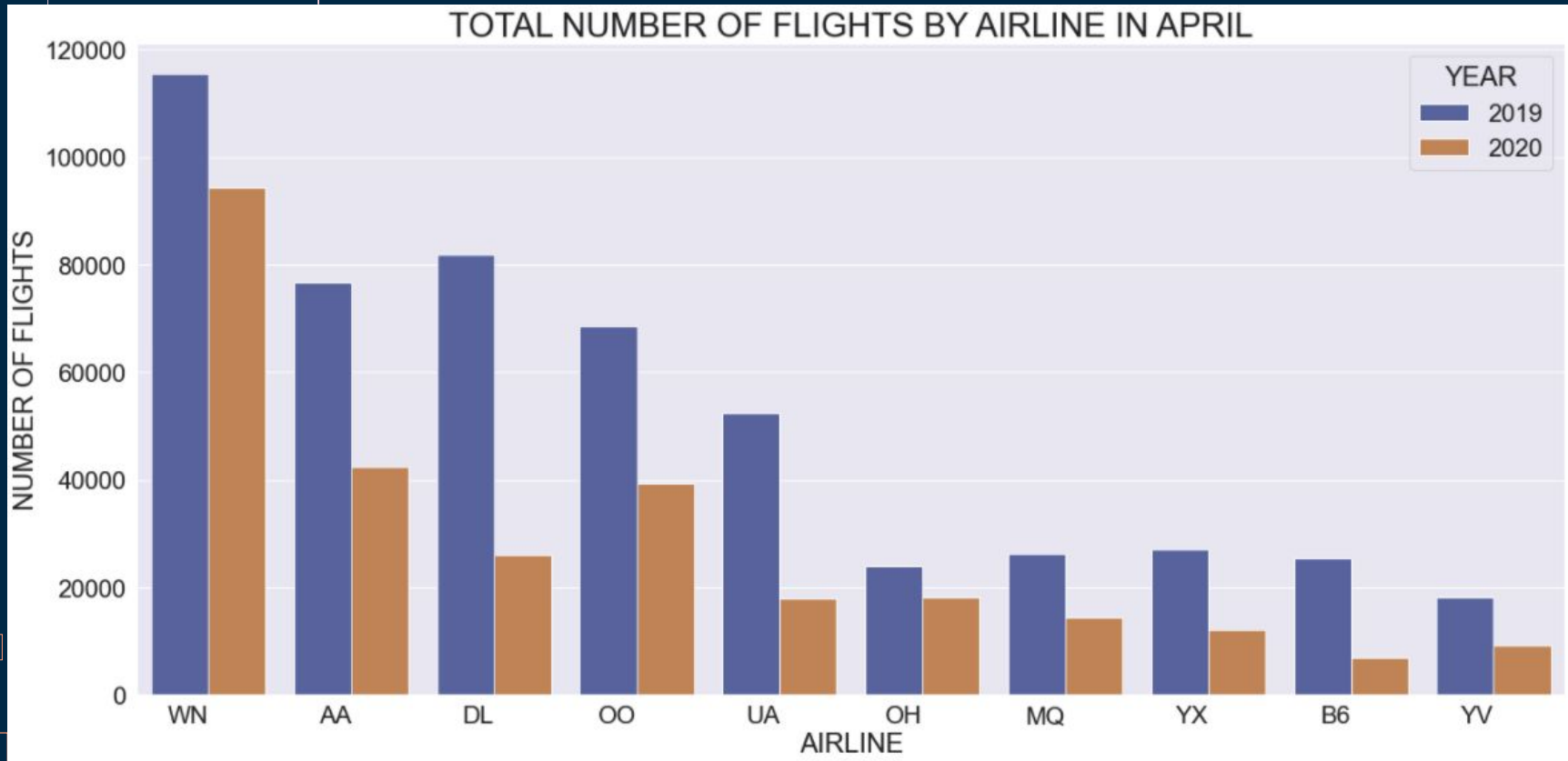
Inconvenience to airlines and passengers

Design a model that predict arrival delay for Delta Airlines

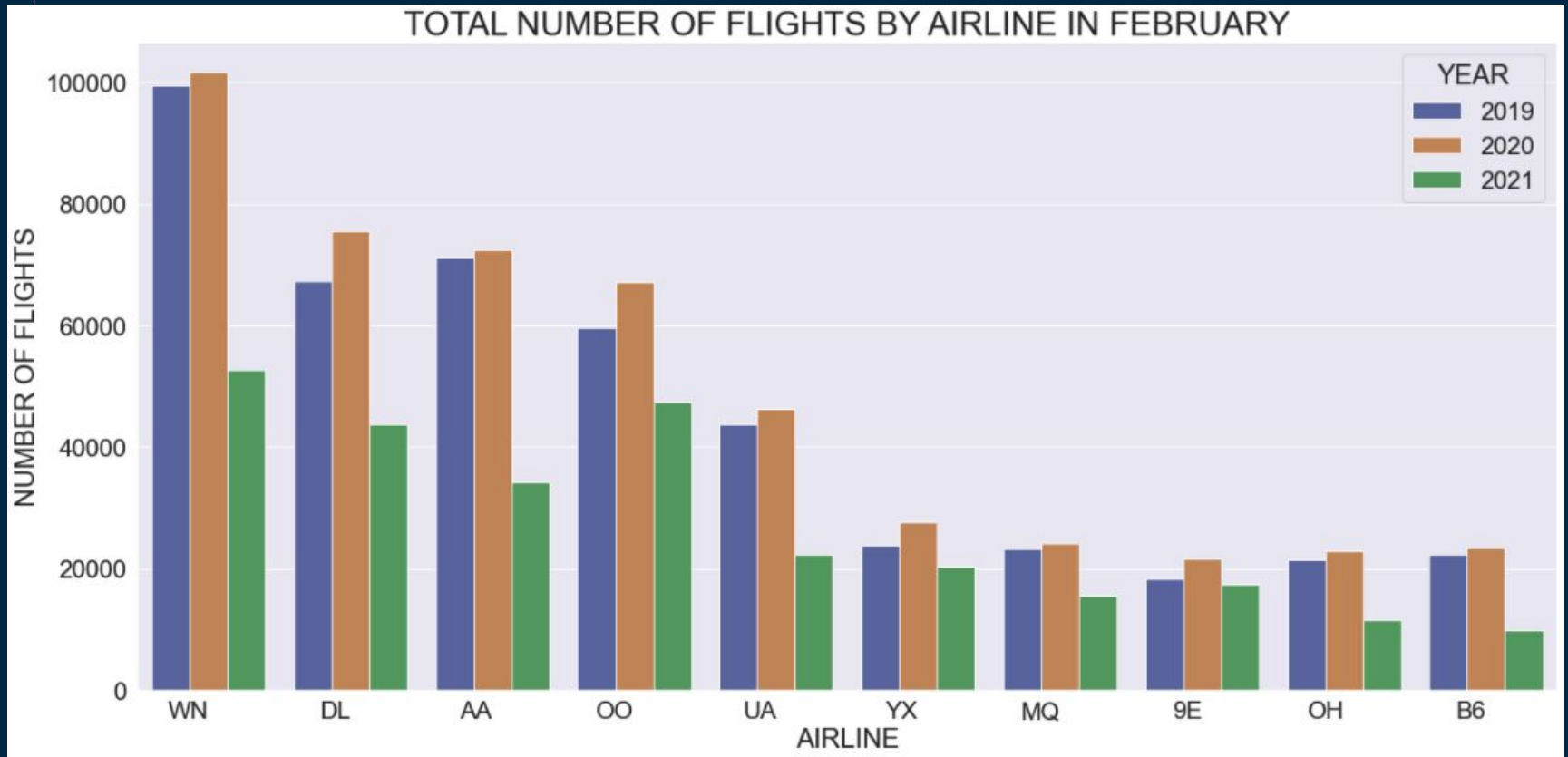
Choice of 2019 Data for ML models



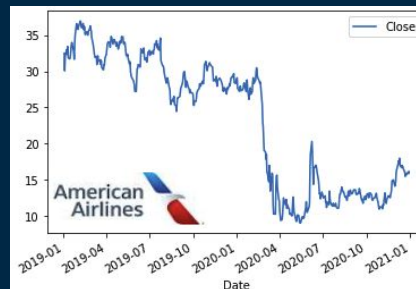
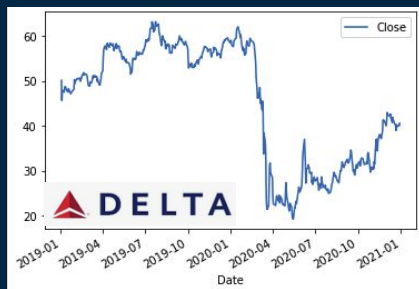
Choice of 2019 Data for ML models



Choice of 2019 Data for ML models



Choice of 2019 Data for ML models



Data Gathering and Preprocessing

Data Source:

2019-2020 and 2021 Data

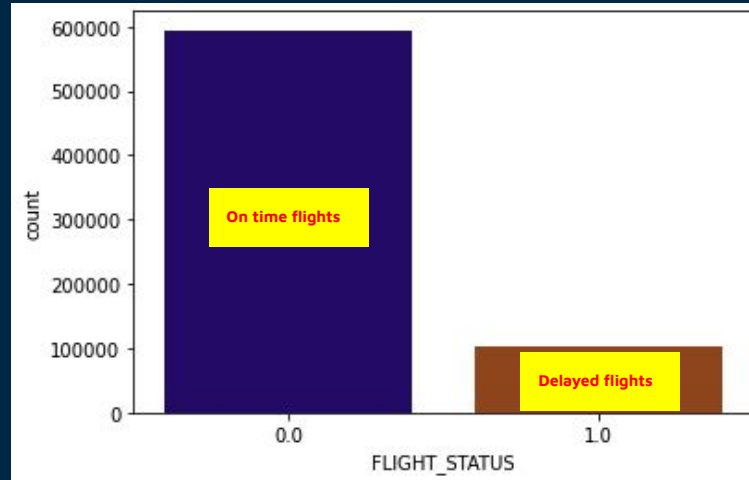


Bureau of Transportation Statistics

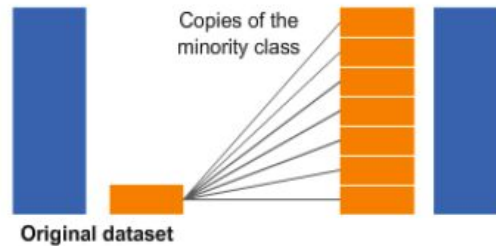
- ❑ 26 consecutive months of Data collected
- ❑ Over 12 million rows
- ❑ 17 Airlines
- ❑ 346 flight destinations
- ❑ 43 features

Exploratory Data Analysis

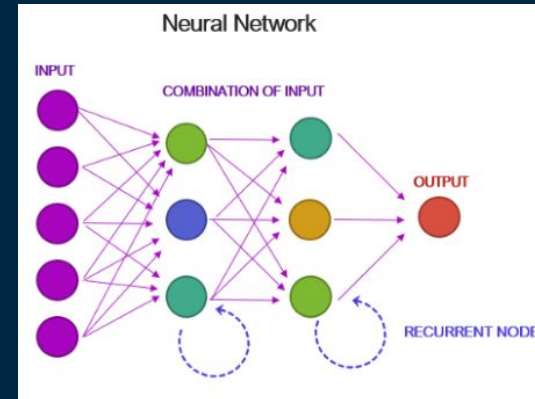
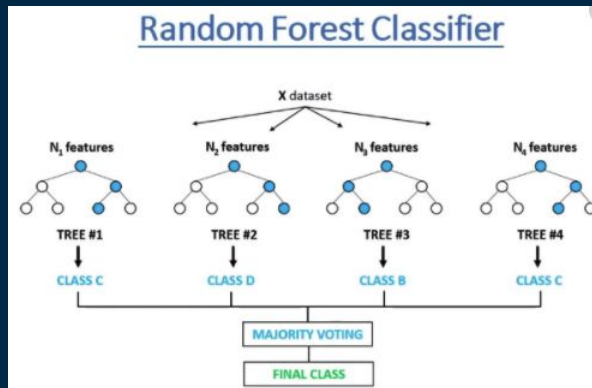
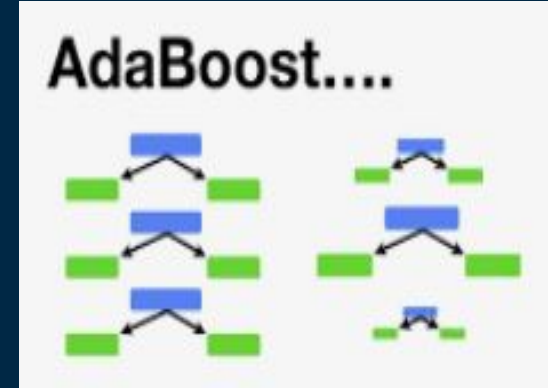
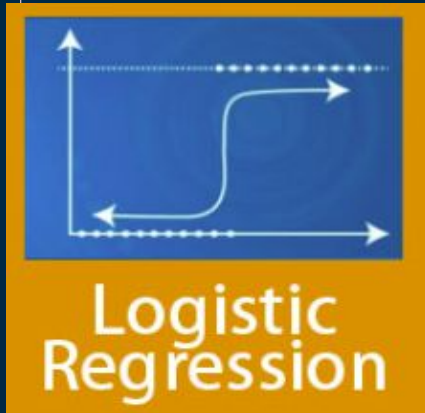
Data Distribution



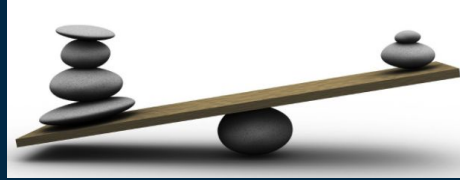
Oversampling



Machine Learning Models Used



Models Evaluation and Selection



Baseline F1-Score = 0.67

Machine Learning Algorithm	Balanced Dataset		
	F1-Score	Precision	Recall
1. Logistic Regression	0.63	0.64	0.62
2. Decision Trees	0.61	0.53	0.72
3. AdaBoost	0.61	0.53	0.72
4. Random Forest	0.71	0.67	0.76
5. Neural Networks	0.71	0.70	0.72



Conclusions and way forward

Complex and hard-to-predict business

Models performance improvement

Focus on one city destination

ACKNOWLEDGEMENTS

JEFF HALE

THANK
YOU!

JACOB KOEHLER

THANK
YOU!

ERIC BAYLESS

THANK
YOU!

THANK YOU!