Rajwant Kaur, Philipp Werdenbach, Oliver D. Mücke

Table of Contents

Introduction

Why are flights delayed?

- Aircraft delay from previous flight
- Extreme weather conditions
- Air traffic control restrictions
- Waiting for crew or staff strike
- Mechanical issues
- Bird strike etc…

Objective

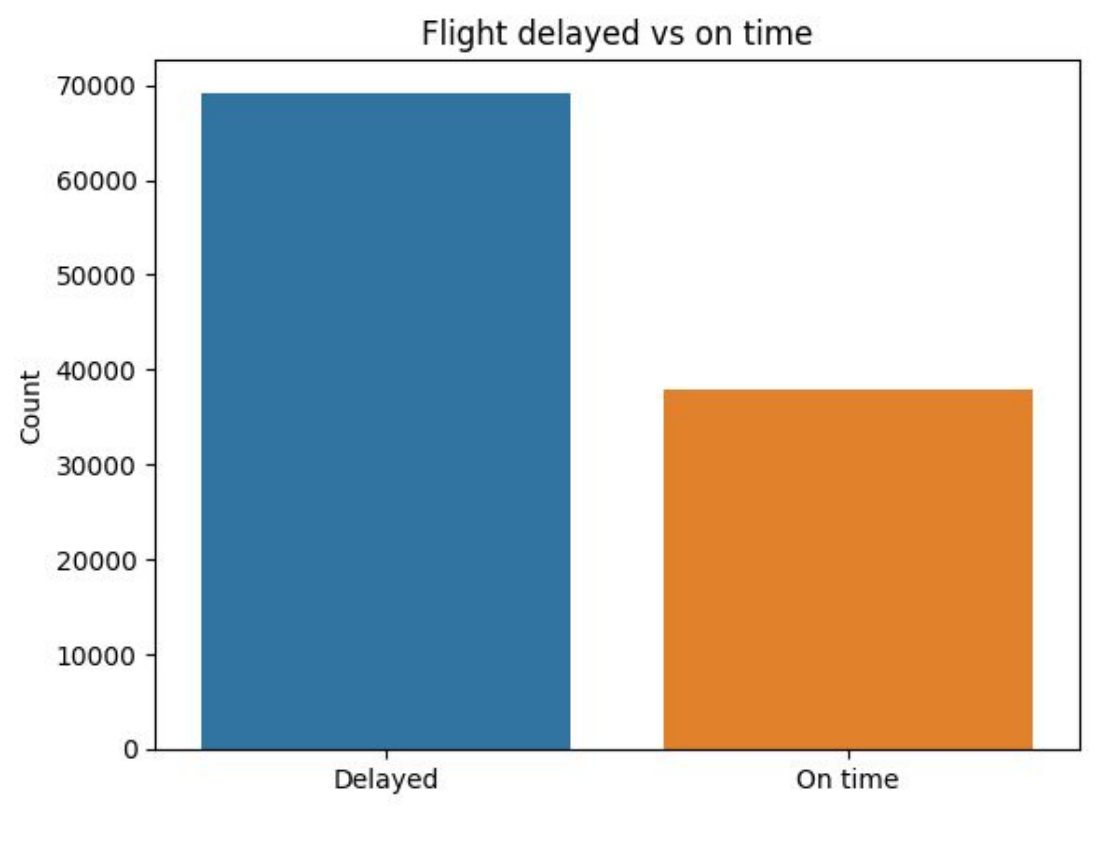
Goal:

To predict the estimated duration of **flight delay in minutes** for each flight mainly flying from or to the Tunis-Carthage International Airport

Why do we need predictions?

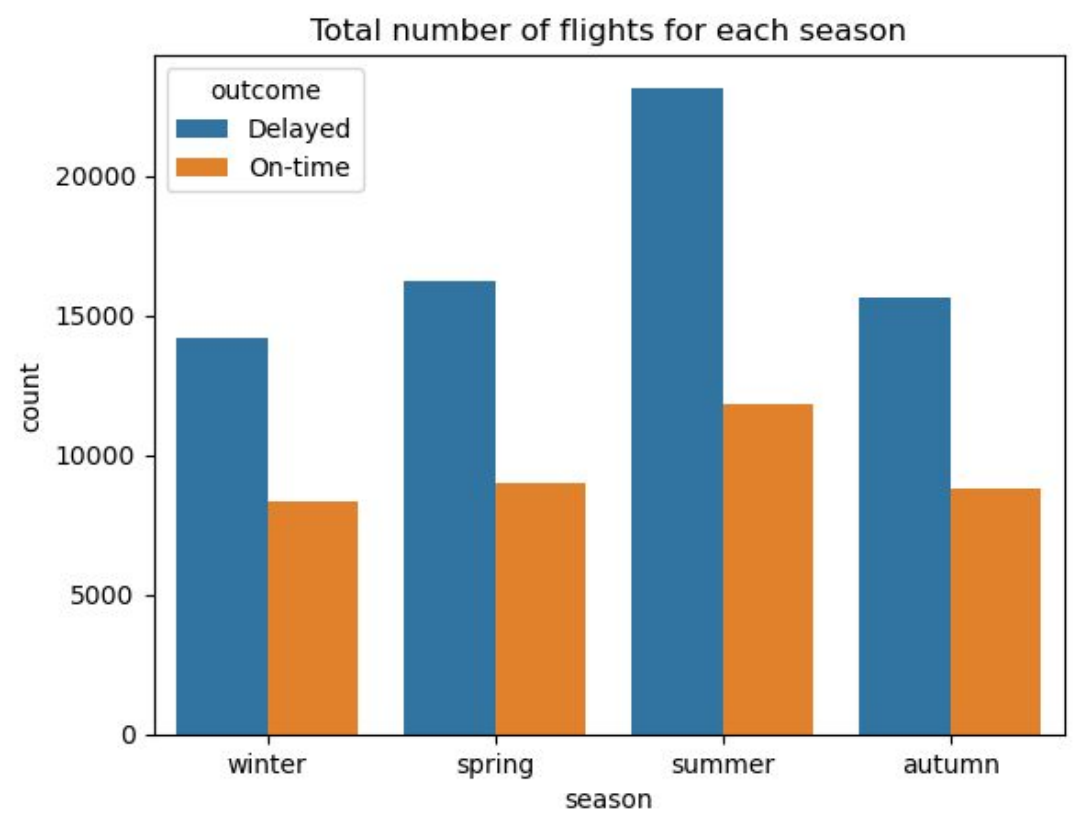
- To decrease of capital costs caused by reallocation of flight crews and aircraft
- To minimize the negative impact on passenger demand

Data Analysis



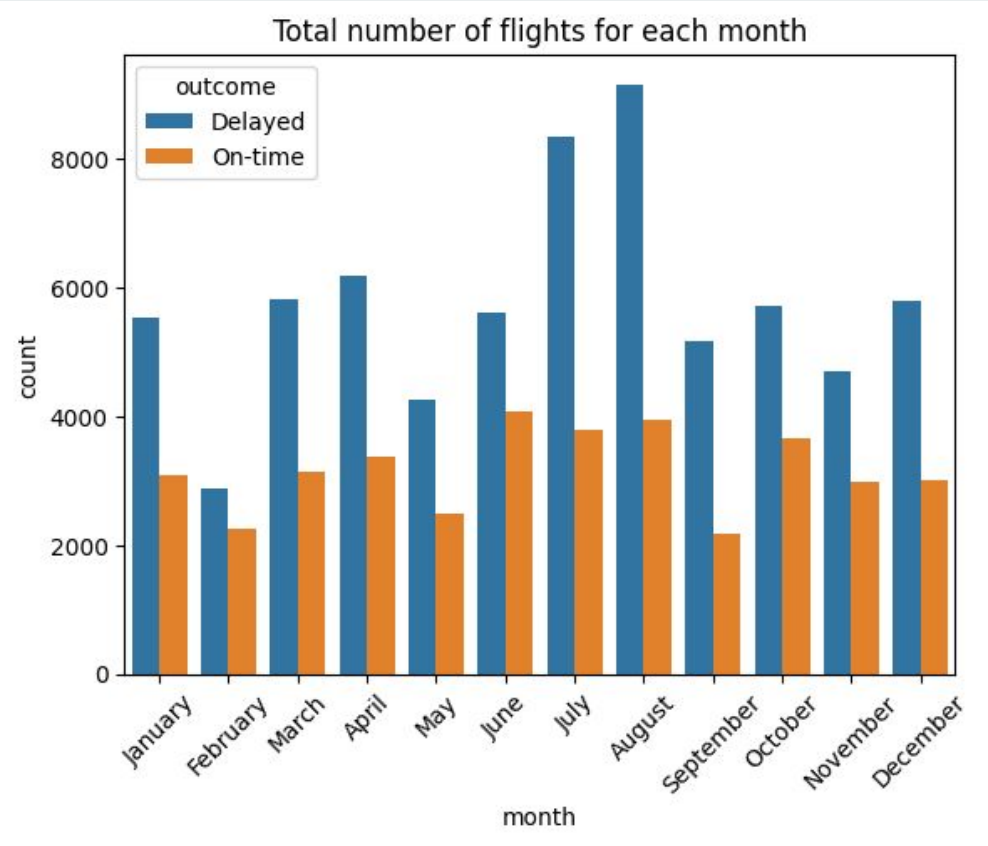
- years 2016 to 2018
- ca. 100,000 samples
- 9 attributes

Data Analysis



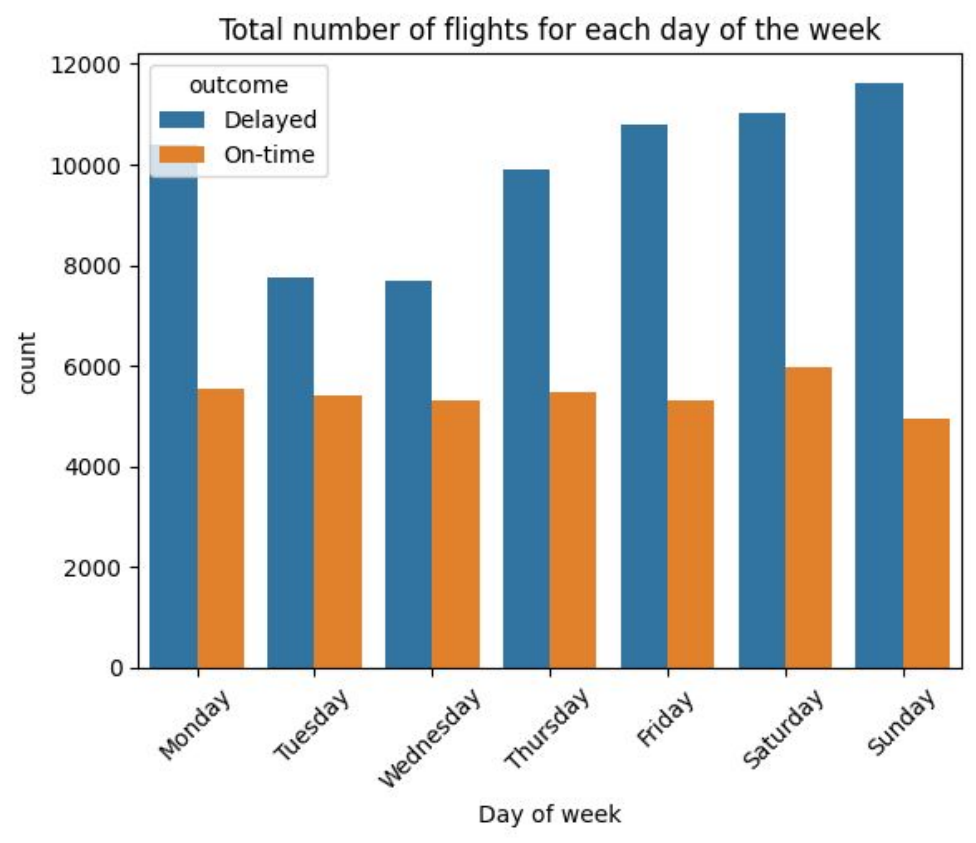
Absolute number of delayed flights is larger in the summer season

Data Analysis



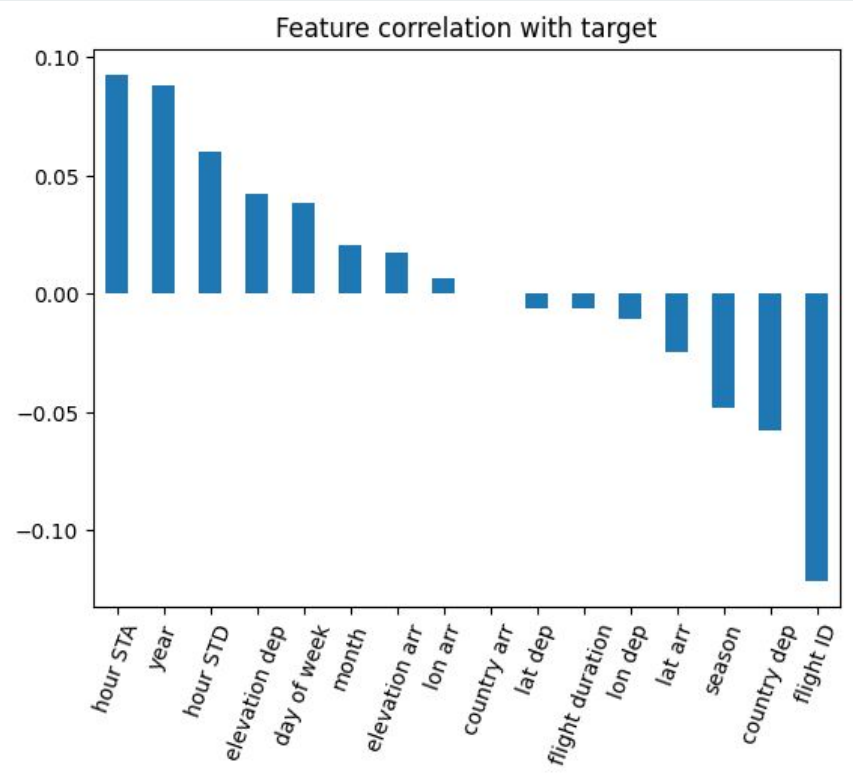
Absolute number of delayed flights is larger for months July and August

Data Analysis



Absolute number of delayed flights is smallest for Tuesdays and Wednesdays, increases afterwards

Data Analysis



Attributes are mainly uncorrelated and the correlation with true label is very small (<0.15)

Predictions

How good are Models?

- Lower the values of Error ⇩
- Closer the predicted values are to the test 
- Better is the model 👍

Our Models:

- Regression analysis performed
- Different regressors/models are chosen
- Error analysis

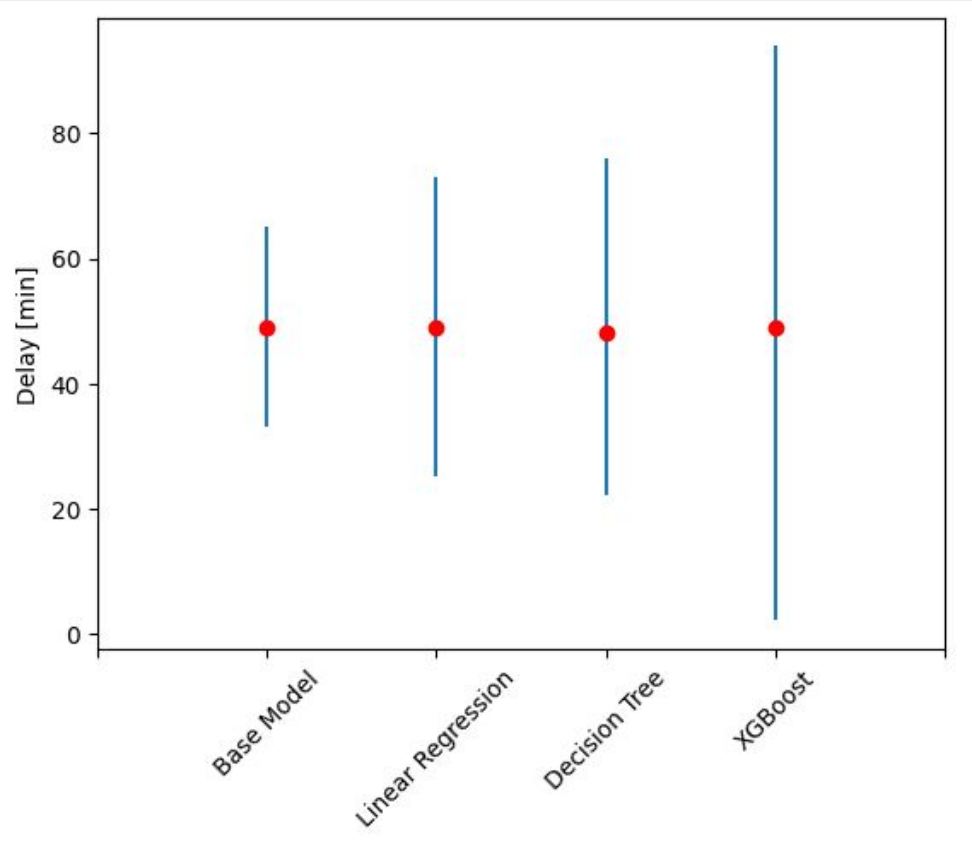
Predictions

Linear regression (base model)	RMSE (test): 112.60 min
Linear regression (optimized)	RMSE (test): 111.311 min
Decision Tree	RMSE (test): 110.675 min
XGBoost	RMSE (test): 105.045 min

Predictions - Average

Actual delay	48 min
Linear Regression (base model)	49 min (+/- 16 min)
Linear Regression (optimized)	49 min (+/- 24 min)
Decision Tree	49 min (+/- 27 min)
XGBoost	48 min (+/- 46 min)

Predictions - Average



Conclusions

- Predictions contain large errors
- Attributes are no good indicator for delay
- More meaningful attributes might be gathered