Flight Delay Prediction for Tunisair

ML-Project

Kevin Dietrich
Moritz Baur
Grese Berisha

Problem & Motivation

Problem ✈️

- Flight delays are **disruptive** and **costly**
- Passengers lose **time** ⏱️and **confidence**
- Airlines and airports face **financial losses** 💼 and reduced **operational efficiency**

Motivation 📈

- Tunisair aims to implement a **predictive solution** to **anticipate delays** and **mitigate their impact**

Project Objective 🎯

Goal:
Use machine learning to predict the **length of flight delays (in minutes)**.

Impact:
✔ Better scheduling
✔ Reduced operational inefficiencies
✔ Improved passenger satisfaction

Dataset & Evaluation

Data Source:
📂 Flight data provided by Zindi, consisting of a train/test format for model development

Prediction Target:
🕒 *Delay duration in minutes*

Performance Metric:
📉 *Root Mean Square Error (RMSE)*

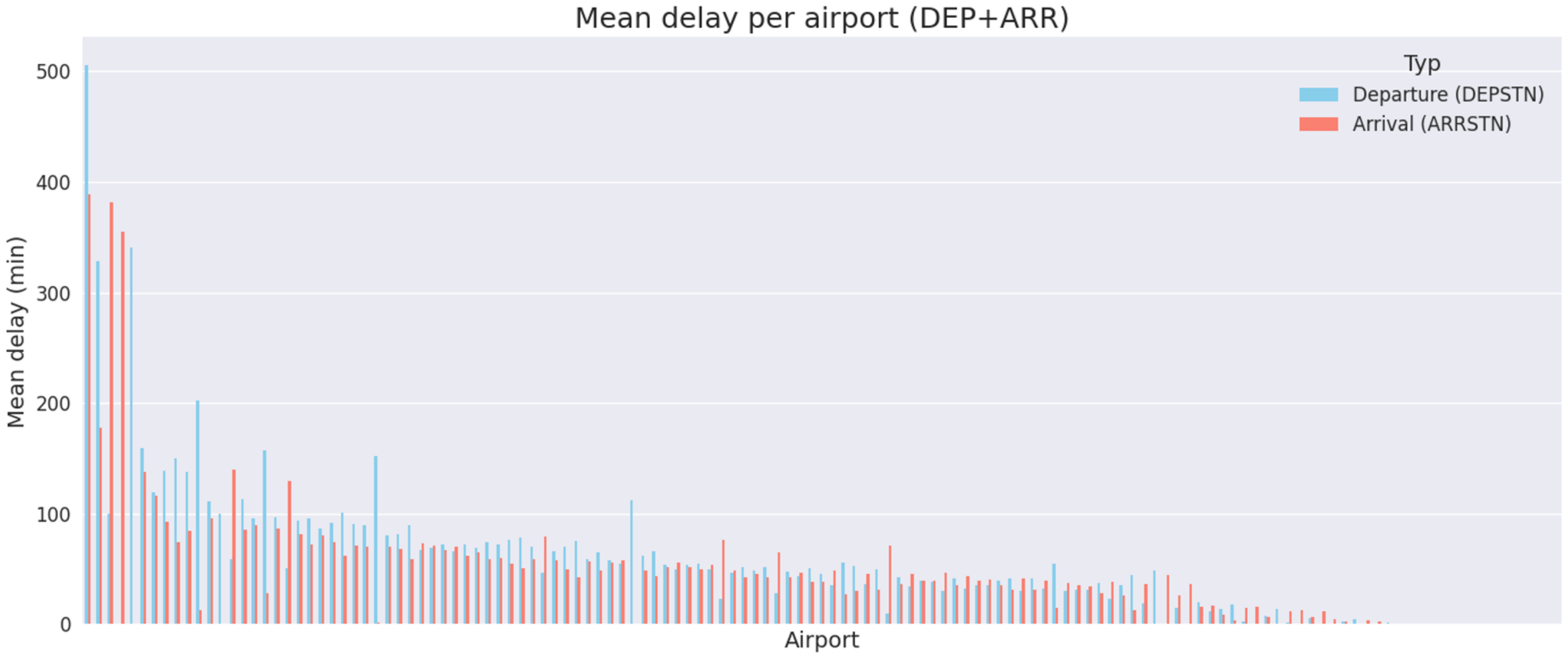
Exploratory Data Analysis (EDA)

🔍 **Initial Insights from EDA:**
We analyzed how delays are distributed based on:

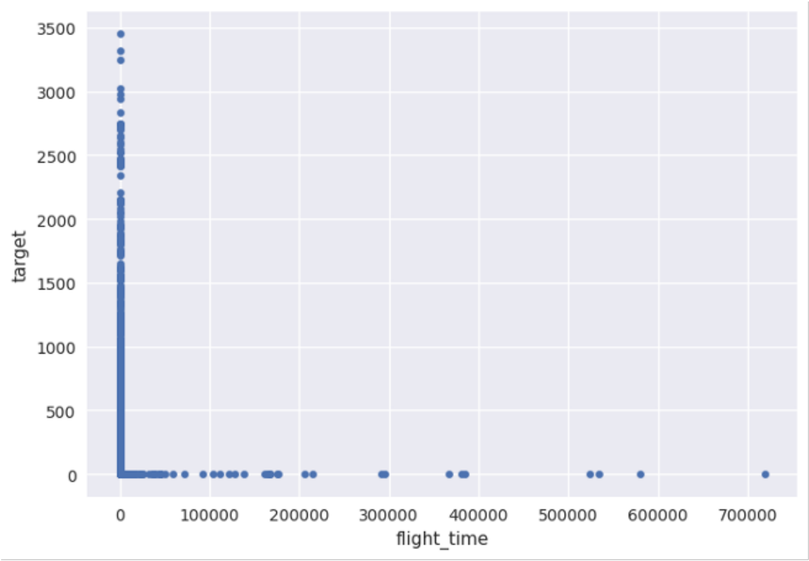
- **Departure airports**
- **Arrival airports**
- **Temporal trends across the years 2016–2018**

Data Table

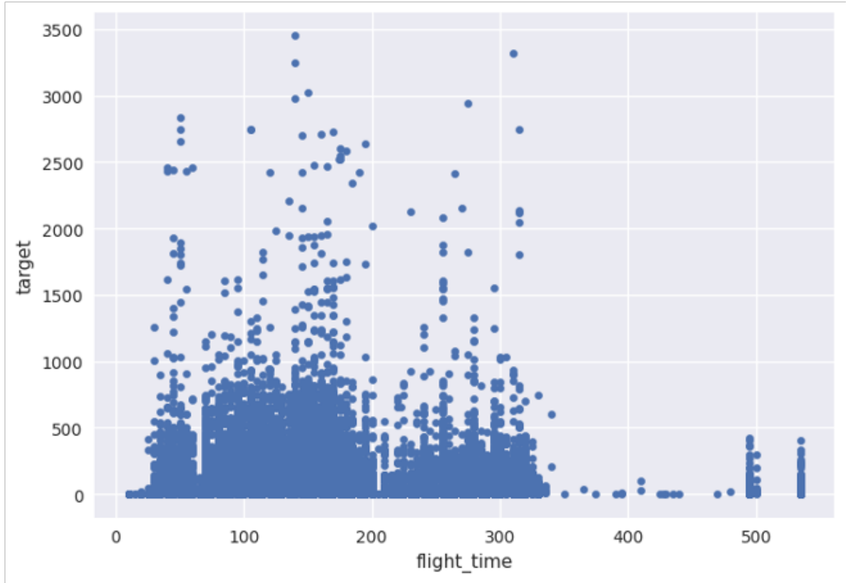
Column	Description
ID	Unique flight identifier
DATOP	Date of flight
FLTID	Flight number
DEPSTN	Departure point
ARRSTN	Arrival point
STD	Scheduled time of departure
STA	Scheduled time of arrival
STATUS	Flight status
AC	Aircraft code
target	Flight delay (min.)


Mean delay per airport (DEP+ARR)
Typ
Departure (DEPSTN)
Arrival (ARRSTN)
Mean delay (min)
500
400
300
200
100
0
Airport

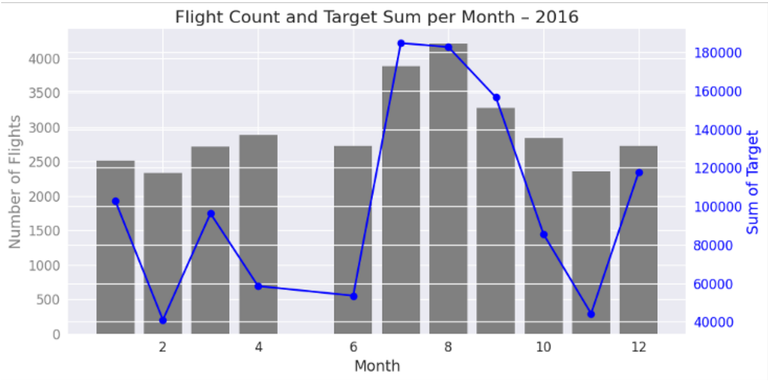
Deriving flight time from STD and STA

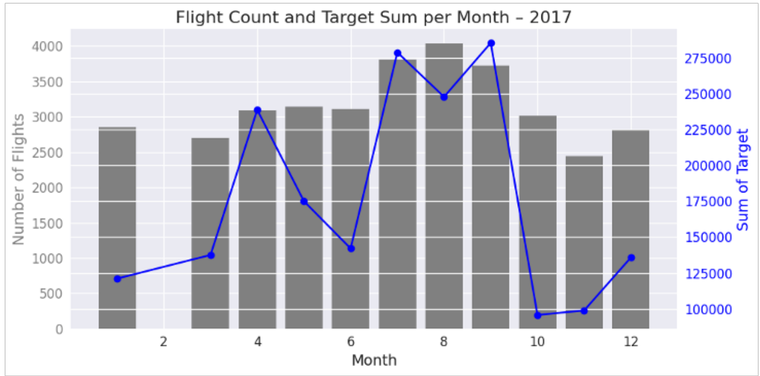


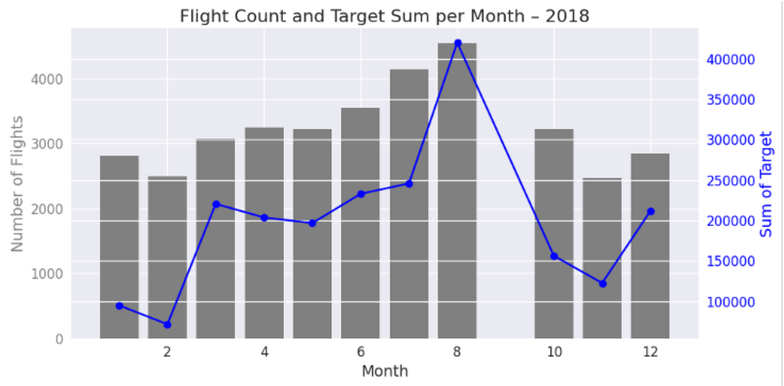
Removing what could be service flights?
Flights where departure and arrival airports are the same

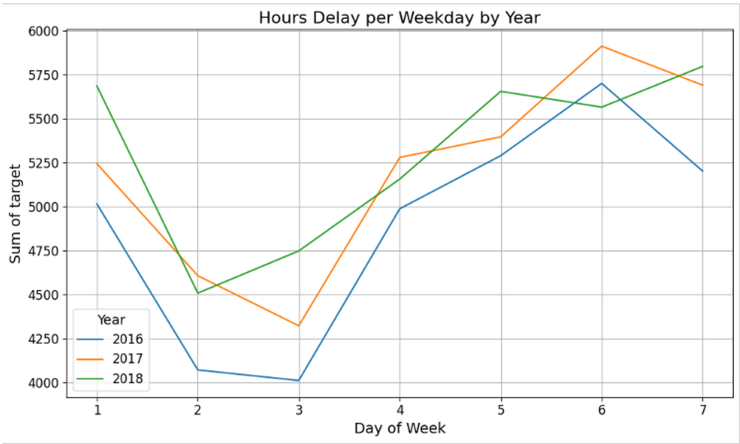


Dissecting DATOP into YEAR, month and day of the week







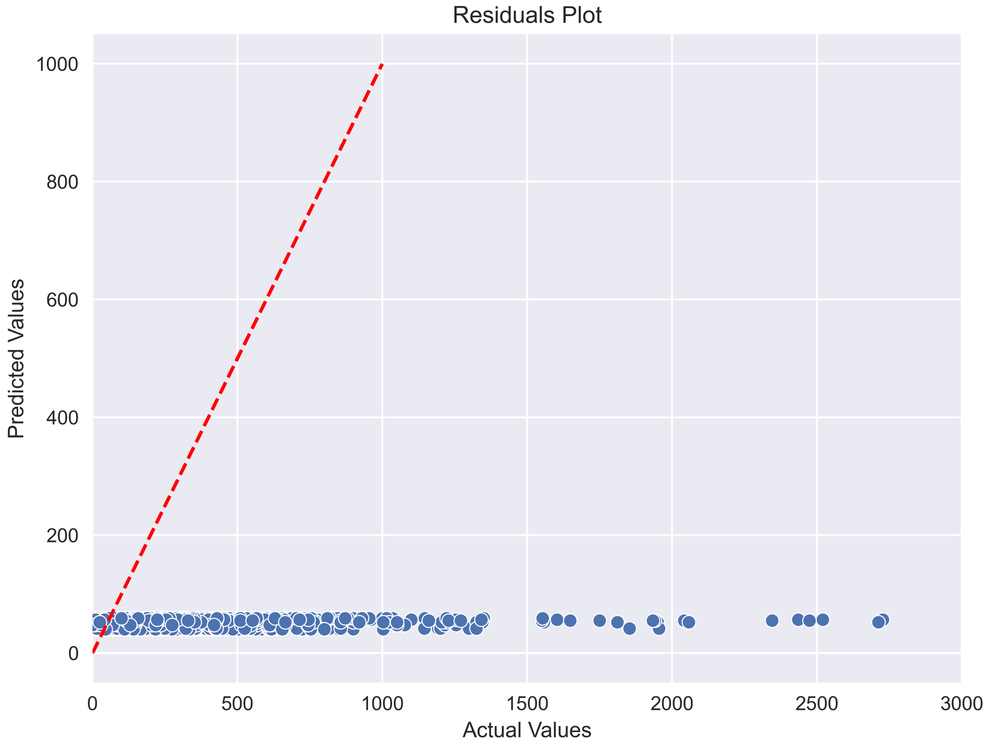


Baseline Model

Initial Approach:
A simple linear regression model using only the **day of the week (or aircraft code)** as the predictor.

Baseline Performance:
📊 RMSE ≈ **114.69**

R2 ≈ **3.01 %**



ML Model

Many categorical variables ...

... but there is:

ML Model

Approach:
A regression model using the CatBoost method with the predictors:

Flight Status | Aircraft Code | Departure and Arrival Point | Year, month and weekday of Departure Time

Baseline Performance:
📊 RMSE ≈ **96.14 (< 100)**

🤖 R2 ≈ **30.48 %**

💬 *We’re happy to answer any questions and look forward to your feedback.*