# Flight Delay Predictions

Amirhossein Arefzadeh
Aria Azem
Amin Aghakasiri

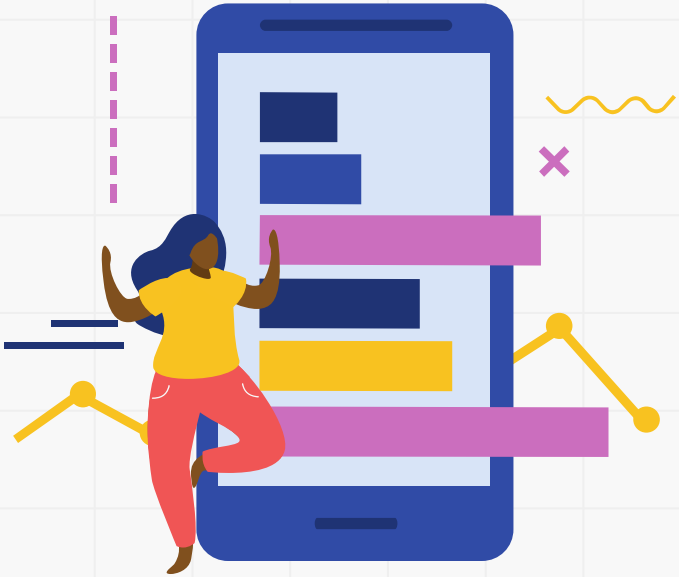
# Table of contents

# 01

# Introduction

# Why Flight Delay

- Per the Bureau of Transportation Statistics, 18.6 % of flights experienced delays (arrival > 15 minutes).
- 1.5 % of flights were canceled
- Flight delays can lead to missed connections, increased costs, and significant passenger inconvenience.



# Flight Delay Dataset

**5,819,079**

Number of flights

**322**

Number of airports

**54**

Number of state

# Flight Delay Dataset

**Source**
www.kaggle.com

**Structure**
36 features

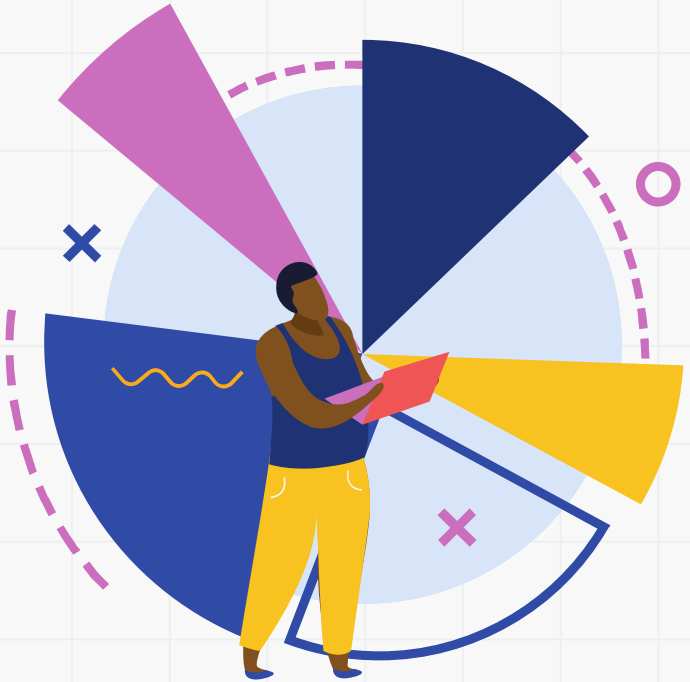
**Target**
Estimating the delay time for flights

# Flight Delay Dataset

- **Date** :
  day of the flight trip
- **Airline** :
  airline identifier
- **Origin_Airport** :
  starting airport
- **Destination_Airport** :
  destination airport
- **Distance** :
  distance between two airports
- **Air_Time** :
  the time in the sky

# 02

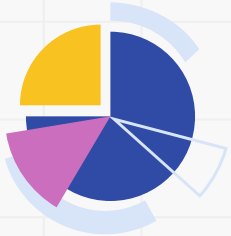
# Preprocessing



# Preprocessing

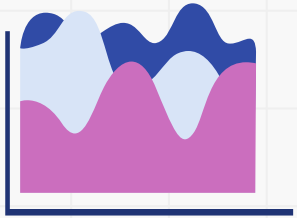
## Data cleaning

Removing duplicated and invalid data and add new features



## Label encoding

Encoding nominal and ordinal features accordingly

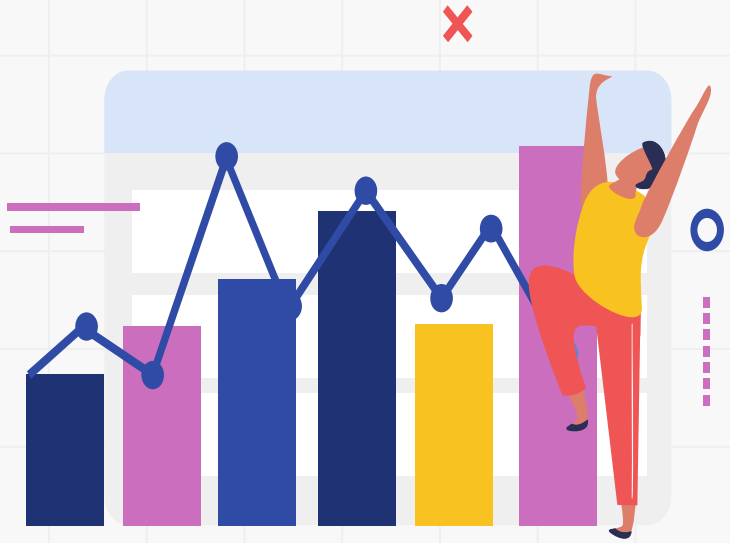


## Normalizing

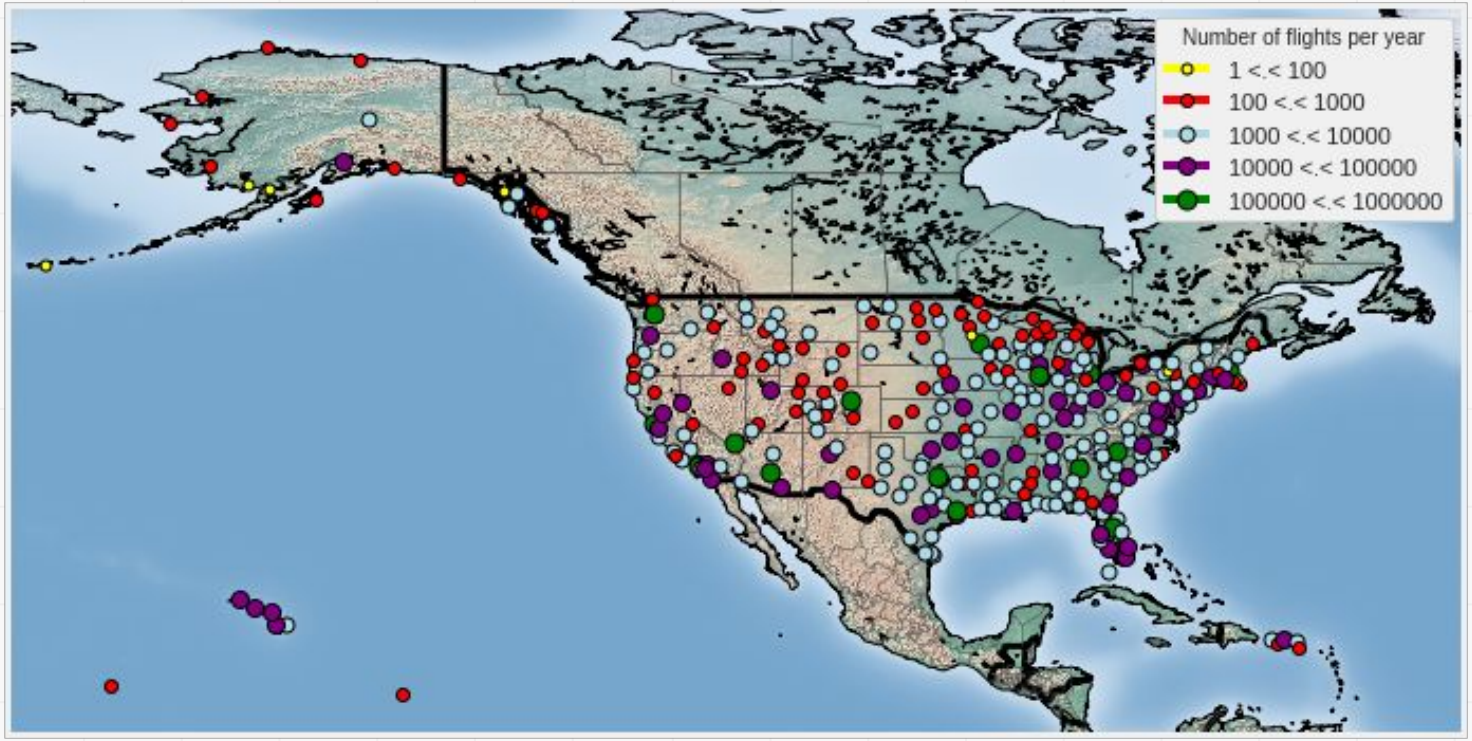
Scaling data for better ML convergence speed

# 03

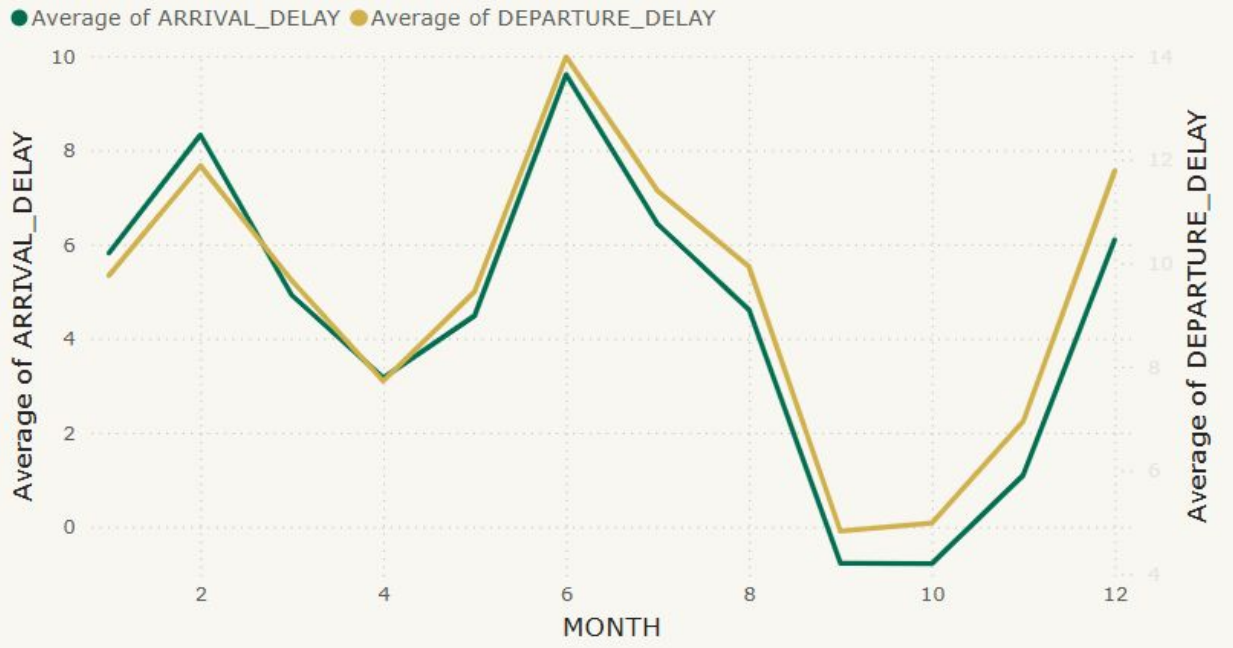
# Data Exploration



# 2015 U.S. Airport Flight Distribution

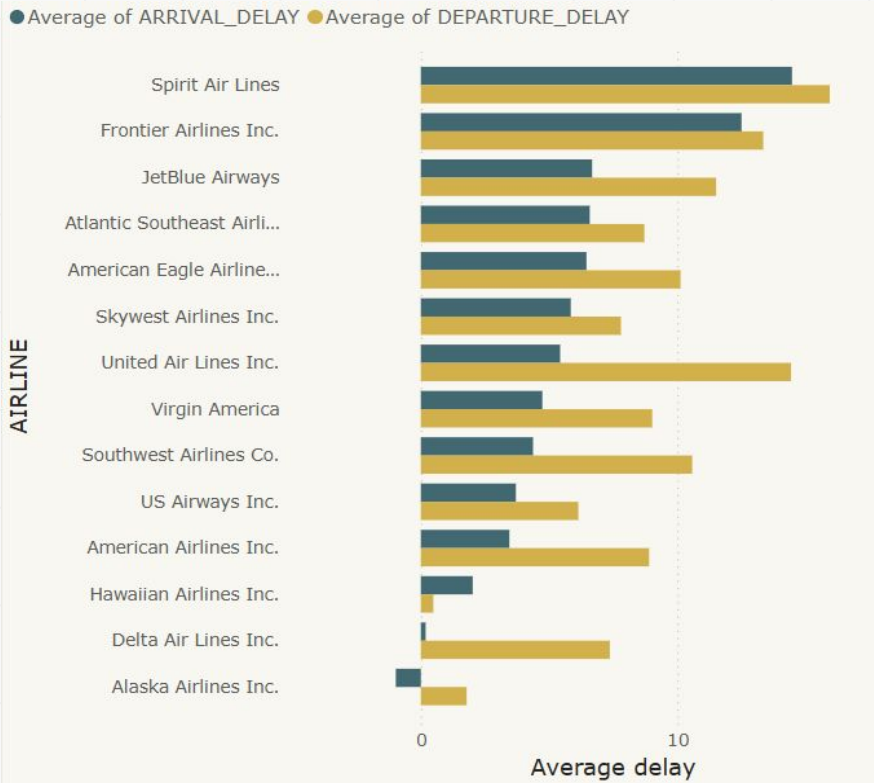


# Average of delay per month



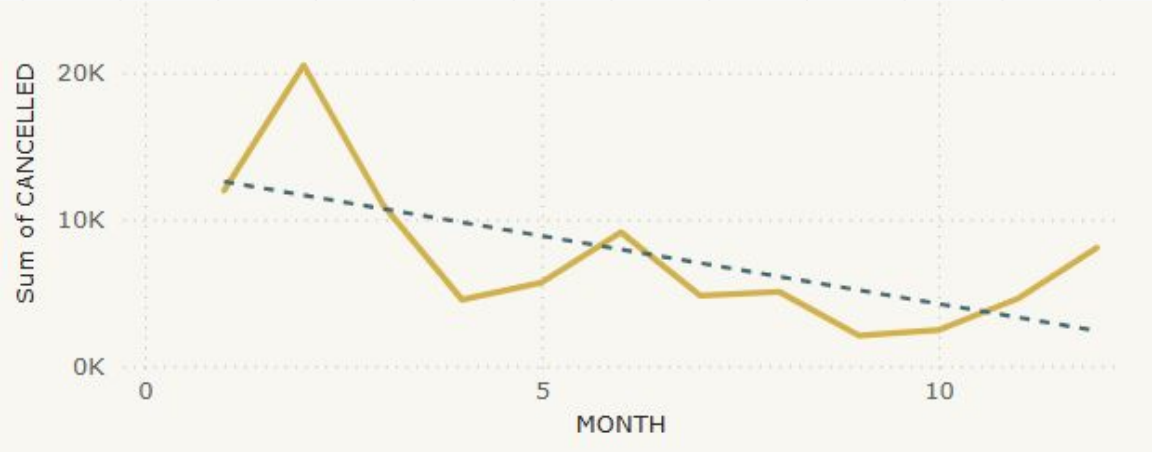

This chart shows the average arrival and departure delays across difference months. Seasonal trends, weather conditions and holidays may  contribute to variations in delays

# Average of delay by AIRLINE



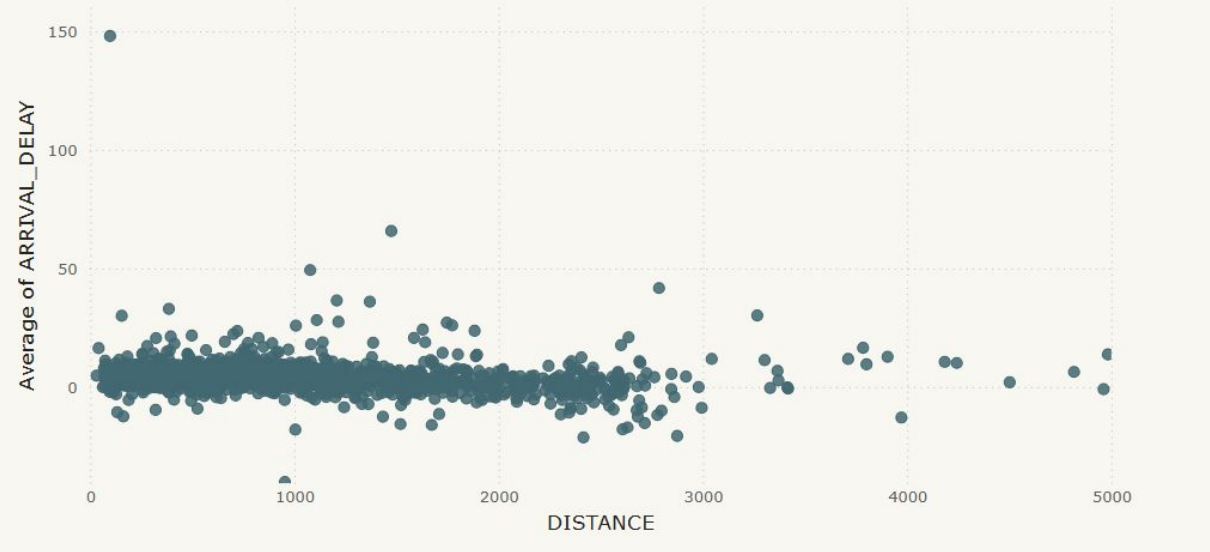

This visualization compares the average delays for different airlines. Some airlines may have better on-time performance due to fleet efficiency, airport hubs, or operational strategies.

# number of cancellation per month



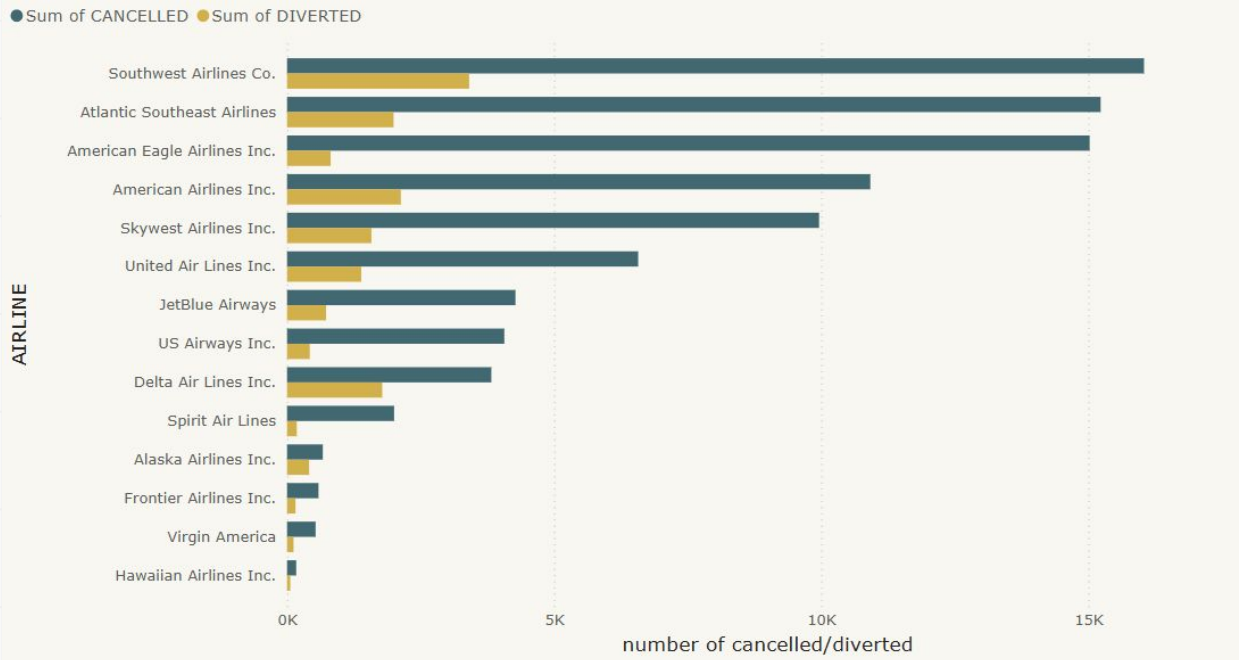

Displays the monthly trend of flight cancellations. Weather conditions, mechanical issues, and airline policies could influence these numbers.

# Average of DELAY by DISTANCE



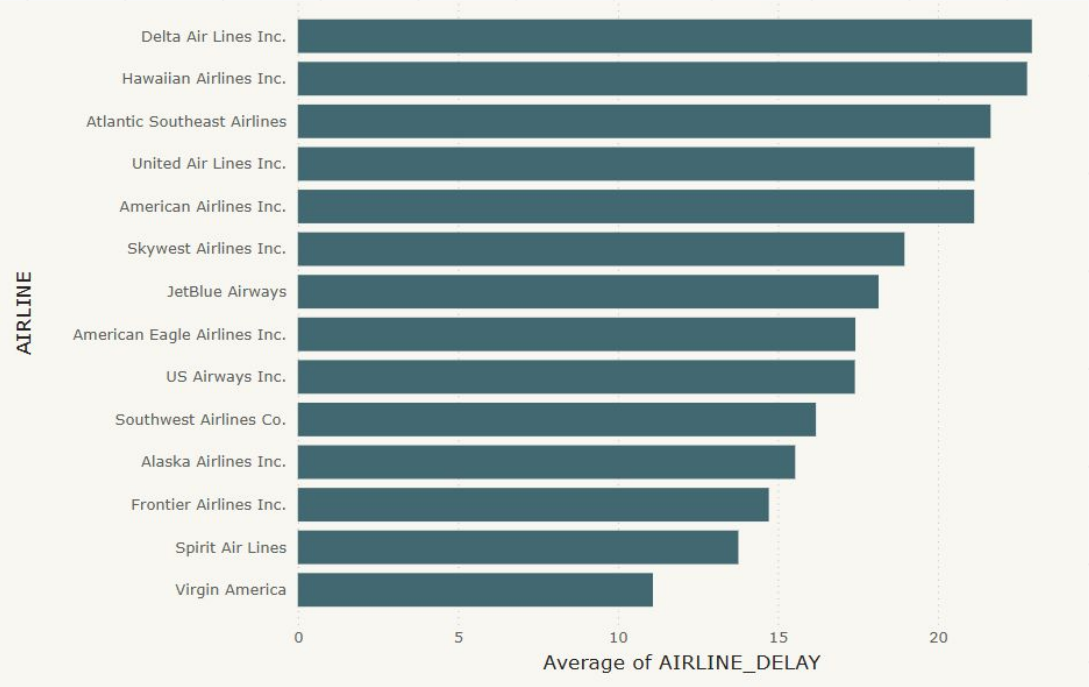

A plot of average arrival delay versus flight distance shows no clear correlation, indicating that delay times are not strongly related to how far a plane travels.

# Sum of CANCELLED and DIVERTED by AIRLINE



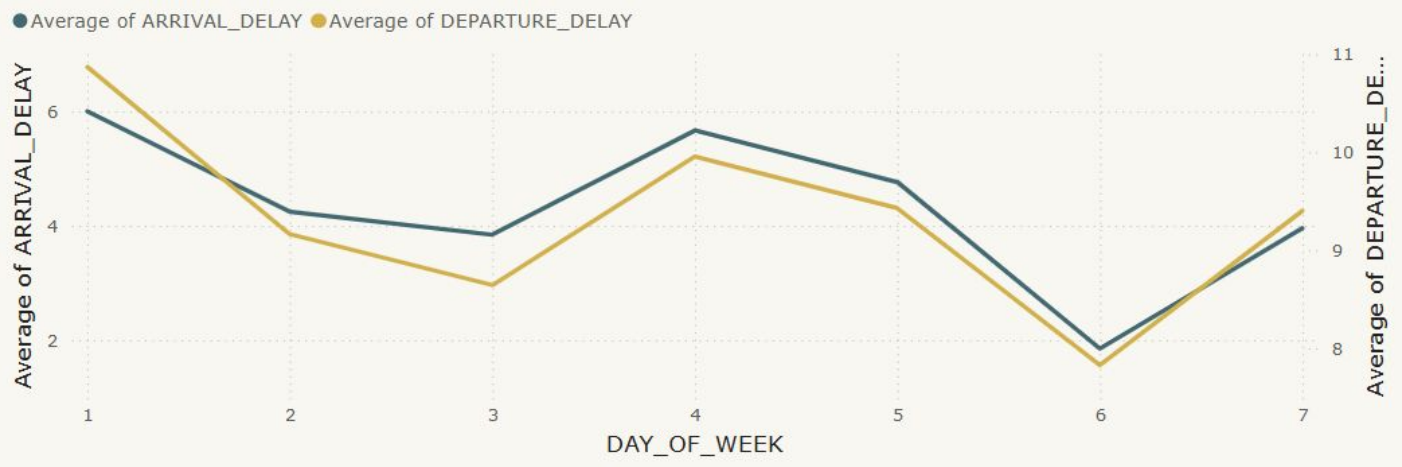

Compares airlines based on the number of cancelled and diverted flights. Some airlines may handle disruptions better than others

# Average of AIRLINE_DELAY by AIRLINE




This chart ranks airlines based on their overall flight delays. Operational efficiency and airport hubs play a role in these rankings.

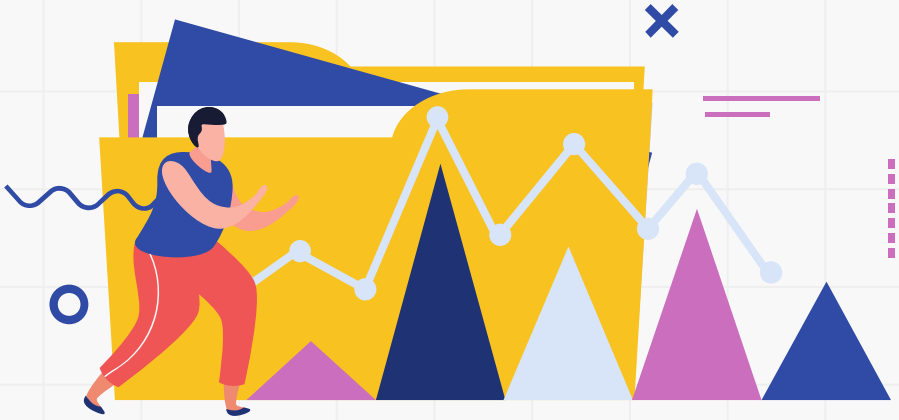
# Average of delay by DAY_OF_WEEK




Examines whether certain days have higher delays. Weekends or peak business travel days might see different delay patterns.

# 04

# Learning



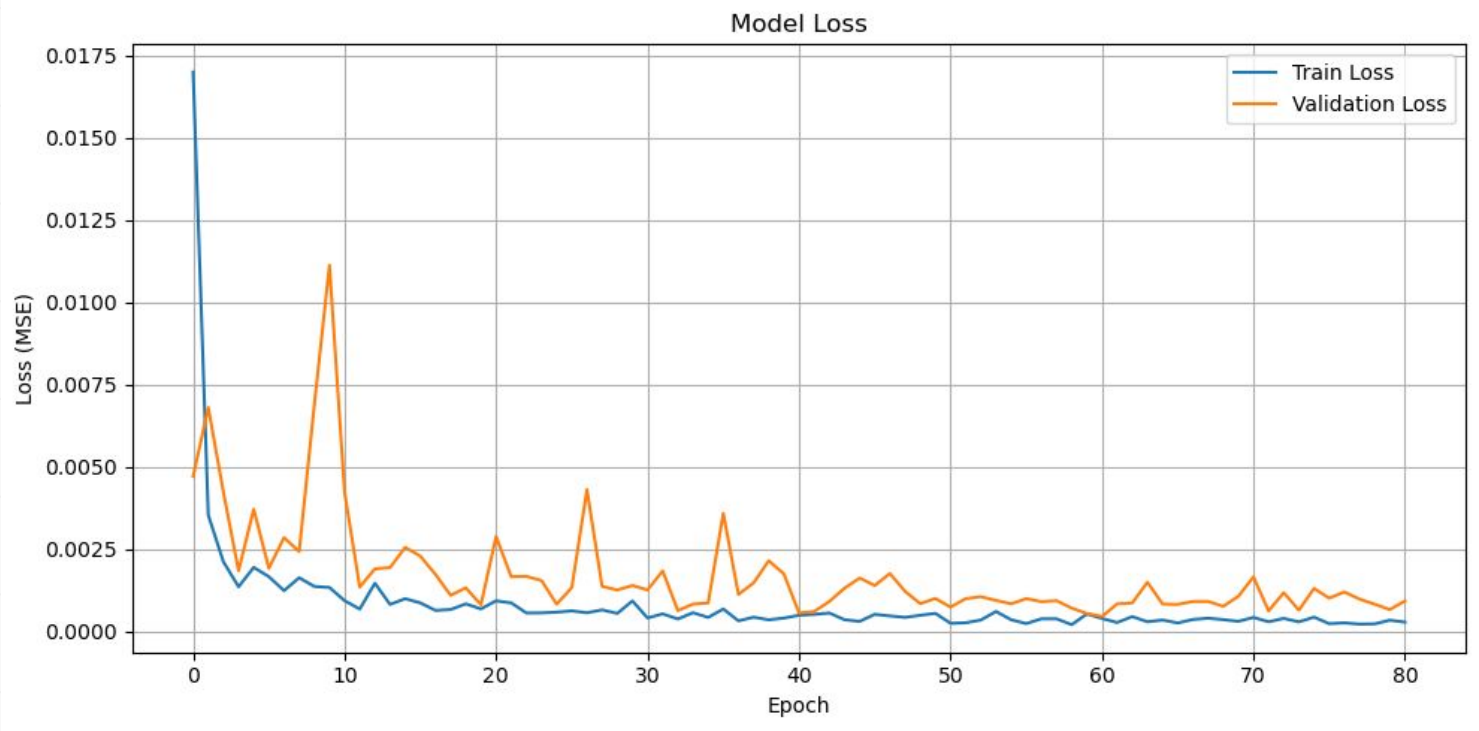
# Model Overview

| Layer (type) | Output Shape | Param # | Connected to |
|---|---|---|---|
| cat__DAY_OF_WEEK_i… (InputLayer) | (None, 1) | 0 | - |
| cat__WEEK_OF_YEAR_… (InputLayer) | (None, 1) | 0 | - |
| cat__IS_WEEKEND_inp (InputLayer) | (None, 1) | 0 | - |
| passthrough__AIRLI… (InputLayer) | (None, 1) | 0 | - |
| passthrough__ORIGI… (InputLayer) | (None, 1) | 0 | - |
| passthrough__DESTI… (InputLayer) | (None, 1) | 0 | - |
| passthrough__TAIL_… (InputLayer) | (None, 1) | 0 | - |
| cat__DAY_OF_WEEK_e… (Embedding) | (None, 1, 4) | 28 | cat__DAY_OF_WEEK… |
| cat__WEEK_OF_YEAR_… (Embedding) | (None, 1, 4) | 24 | cat__WEEK_OF_YEA… |
| cat__IS_WEEKEND_emb (Embedding) | (None, 1, 2) | 4 | cat__IS_WEEKEND_… |
| passthrough__AIRLI… (Embedding) | (None, 1, 8) | 112 | passthrough__AIR… |
| passthrough__ORIGI… (Embedding) | (None, 1, 50) | 15,650 | passthrough__ORI… |
| passthrough__DESTI… (Embedding) | (None, 1, 50) | 15,650 | passthrough__DES… |
| passthrough__TAIL_… (Embedding) | (None, 1, 50) | 220,600 | passthrough__TAI… |
| num_inp (InputLayer) | (None, 6) | 0 | - |
| flatten (Flatten) | (None, 4) | 0 | cat__DAY_OF_WEEK… |
| flatten_1 (Flatten) | (None, 4) | 0 | cat__WEEK_OF_YEA… |
| flatten_2 (Flatten) | (None, 2) | 0 | cat__IS_WEEKEND_… |
| flatten_3 (Flatten) | (None, 8) | 0 | passthrough__AIR… |
| flatten_4 (Flatten) | (None, 50) | 0 | passthrough__ORI… |
| flatten_5 (Flatten) | (None, 50) | 0 | passthrough__DES… |
| flatten_6 (Flatten) | (None, 50) | 0 | passthrough__TAI… |
| concatenate (Concatenate) | (None, 174) | 0 | num_inp[0][0], flatten[0][0], flatten_1[0][0], flatten_2[0][0], flatten_3[0][0], flatten_4[0][0], flatten_5[0][0], flatten_6[0][0] |
| dense (Dense) | (None, 128) | 22,400 | concatenate[0][0] |
| dense_1 (Dense) | (None, 64) | 8,256 | dense[0][0] |
| dense_2 (Dense) | (None, 32) | 2,080 | dense_1[0][0] |
| regression_output (Dense) | (None, 1) | 33 | dense_2[0][0] |

# 284,837

total number of parameters

# Test vs. Validation Loss per Epoch



# Model Evaluation

```
----- Test metrics
MSE    : 0.000372
RMSE   : 0.019299
MAE    : 0.007517
R²     : 0.9996
SMAPE  : 4.3092%
```

```
----- Test metrics (original scale)
MSE    : 0.590975
RMSE : 0.7687
MAE   : 0.2994
R²    : 0.9996
SMAPE  : 12.6284%
```

# Sample Predictions

| | flight_id | actual_arrival_delay |
|---|---|---|
| 0 | 11 | -30.0 |
| 1 | 27 | -3.0 |
| 2 | 30 | 2.0 |
| 3 | 68 | -4.0 |
| 4 | 80 | 9.0 |
| 5 | 89 | -18.0 |
| 6 | 134 | -6.0 |
| 7 | 169 | 27.0 |
| 8 | 194 | 11.0 |
| 9 | 213 | 3.0 |

| | flight_id | predicted_arrival_delay |
|---|---|---|
| 0 | 11 | -30.336096 |
| 1 | 27 | -3.222282 |
| 2 | 30 | 1.942082 |
| 3 | 68 | -4.065832 |
| 4 | 80 | 9.057667 |
| 5 | 89 | -18.058674 |
| 6 | 134 | -5.812082 |
| 7 | 169 | 26.775656 |
| 8 | 194 | 10.948182 |
| 9 | 213 | 3.181458 |

# Thanks!

**Do you have any questions?**

amirrfz2003@gmail.com
ariaazem@gmail.com
kasiri.amin@gmail.com