On-Time Flight Performance Prediction

IDS 561 | Group 24 | Spring 2022

Team

Giovanni Alvin Prasetya
Anupriya Rastogi
Rajaram Ramesh

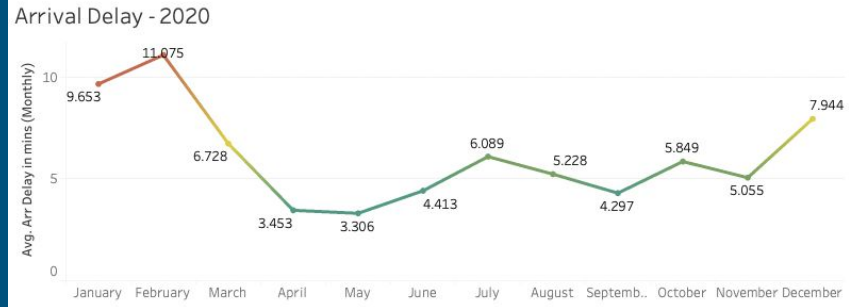
Problem Setting

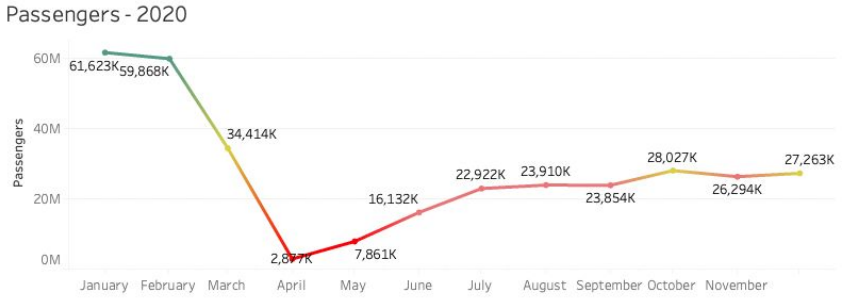
- Flight delays are something that every traveler has experienced at some point in their journeys.
- Being able to predict flight delays ahead of departure and arrival time can potentially reduce the last minute change of schedules.
- It can also help airlines make the necessary preparations in advance to ensure that they make the journey for their passengers as smooth as possible.

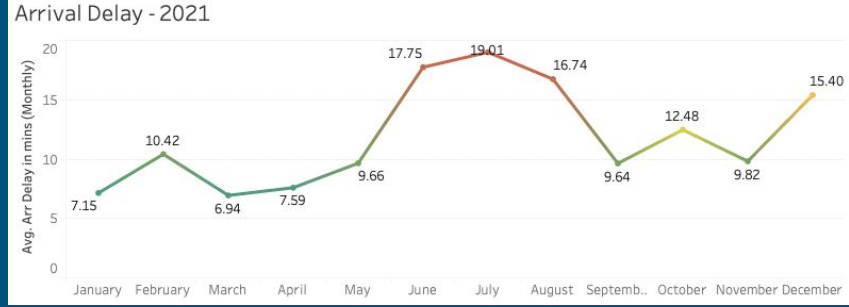
Data

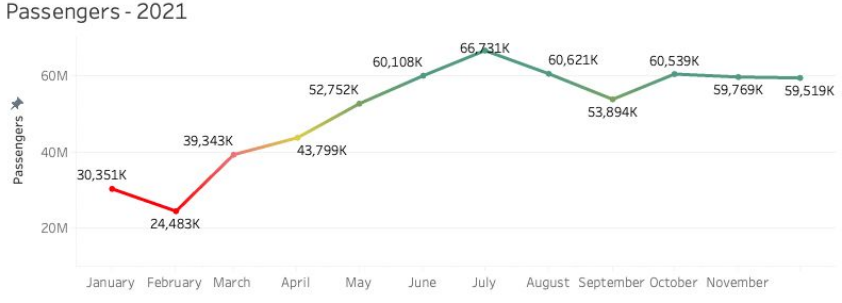
- Airline On-Time Performance (Primary Dataset)
  - Source: **Bureau of Transportation Statistics (BTS)**
  - Time Period: **January 2019 - December 2021**
  - Size of Dataset: **6.6GB | 18,105,787 Instances**
- Passengers
  - Source: **Bureau of Transportation Statistics (BTS)**
  - Time Period: **January 2019 - December 2021**
  - Dataset: **1,758,425,690 Passengers**
- Airline Employment Data
  - Source: **Bureau of Transportation Statistics (BTS)**
  - Time Period: **January 2019 - December 2021**
  - Dataset: **16 Airlines (e.g. United Airlines, Southwest Airlines, American Airlines, Delta Airlines)**
- COVID Data: (2020-2021)
  - Source: **Centers for Disease Control and Prevention (CDC)**
  - Time Period: **January 2020 - December 2021**
  - Dataset: **Daily Cases**

Exploratory Data Analysis & Visualization (Tableau)

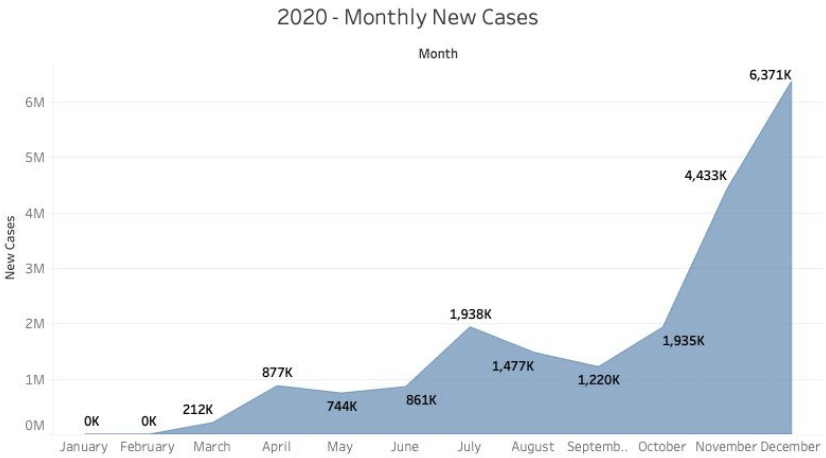


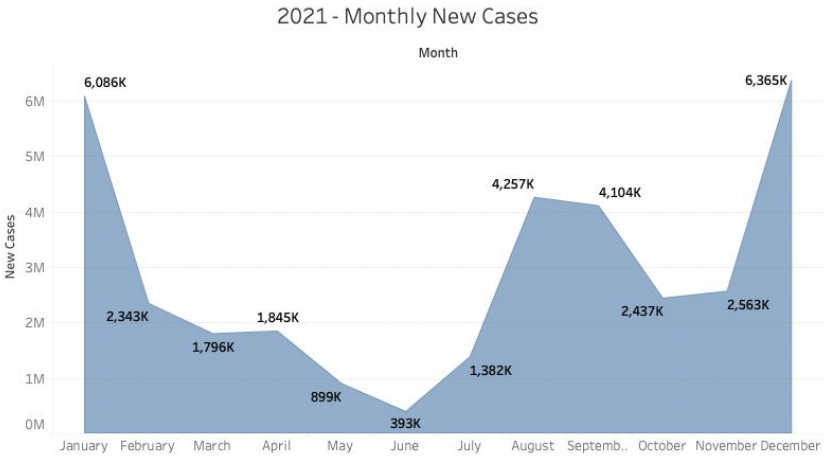


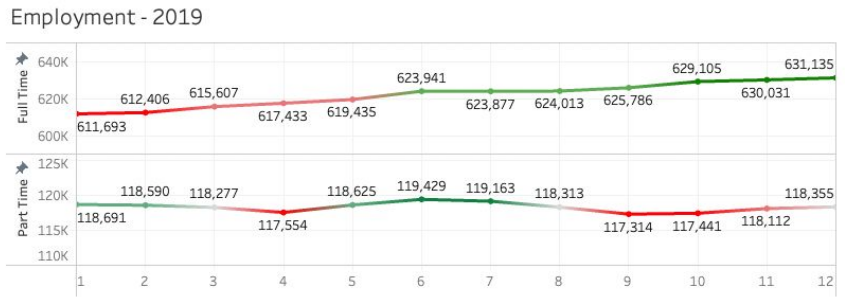


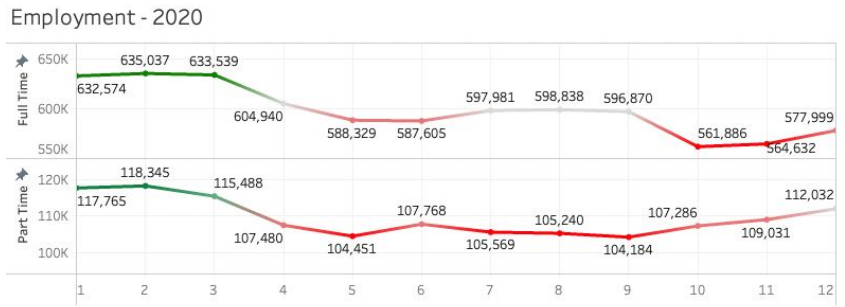


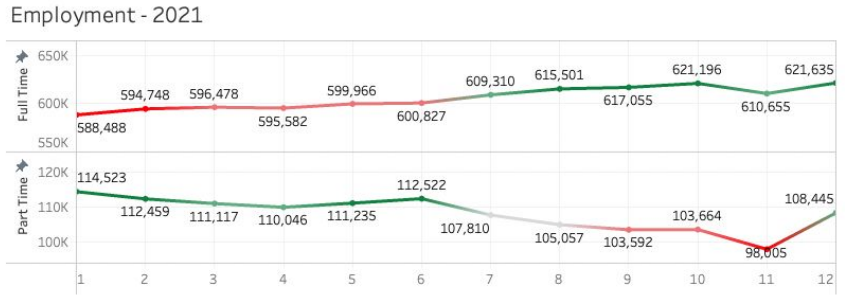
Exploratory Data Analysis & Visualization (Tableau)











Methods & Results: Machine Learning Models

- **Random Forest**

Arrival Delay			
Year	2019	2020	2021
Accuracy	0.81	0.90	0.83
Test-error	0.19	0.10	0.17

- **Linear Regression**

(with and without regularization)

(For all period combined)

	Linear Regression w/o Regularization	Linear Regression w Regularization
R2	0.938	0.938
RMSE	11.729	11.729

- **XGBoost**

(For all period combined)

	XGBoost
R2	0.947
RMSE	12.004

Thank you!