Flight Status Prediction

Group 6: Ardavasd Ardhaldjian, Andrew Fogarty, Calisto Betti, Michael Fawehinmi, Shreedhar Kodate, & River Yan
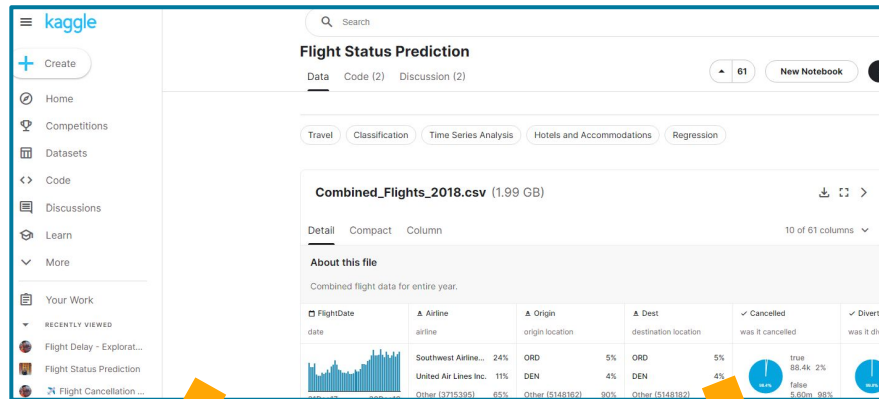
Project Organization

1. The Problem
2. Project Objective
3. Exploratory Data Analysis
4. Feature Engineering
5. Risk Analysis

The Problem

- Every year thousands of flights are delayed
  - Who does this effect?

3

Suitable for Machine Learning?
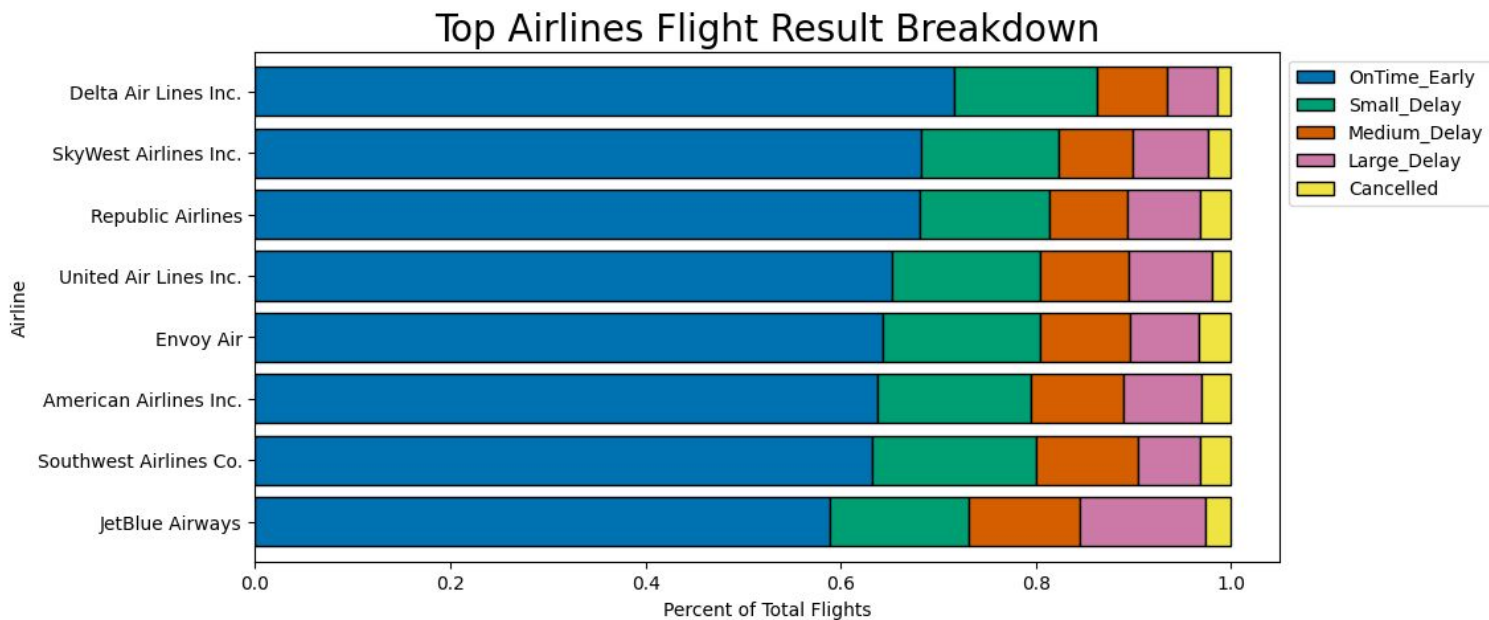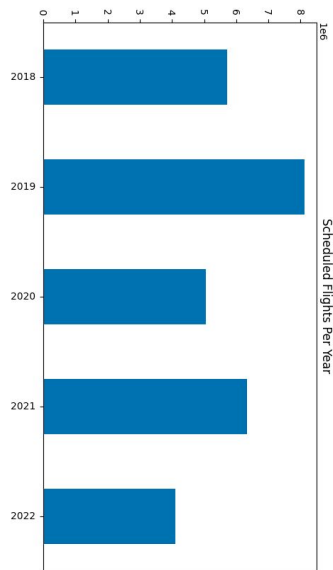

Instances

Features

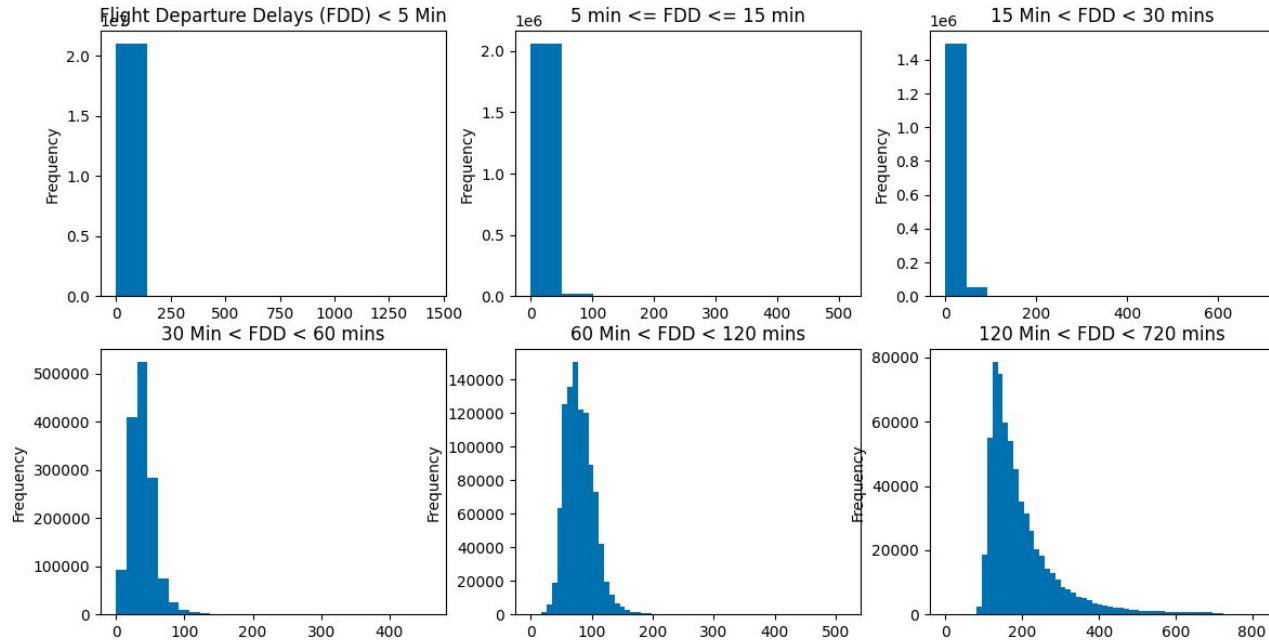# Target Variables

Arrival Delay

Departure Delay

Model Usage

Exploratory Data Analysis
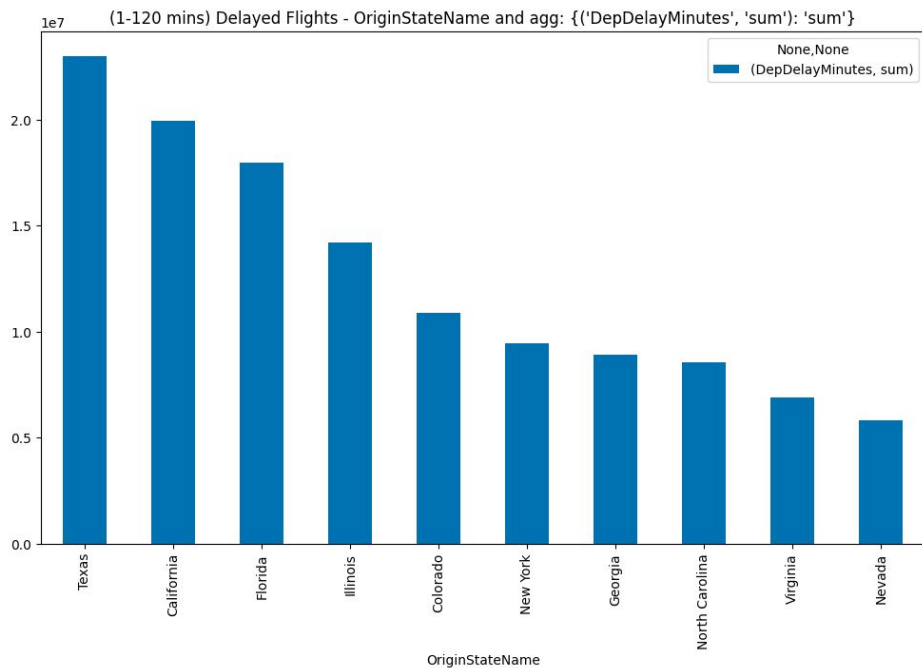

Top Airlines Flight Result Breakdown
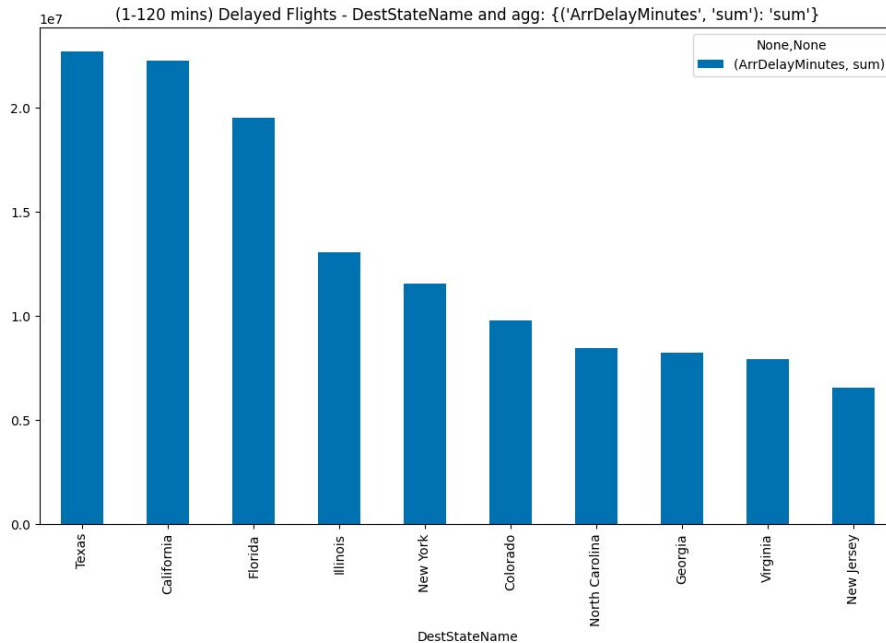
Exploratory Data Analysis

Exploratory Data Analysis

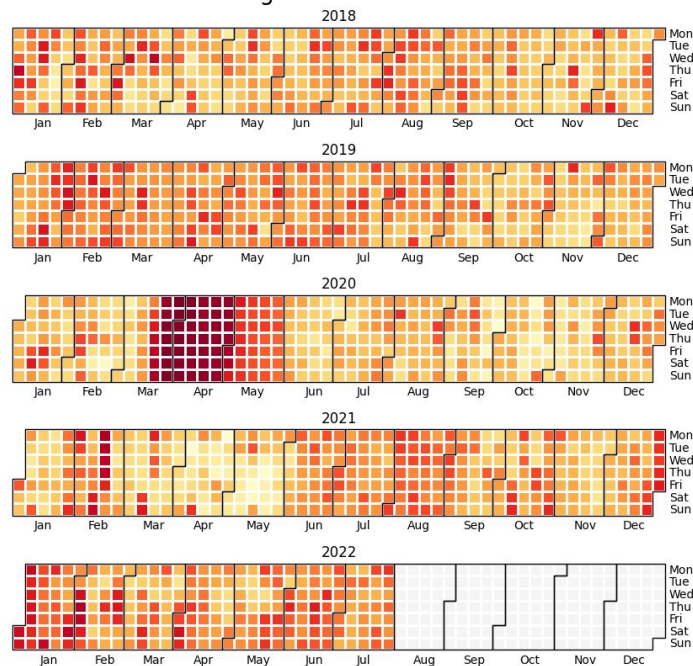More 2,000,000 minutes wasted due to Departure delays

Exploratory Data Analysis

More 2,000,000 minutes wasted due to Arrival delays

Exploratory Data Analysis

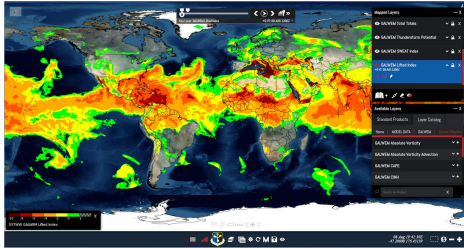| Year	DelayGroup Quarter	OnTime_Early	Small_Delay	Medium_Delay	Large_Delay	Cancelled
2018	1	64.138542	15.995264	9.863051	7.498896	2.504246
	2	62.013049	17.209520	10.497789	9.108815	1.170827
	3	62.048764	16.360446	10.067241	9.702782	1.820767
	4	63.734221	17.690728	10.114689	7.397290	1.063072
2019	1	62.553600	16.165487	10.097043	8.484429	2.699441
	2	62.072775	15.775130	10.029599	9.950298	2.172199
	3	65.546440	14.659965	9.020203	8.919005	1.854386
	4	66.449972	15.913992	9.340979	7.362766	0.932290
2020	1	69.278144	12.152639	6.714420	5.153751	6.701047
	2	68.902406	7.100452	2.572866	1.665322	19.758953
	3	80.512392	11.413842	4.334026	2.860944	0.878797
	4	76.471836	14.059111	5.646849	3.102190	0.720014
2021	1	73.604608	13.237208	6.427152	4.197874	2.533157
	2	68.148487	15.342233	8.454349	7.152819	0.902112
	3	61.424422	16.730648	10.488289	9.305620	2.051022
	4	64.164475	16.723481	10.043912	7.355914	1.712218
2022	1	61.962753	15.195123	10.151889	8.608339	4.081897
	2	59.145074	17.335065	11.358178	9.698266	2.463416
	3	59.472179	16.386961	11.566732	10.769278	1.804850

US Flight Cancellations

11
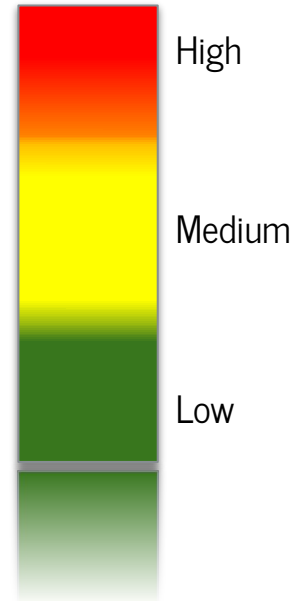
Feature Engineering

- **Numeric Column Grouping**
  - One-Hot Encoding

- **Supervised Learning**

- **Dimensionality Reduction**
  - Feature Selection

- **Alter Feature Variables**

Risk Analysis

- Model Variance
- Dataset Outliers
- Model Accuracy
- Model Misuse

Risk Level

High

Medium

Low

Roles & Responsibilities

- **Recorder:** Takes notes summarizing team discussions and decisions and keeps all necessary records.
- **Reporter:** Serves as group spokesperson to the class or instructor
- **Timekeeper/ Checker:** Keeps the group aware of time constraints and deadlines and makes sure meetings start on time.
- **Business Analyst:** Defines project needs, assists in requirement gathering and hypothesis formulation, documents technical and qualitative.
- **Data Scientist:** Tests hypothesis, develops models, validates and develops code in collaboration with the team.
- **Project Facilitator:** Submits all project assignments and materials in accordance to project deadlines.

Rules

- Attend weekly meeting and in a timely manner
- Maintain communication throughout the week and update team in case of cancellations
- Complete assigned work in respect to due dates
- Be Attentive and collaborate during discussions and group meetings
- Be respectful to fellow teams during virtual discussions/ meetings
- Follow project guidelines and reference canvas guidance
- For decision making a majority vote will taken
- Weekly meetings will take place on Mondays at 6:00pm and Thursday at 6:00 pm(as needed)
- No weekend meetings
- Communications will take place via WhatsApp
- Meeting will be virtual via Zoom.

Thank You !