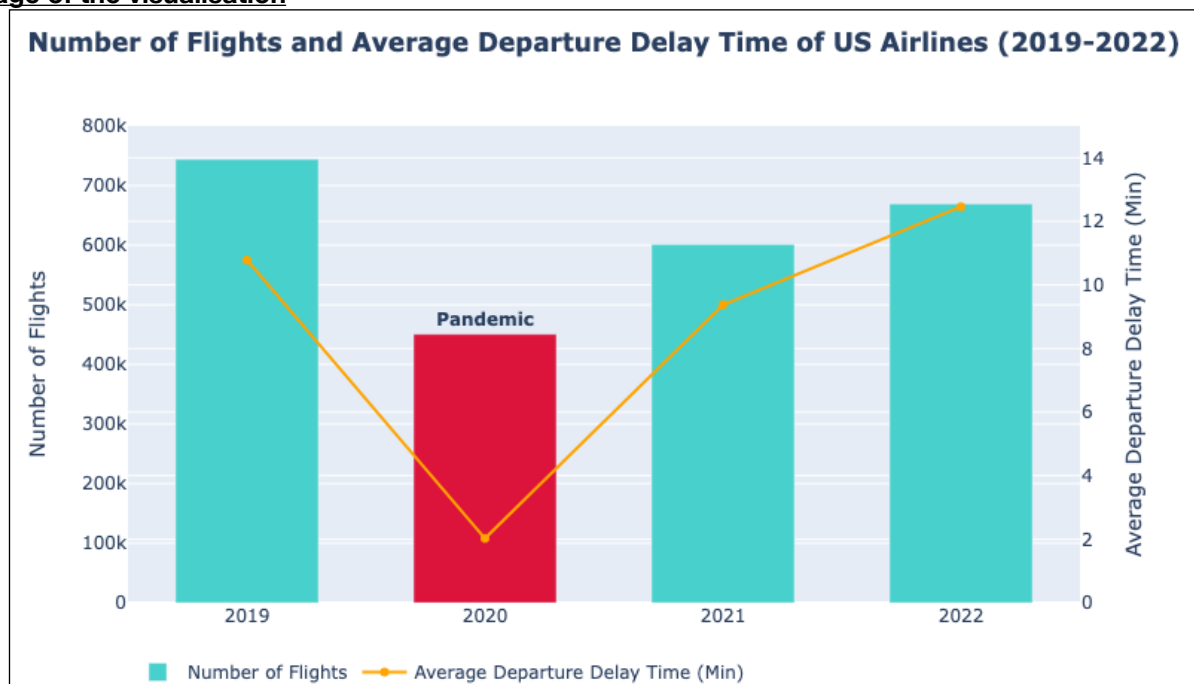# Visualisation 1

## Goal

The target users of this visualisation are US airline/airport operation teams responsible for flight optimisation. The number of flights operated by US airlines from 2019 to 2022 and the average departure delay time (in minutes) of those flights are depicted in a bar-line chart.

The above data is from 2019 to 2022, which included a very critical world event - the pandemic breakout in 2020. The variation in the number of flights during this period is worth studying to understand how the global crisis impacted the US aviation industry, which is the first goal of this visualisation.

This dataset also depicts how the number of flights affects the effectiveness of flight operations. It becomes our second goal since flight delay time is one of the most vital indexes of efficiency measurement and greatly impacts the aviation business due to the extremely high cost on every delaying minute. Knowing the correlation between the number of flights and average departure delay can help airline/airport operators to further study on how to optimise performance regarding the busyness of flight operation. The results of this visualisation provide a high level picture of how the data is correlated.

## Image of the visualisation



## Insight

From the above visualisation, it shows that there was a significant drop in flights operations in 2020 which was the pandemic occurrence. There was steady growth in the following 2 years though the traffic in 2022 was still lower than in 2019. This implies that the covid-19 health crisis impacted US aviation operations.

Similar to the US flights variation, the trend of average departure delay time dropped significantly in 2020 and gradually increased afterward. It indicates that the number of US flights and their average departure delay time have a direct relationship that lower air traffic leads to better performance. The target users can then review strategies related to manpower, airport hardware and security processes to optimise efficiency.

## Data Abstraction

The raw dataset is a table with 32 columns and around 3,000,000 records on flight delay and cancellation from January 2019 to August 2023 operated by US airlines. The dataset is preprocessed using Python to keep only relevant columns and rows (e.g. 2023 records are from Jan to Aug only which cannot provide the

full picture of the entire year so we will not present them in the visualisation) and also restructure them into a new table. Detailed steps will be included in the coding comments.

The abstract dataset types of the preprocessed data is a table with attributes and items.
The abstract data type and metadata are as follows:

| Attribute Name | Abstract Data Type | Metadata (Data Semantics) | |
| --- | --- | --- | --- |
| | | Data Type | Description |
| Year | Categorical key attribute | object | The year of flight date |
| Number of Flights | Quantitative sequential value attribute | int64 | The number of flight operations of US airlines per year |
| Average Departure Delay Time | Quantitative diverging value attribute | float64 | The average of delay time of all flights operated by US airlines per year |

The visual encodings are as follows:

| Mark |
| --- |
| 1. Line (Bar Chart - for Number of Flights) <br> 2. Line & Point (Line Chart - for Average Departure Delay Time) |

| Channel | Effectiveness & Expressiveness |
| --- | --- |
| 1. Position <br> - Vertical Position (2 y-axes: Number of Flights, Average Departure Delay Time) <br> - Horizontal Position (Year) | - The 2 value attributes with 2 separate y axes positioned on the left and right improve accuracy and separability <br> - The vertical position has quantitative scales for showing numbers and minutes respectively which improves the discriminability |
| 2. Colour Hue <br> - Distinguish 2 types of graphs <br> - Highlights bar colour for extreme values | - 2 different colour hues representing 2 different marks for two value attributes improves separability <br> - The highlight colour for the lowest value bar in the pandemic year enhances separability |
| 3. Additional Annotations <br> - Pandemic text label <br> - Hover labels on points for discrete years | - Annotation is used to add a pop out effect indicating the pandemic year, improving separability <br> - The hover labels enhance discriminability on numbers |

**Task Abstraction**

| Action | Target | Description | Respective Goal |
| --- | --- | --- | --- |
| High-Level: Discover | Trend | To obtain a general idea about the trend of number of flight operations and the average departure delay from 2019 to 2022, covering the 'before-during-after' period of covid-19 | Both high-level and mid-level tasks help the target users to achieve the goal of discovering how the global crisis impacted the US aviation industry. |
| Mid-Level: Lookup | Outliers | To lookup the special case - pandemic impact on the 2 value attributes | |
| Low-Level: Compare | Correlation of 2 value attributes | To compare how the average departure delay time changes with the number flights | The target users can also achieve the goal of understanding how number of flights affect the effectiveness of flight operations |

**Data sources**
https://www.kaggle.com/datasets/patrickzel/flight-delay-and-cancellation-dataset-2019-2023/data?select=flights_sample_3m.csv
(The data set is a table in csv format with 32 columns and around 3,000,000 rows)