

MSc Business Analytics

MIS41040 Business Decision Support System

Team 26 – Visualization

|  |  |
| --- | --- |
| **Student Name** | **Student ID** |
| Sathyanarayanan Jayagopi | 23200072 |
| Mouleswar Mani Perumal | 23201655 |
| Sauhrid Mannurkara | 23200698 |

Submission date:17/04/2024

URL: [Dashboard](https://public.tableau.com/app/profile/sathya.jayagopi/viz/MIS41040Team26_17132143404600/Landing)

**Introduction:**

The global Covid pandemic has had significant impacts on various industries, with the travel and aviation sector being one of the hardest hit. In order to gauge the impact accurately by stakeholders, Decision Support Systems (DSS) are crucial tools in analyzing these impacts.

This report outlines the development of a DSS aimed at understanding the pandemic's effect on air travel, focusing on the airline industry and the changes in aircraft utilization. Utilizing data from the OpenSky Network, this DSS provides an in-depth look at the extent to which airlines, aircraft mixes, and city/regions have been affected.

**Information Requirements for Decision-Making:**

The primary objective of a DSS in this context is to assist decision-makers in understanding and responding to the pandemic's impact on travel. To achieve this, the system must meet several critical information requirements:

Historical Flight Activity: The DSS needs access to comprehensive historical data on flight activities, including the frequency of flights, airlines and origins and destination of flight involved

Time Series Data: It should enable analysis over various timeframes to compare pre-pandemic, during-pandemic, and post-pandemic periods. This will help in understanding recovery patterns and predicting future trends.

Geographical Information: Since the impact of the pandemic varied by location, the DSS should facilitate regional and country-specific analyses. This requires integrating geographical data with flight data.

Aircraft Data: Types and model of aircraft used by airlines during the pandemic and recent events concerning the safety of Boeing 737 Max and are vital indicators of how airlines adapted to changes in travel demand, operational challenges and regulations.

**Data Handling and Preparation:**

Step 1: Downloading and Preparing Data from OpenSky Network

The first step involved acquiring a comprehensive dataset from the OpenSky Network for the period between January 2019 and December 2021, accessed from OpenSky Network dataset. This dataset was chosen for its extensive coverage of flight activities, capturing a broad spectrum of data necessary for a thorough analysis of the pandemic's impact on air travel.

The dataset was organized into monthly records stored in compressed .gz files. We downloaded and decompressed a total of 36 CSV files, each file containing on average more than 2 million records. The following are the details columns that are mapped to the CSV file from the OpenSky Network documentation.

Callsign: Identifier of the flight displayed on ATC screens, crucial for identifying the airline.

Number: Commercial number of the flight, useful for tracking specific flights.

ICAO24: Transponder unique identification number, vital for unique aircraft identification.

Registration: Aircraft tail number, important for aircraft-specific analyses.

Typecode: Aircraft model type, necessary for understanding changes in aircraft usage.

Origin and Destination: Airport codes of the flight's origin and destination, crucial for geographical analyses.

First-seen and Last-seen: Timestamps for the start and end of tracking, important for temporal analysis.

Day: Specific day of the flight, useful for daily traffic analysis.

Latitude\_1, Longitude\_1, Altitude\_1; Latitude\_2, Longitude\_2, Altitude\_2: Coordinates and altitude at start and end of flights, initially considered for geographical tracking.

These data were then systematically imported into a PostgreSQL database, which are designed specifically to handle tabular data. Indexes were created on several columns to enhance performance, particularly for real-time querying and complex joins needed for the analysis.

**Step 2: Data Cleansing and Input Feature Selection:**

Further inspection of the dataset led to a strategic decision to drop certain columns to focus our analysis on few critical input features. Specifically:

Latitude\_1, Longitude\_1, Altitude\_1, Latitude\_2, Longitude\_2, Altitude\_2: These columns were removed because geographical analyses were to be conducted using more aggregated geographical data available from Tableau, which recognizes Country, Region/State, and City names.

Destination: This column was also dropped because the decision support system (DSS) being built was designed to analyze overall flight activities rather than specific airline routes. Focusing on the origin of flights was deemed sufficient for this purpose, given that flights return to their operating hubs. Additionally, the return journey of a flight will be captured as a departure from the previous leg's destination which will be the origin, which further supports the decision to concentrate on the origin data only.

First-seen and Last-seen: This column was removed as there is a Day column in the dataset and we are tracking flight operations at Day level

Enhancements were made to the dataset to improve the utility and readability of the information:

Typecode values were translated into actual aircraft manufacturer and model numbers using a lookup table created from data at <https://www.flugzeuginfo.net/table_accodes_en.php> , with a left join performed to enrich the dataset.

Origin codes were converted into more descriptive labels (Airport Name, City Name, Country) using another lookup table sourced from <https://www.flugzeuginfo.net/table_airportcodes_country-location_en.php>.

Callsigns were parsed to extract airline names. The first three characters of each callsign, which typically correspond to an airline, were used to fetch airline names from a curated list found at <https://123atc.com/call-signs> through SQL queries and additional joins.

**Step 3: Handling Records with Null Values:**

To ensure the integrity and usability of our dataset, we implemented specific strategies to handle records with null values in key fields such as origin, typecode, and callsign. Our methods leveraged available data to infer missing values where possible, and selectively excluded records where necessary, as described below:

a) Origin Having Null Values:

For records with null values in the origin field, we implemented a method to infer the missing origin based on the callsign. Specifically, for each flight record with a null origin, we examined other records within the same calendar month featuring the same callsign where the origin was not null. We then used this non-null origin to populate the missing values. This approach rests on the assumption that airlines are unlikely to change the origins associated with a particular callsign within a short period, as callsigns are typically linked to ticket reservation systems used across multiple platforms.

b) Records Having Null Typecode:

Similarly, for records with null typecode values, we sought out other records within the same calendar month that shared the same callsign and had a valid typecode. By applying the non-null typecode from these records to those missing this data, we addressed gaps in our dataset. The underlying assumption here is that airlines are likely to maintain consistency in aircraft types used for specific callsigns over short periods, due to the high costs associated with changing flight and maintenance crews.

c) Handling Null Callsigns or Unmatched Airline Information:

Records with null callsigns or those whose callsigns could not be matched to any airline from our lookup process were excluded from the dataset. This decision was based on the rationale that such flights could belong to private, cargo, or military operations, which do not align with our analysis focus on commercial airline operations. Removing these records helps maintain the relevance and accuracy of our analytical outputs.

**Step 4: Data Aggregation for Analysis:**

To facilitate comprehensive analysis, we structured the data into two distinct aggregated sets based on different analytical needs:

Airline and Airport Analysis: Data was aggregated at the level of airline companies, airports, and days to assess the daily flow of flights. This grouping allowed us to analyze trends and patterns in flight activities at a granular daily level, enabling detailed temporal and spatial insights into how different airlines and airports have been impacted over time.

Airline and Aircraft Manufacturer Analysis: A separate aggregation was performed at the level of airline, aircraft manufacturer, and aircraft model, but on an annual basis. This broader temporal grouping was designed to evaluate long-term trends in the use of different aircraft types by airlines, providing insights into strategic decisions made by airlines regarding fleet size, composition and manufacturer preferences.

**Step 5: Data Extraction for Visualization**

Due to the limitations imposed by Tableau Public, which restricts uploads to worksheets containing no more than 10 million records, we implemented a selective extraction strategy:

We chose to limit our analysis to the top five airlines by flight volume per country. This approach ensures we capture the most significant contributors to air traffic in each country while adhering to Tableau's data capacity constraints.

Two sets of CSV files were prepared: one containing the aggregated Airline Airport Flight Data and the other containing Airline Aircraft Manufacturer Data. These files were specifically tailored for efficient analysis and visualization in Tableau, enabling us to leverage Tableau's powerful visual analytics capabilities to draw actionable insights from the data.

**Dashboard Design and Layout:**

The landing page of the Dashboard provides three options to the user:

a) Covid’s impact on Airline Flight Operations

b) Covid’s impact on Airline Aircraft Manufacturers' Mix

c) Covid’s impact on City/Region Flight Operations

We decided to consider Airline Flight Operations and City/Region Flight Operations separately because airline operations are widespread and do not reflect the impact of Covid-19 at a smaller, regional or city level.

**The Landing Page:**

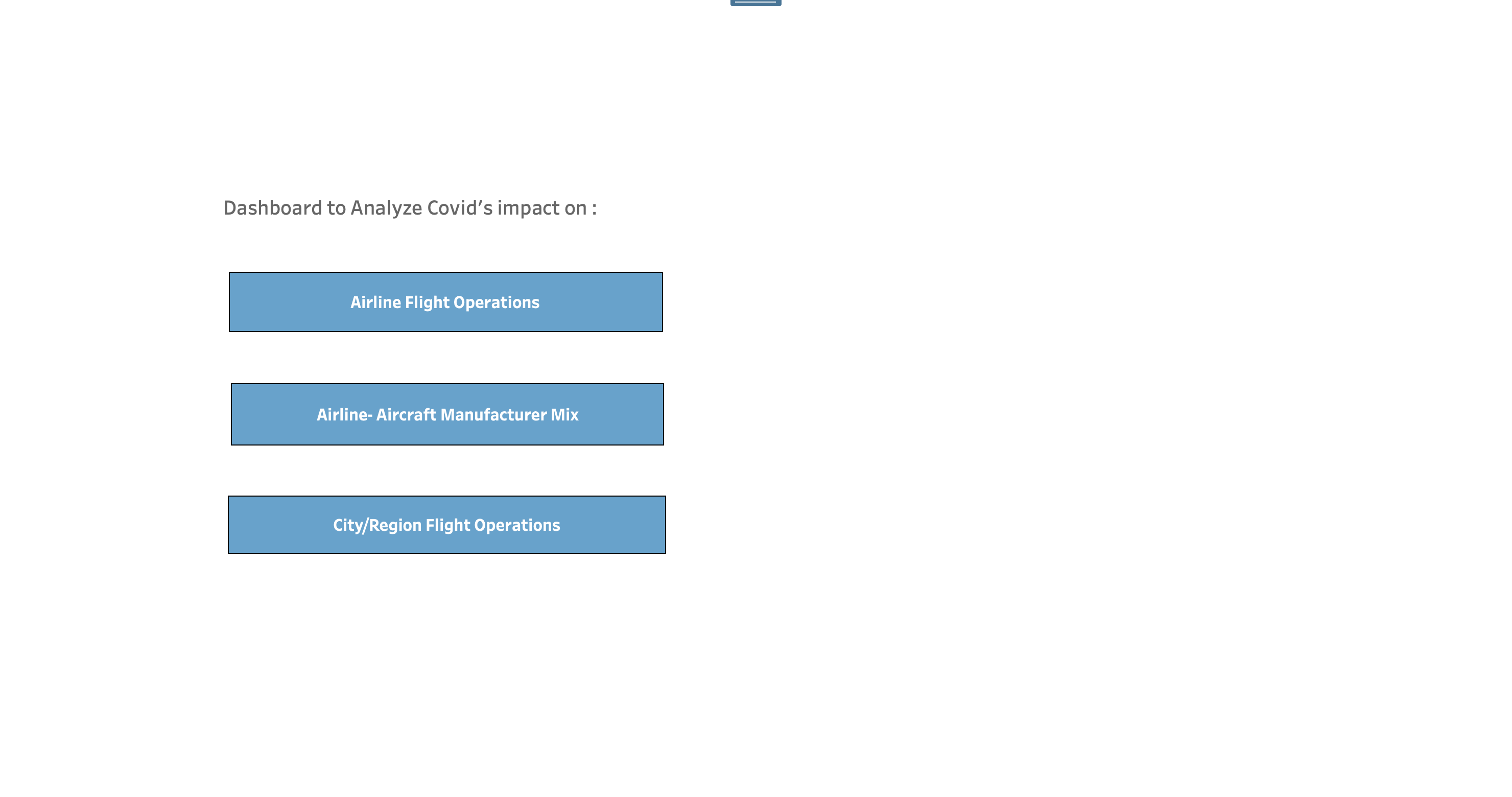


Fig 1.

**Airline Flight Operations Dashboard:**

Once the user clicks on Airline Flight Operations Navigation button, they will be taken to dashboard where they can analyze impact of covid on airline operations.

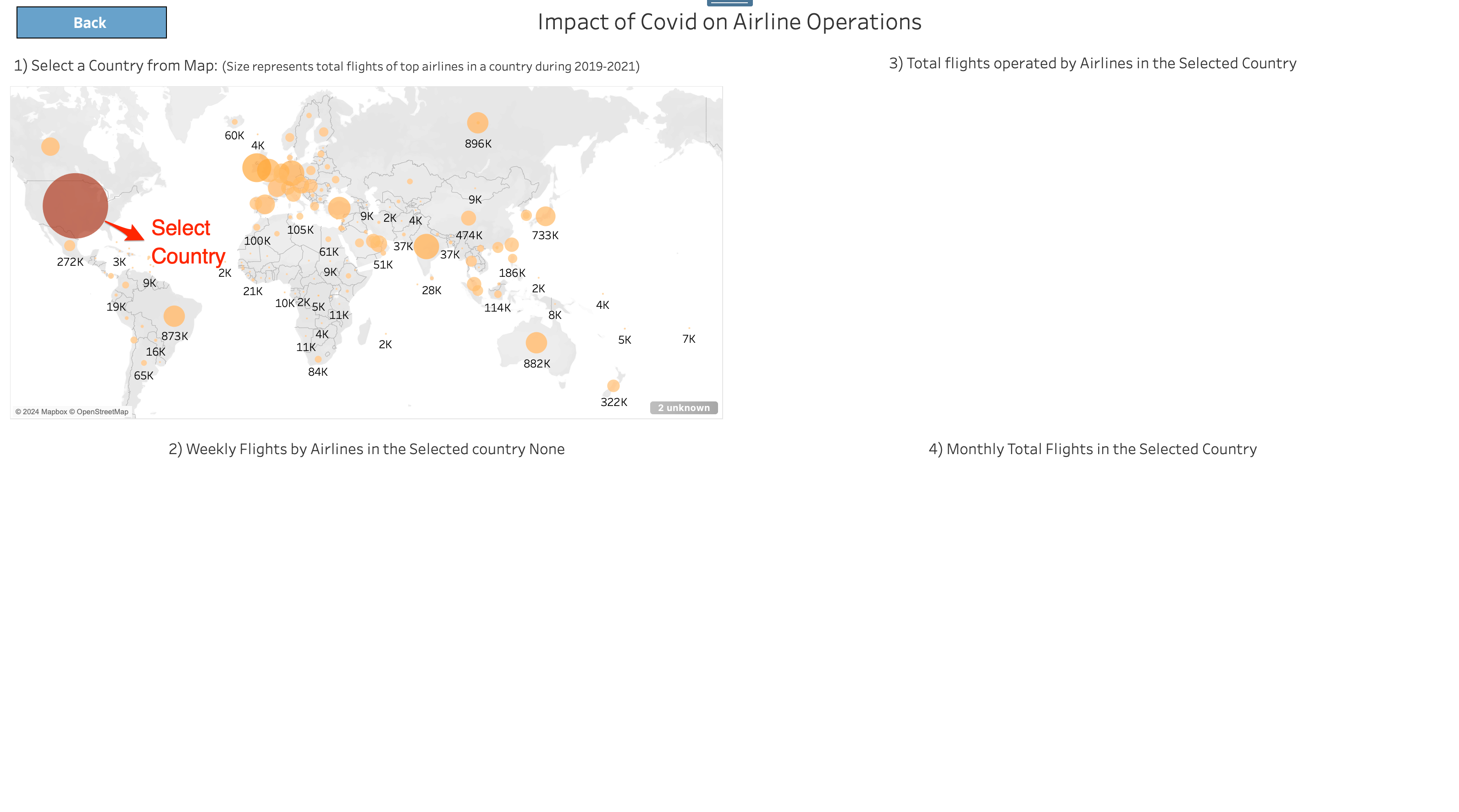


Fig 2.

On this dashboard, users can select country in Section 1 to get the details of flight operation of top airlines within the selected countries.

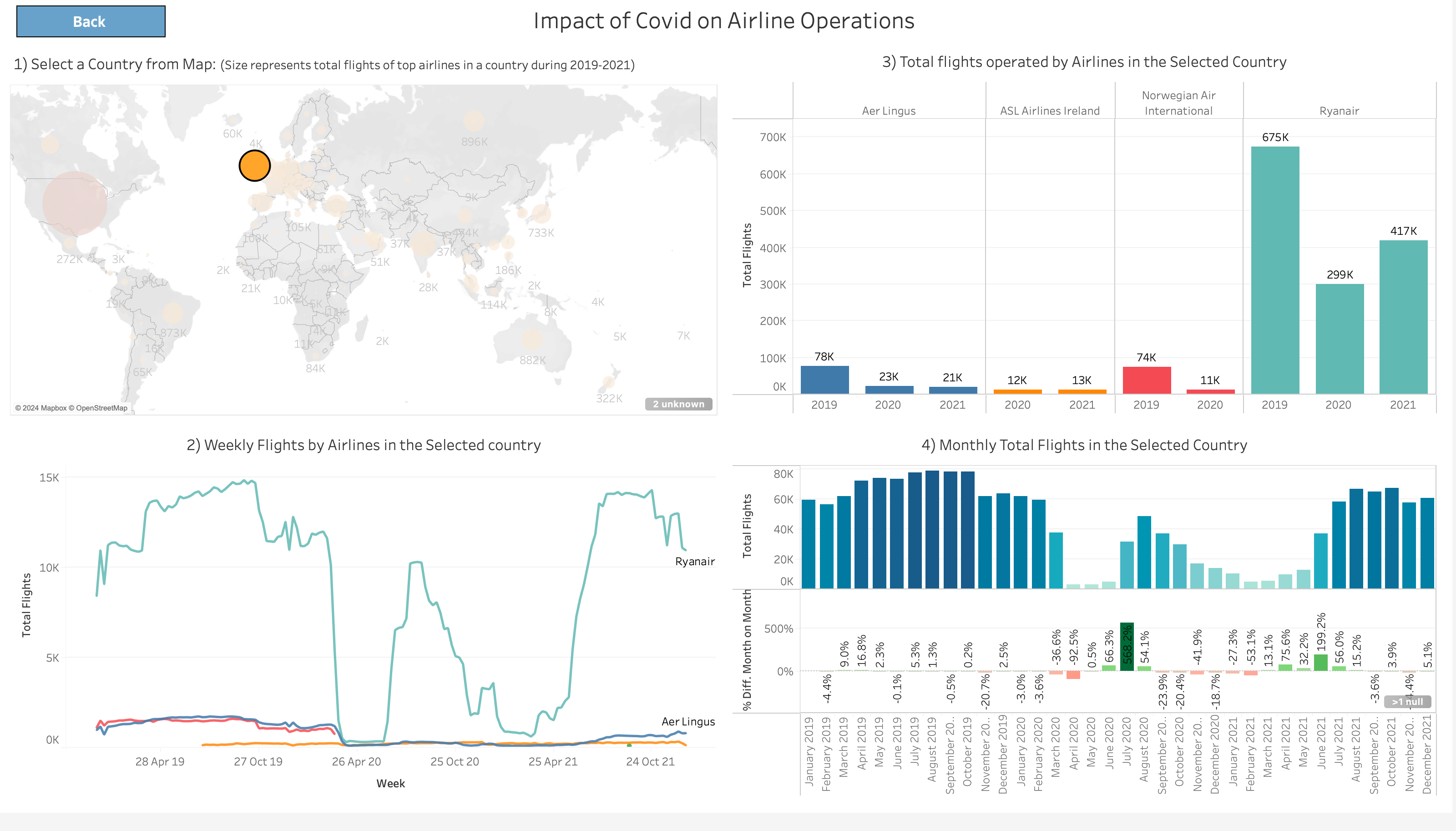


Fig 3

Once a country is selected, Section 2 will display the Total Weekly flights operated by the top airlines in the selected country from Jan 2019 to Dec 2021. We chose to display Weekly total flights instead of daily in order to get a smoother line. Airlines tend to operate more flights during Friday to Monday to address the weekend travel demand and lower numbers are operated during Tuesday to Thursday. Users could observe the sharp decline in flight operations at the end of first quarter of 2020 and the recovery in the flight operation starting in mid to late 2021.

Section 3 will display the total flights operated by top airline in the selected country year wise as a column chart.

Section 4 will display the total flights operated by top airlines in that selected country on monthly basis and also the month on month percentage difference is also displayed.

If the user clicks the Back button, the system will take them back to the landing page.

**Airline- Aircraft Manufacturer Mix Dashboard:**

If the users clicks on the Airline- Aircraft Manufacturer Mix Dashboard button they will be taken to the Airline’s Aircraft model mix page.

In Section 1, A world map will be displayed and user can select a country to get details of top airlines based out of that country and their aircraft mix.

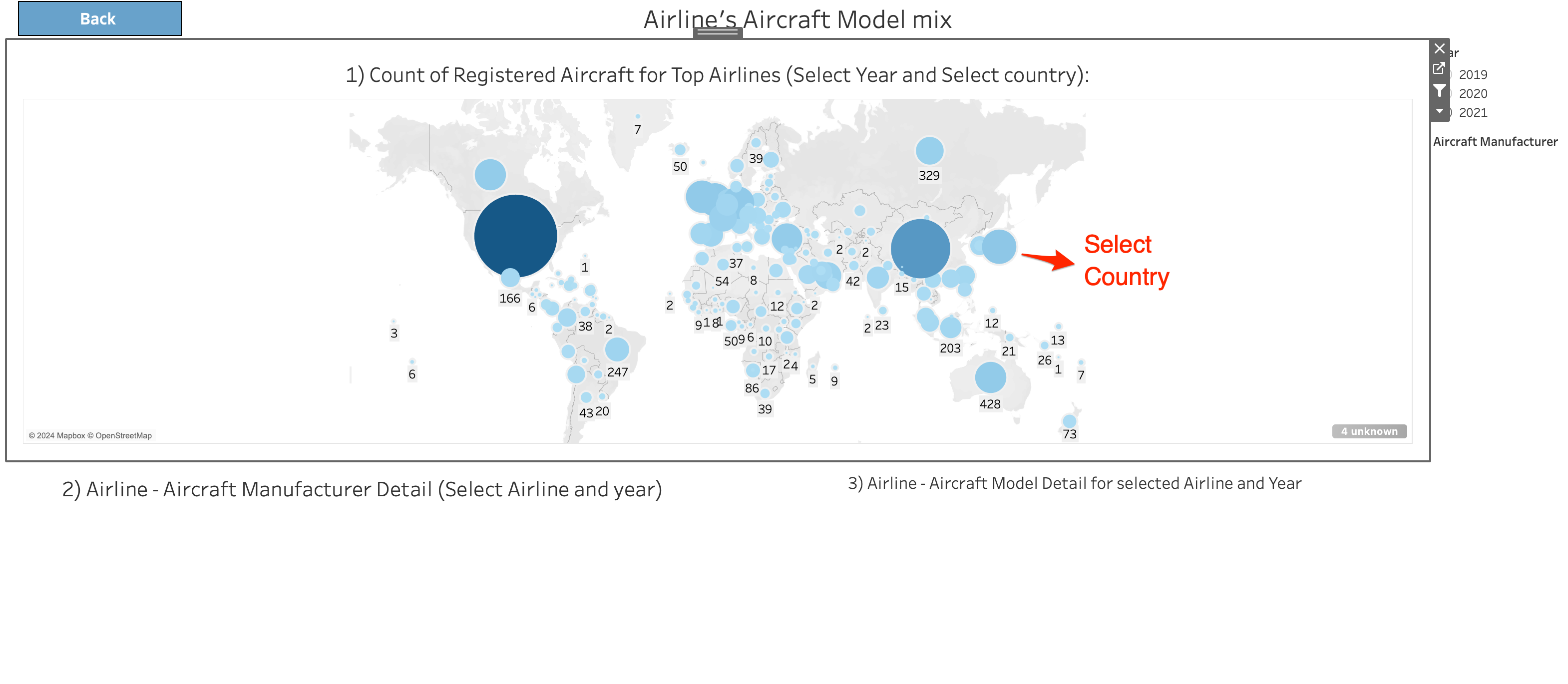


Fig 4

Once the user selects a country on the map, Section 2 will populate with a column graph showing the top airlines in the selected country and a year-wise breakdown of aircraft manufacturers.

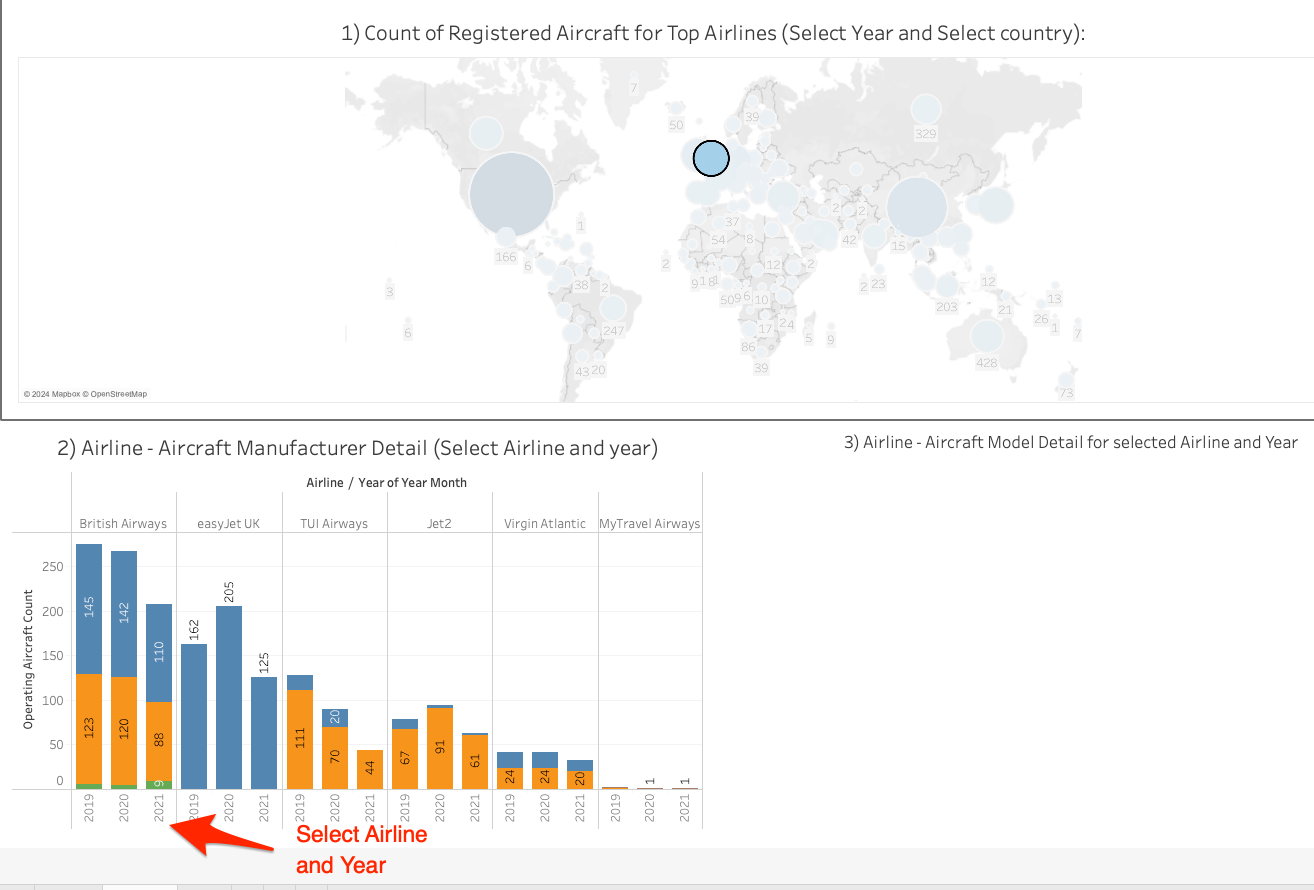


Fig 5

In Section 2, if user selects a year and airline then aircraft model details that the airline had in operation in that corresponding year will be displayed in Section 3.

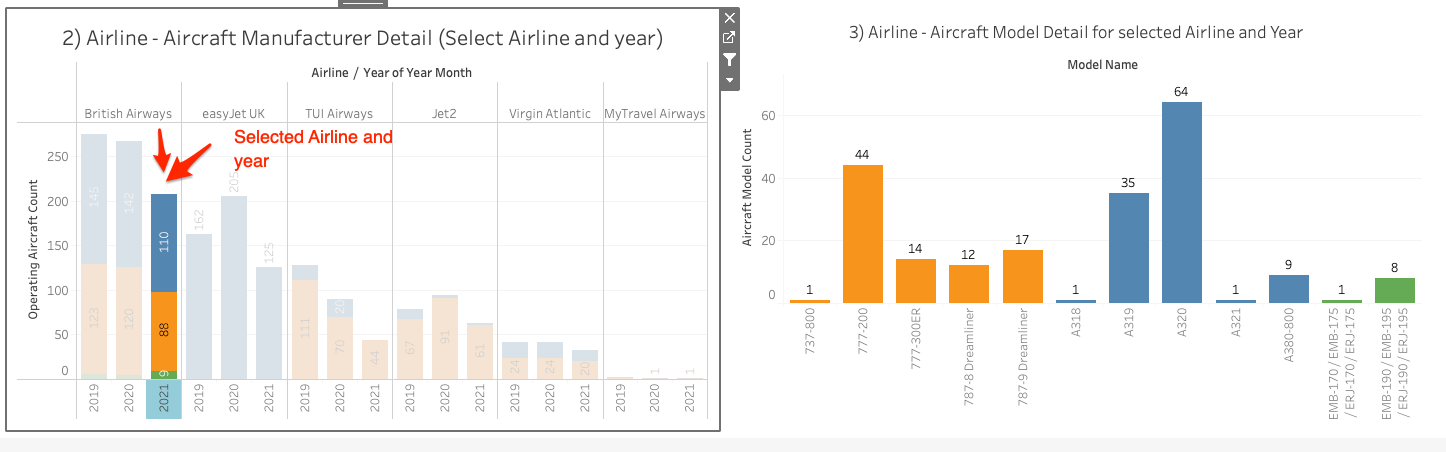


Fig 6

If the user click on the Back Navigation button then the system will take the user back to the Landing page.

**City/Region Flight Operations Dashboard:**

Once the user clicks on Airline Flight Operations Navigation button, they will be taken to dashboard where they can analyze impact of covid on airline operations.

In Section 1, A world map will be displayed and user can select a City/Region to get details of flights operation from all the airport that serve that city/region.

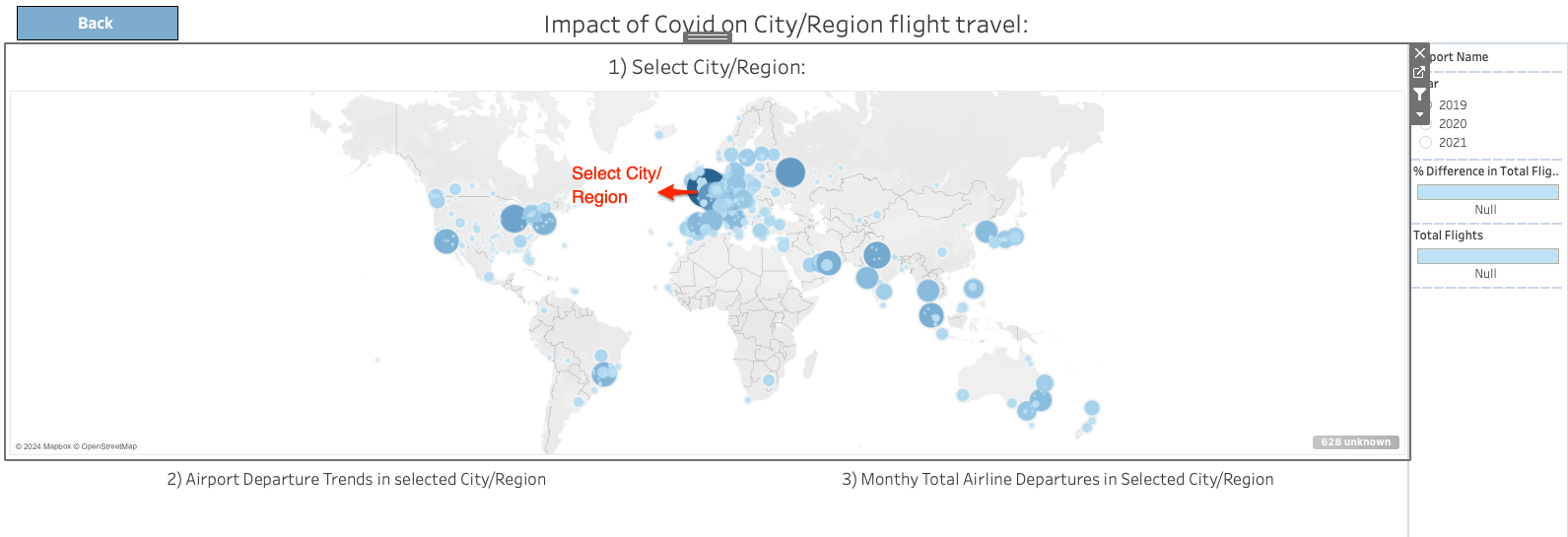


Fig 7

For example, if the user selects London, Great Britain, then Section 2 will display the weekly flight operations of all the airports that server London.

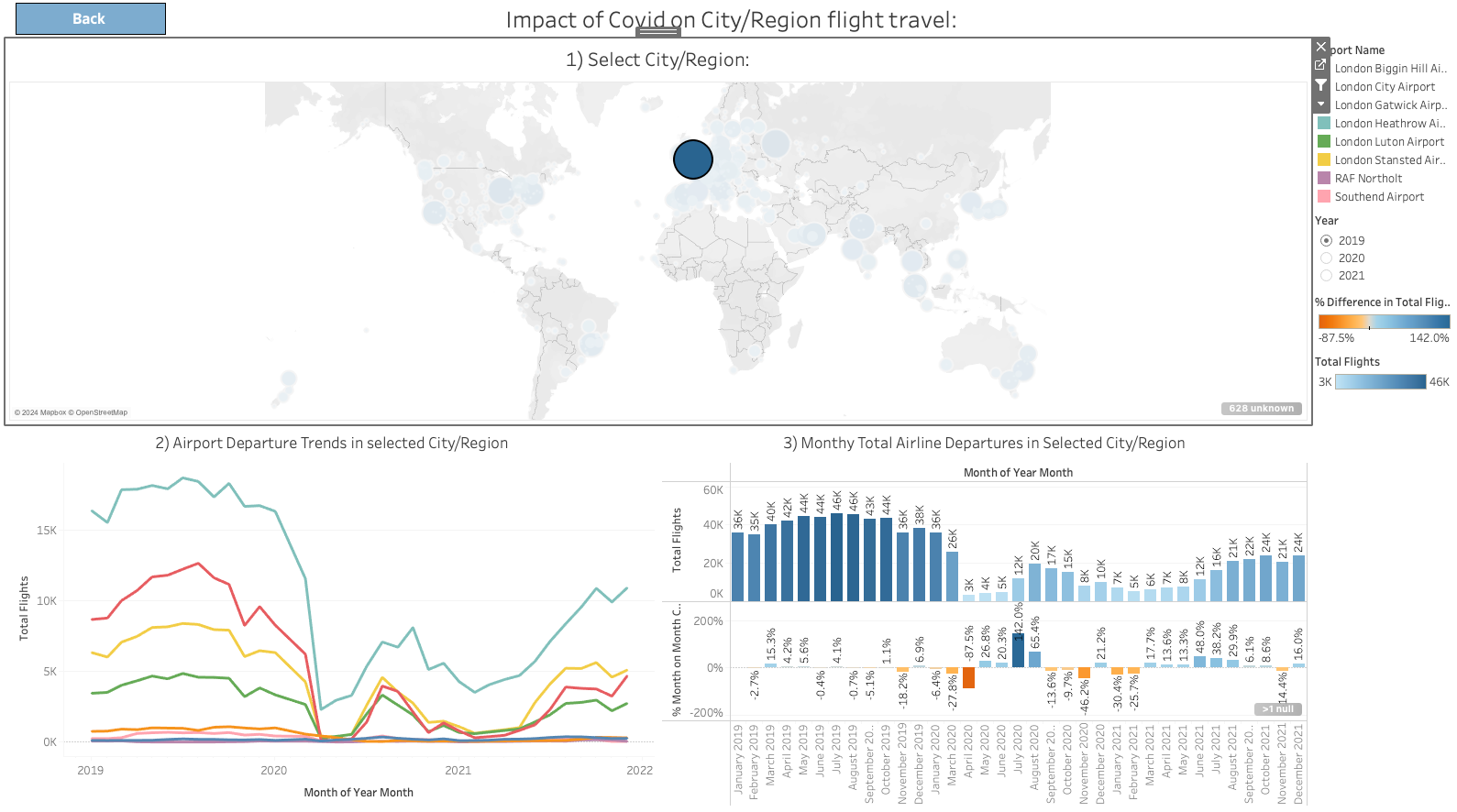


Fig 8

Section 3 will display the total flights operated in that selected city/region on monthly basis and also the month on month percentage difference is also displayed.

**Conclusion:**

The DSS system designed here aims to not only present essential metrics on the impact of Covid-19 on flight operations but also to trace the industry's recovery from the mid-2020 lows to the end of 2021. It offers a robust platform for data-driven decision-making, enabling stakeholders to identify trends, disruptions, and responses within the aviation sector. We designed the system to be a user-friendly interface for stakeholders with varying levels of data expertise who can interact with and extract valuable insights.