A logo of a university

Description automatically generated

**UCD Michael Smurfit Graduate Business School**

**Batch 2023/2024**

**MSc. Business Analytics**

**Module: MIS41040 – Business Decision Support System**

**Lecturer: Peter Keenan**

**Team Number: 3**

**Submitted by:**

|  |  |
| --- | --- |
| **Members** | **Student ID** |
| Karan Gulati | 23200311 |
| Rufus Meera Gomathi Sankar | 23200015 |
| Shivang Ranjan | 23200060 |

**Submitted on: 17th April 2024**

**Topic:**

**Decision Support System - Covid Impact on Air Travel**

A logo of a university

Description automatically generated

**UCD Michael Smurfit Graduate Business School**

**Assessment Project: Declaration of Authorship**

**Team 3**

|  |
| --- |
| **Module Code for which project is submitted:** MIS41040 |
| **Name of the module for which this project is submitted:** Business Decision Support System |
| **Lecturer/Trainer:** Peter Keenan |
| **Title of the Project:** Decision Support System - Covid Impact on Air Travel |

|  |
| --- |
| I have read and understood the **Guidelines for the Preparation of Assessment Projects** issued by the UCD Michael Smurfit Graduate Business School. I hereby confirm that the work submitted for assessment in **this project is my own original work** in accordance with those guidelines. I also confirm that **I have not previously submitted the same work in full or in part at UCD or any other university.**  **Signature**: Karan Gulati, Rufus Meera Gomathi Sankar, Shivang Ranjan |
| **Name in Block Letters:** KARAN GULATI, RUFUS MEERA GOMATHI SANKAR, SHIVANG RANJAN |
| **Student Numbers:** 23200311, 23200015, 23200060 |
| **Date:** 17/04/2024 |

**Data Overview**

**Air Traffic Dataset:** We have utilised air traffic data from The OpenSky Network 2020 highlights air travel during the COVID-19 epidemic, with data collected beginning in January 2019. It includes a set of anonymized flight information, including aircraft models, callsigns, and commercial numbers. Geospatial information on flight trajectories, timestamps, and airport coordinates for origin and destination are included in the collection.

(Link: <https://zenodo.org/records/5815448>)

**Airlines Dataset**: This data is useful for identifying the name of Airlines. We used this data to merge airline information to original data using callsigns.

(Link: <https://en.wikipedia.org/wiki/List_of_airline_codes>)

**Airport Dataset:** We have used this dataset for adding Airport Names to the Air Traffic Dataset. For our system, we used ICAO Codes from this data and merged it with air traffic data for identifying the airport. This data also includes additional fields including country code, country name and continent name.

(Link: <https://github.com/ip2location/ip2location-iata-icao/blob/master/iata-icao.csv>)

**Covid Dataset**: Our World in Data's COVID-19 data provides an overview of the global case situation, updated on a date-to-date basis giving information for each country and continent and their respective new covid cases, covid deaths. It also has additional economic factors like age, gdp, life expectancy and population.

(Link: <https://ourworldindata.org/covid-cases>)

In our data cleaning process using Python within the Jupyter Notebook environment, we processed the dataset to ensure it was primed for analytical tasks.

**Data Loading and Initial Inspection:** Our first step involved loading the dataset, which was stored in a compressed CSV format. Utilizing the Pandas library, we effectively managed the gzip compression to ensure seamless data integration into a DataFrame. This preliminary stage was crucial to lay the foundation for all subsequent data handling operations.

Upon successfully loading the data, we conducted an initial inspection. This involved displaying the first few rows using the df.head() function to get an immediate visual confirmation of the data structure and content. Following this, we performed a systematic check for missing values across all columns by employing the df.isna().sum() method. This check helps in identifying the extent and distribution of missing data, which is essential for determining the next steps in the cleaning process. We've carefully reviewed our dataset for duplicated rows and outliers, addressing them to ensure the data's quality and reliability.

**Streamlining the Dataset:** After assessing the data, we moved to streamline the dataset by removing unnecessary columns. Specifically, columns such as 'number', 'altitude\_1', and 'altitude\_2' were removed. These columns were deemed extraneous as they did not contribute to the primary analytical objectives set for our project. This decision to drop specific columns was a strategic move to reduce memory usage and improve processing time. We also simplified the categorization of aircraft types by slicing off the first character from the 'typecode' column entries. This operation, using basic string manipulation techniques in pandas, allows for a broader classification of aircraft and aids in standardizing the data for easier analysis.

**Calculating and Filtering Journey Data:** A significant step in our data cleaning was the calculation of journey times for each flight. We achieved this by subtracting the 'firstseen' timestamp from the 'lastseen', converting the resulting duration from seconds into hours. This metric is vital for our analysis as it directly relates to flight efficiency and scheduling.

Further data filtration was executed by dropping rows where both 'origin' and 'destination' fields were missing, ensuring that our dataset only included flights with identifiable start and end points. Additionally, to uphold the integrity of our flight data, we removed records where the origin and destination were the same, as such entries could indicate data recording errors.

We also eliminated entries with journey times that could be considered outliers, specifically those exceeding 50 hours. Such extreme values could skew analysis results and often indicate either data entry errors or anomalies not relevant to standard flight patterns.

**Handling Missing Data**: An ongoing challenge in data cleaning is managing missing data. We addressed this by dropping rows that lacked crucial geographical information, such as 'latitude\_2' and 'longitude\_2'. This step is vital as geographical data plays a key role in any analysis involving location-based insights.

Additional data integrity checks involved removing rows that had missing data in essential fields like 'callsign' and 'typecode'. The absence of such critical information can compromise the reliability of the system we have built, thus necessitating their removal.

**Data Enrichment through Merging**: To enhance our dataset further, we merged it with additional external sources providing detailed airport data. This enrichment process added valuable information such as airport names, country codes, and precise coordinates for both origin and destination airports. Such enriched data not only provides a deeper layer of context to our analysis but also facilitates more accurate and insightful geographic visualizations.

During the merging process, we ensured clarity and consistency by renaming columns that were merged into our dataset. This renaming helps avoid confusion in later stages of data analysis, ensuring that each column is distinctly recognizable and correctly interpreted.

We've integrated COVID-19 and economic factors data with our existing datasets to explore correlations between travel patterns and COVID-19 cases. This information has been aligned by date, country, and continent to enhance our analysis.

**Final Cleanup and Data Export:** The culmination of our data cleaning process involved a final sweep to remove any rows that still contained null values in significant columns. Finally, the cleaned and enriched dataset was exported into a new CSV file. This step marks the transition from data cleaning to data analysis, with the dataset now properly formatted, enriched, and ready for advanced analytical tasks.

To stay within the 15 million row limit of Tableau Public, we've tailored our dataset to include data from January to August for both 2019 and 2020. This selection perfectly supports the comparative analysis features of our decision support system.

**DSS Landing Page**

Our introductory page provides a snapshot of our team and outlines the components of our system. To dive in, simply click the "Login" button.

Dashboard Link: [MIS41040 Team 3 | Tableau Public](https://public.tableau.com/app/profile/rufus.m.g/viz/MIS41040Team3_17128417773800/Landing)

**Overview Dashboard**

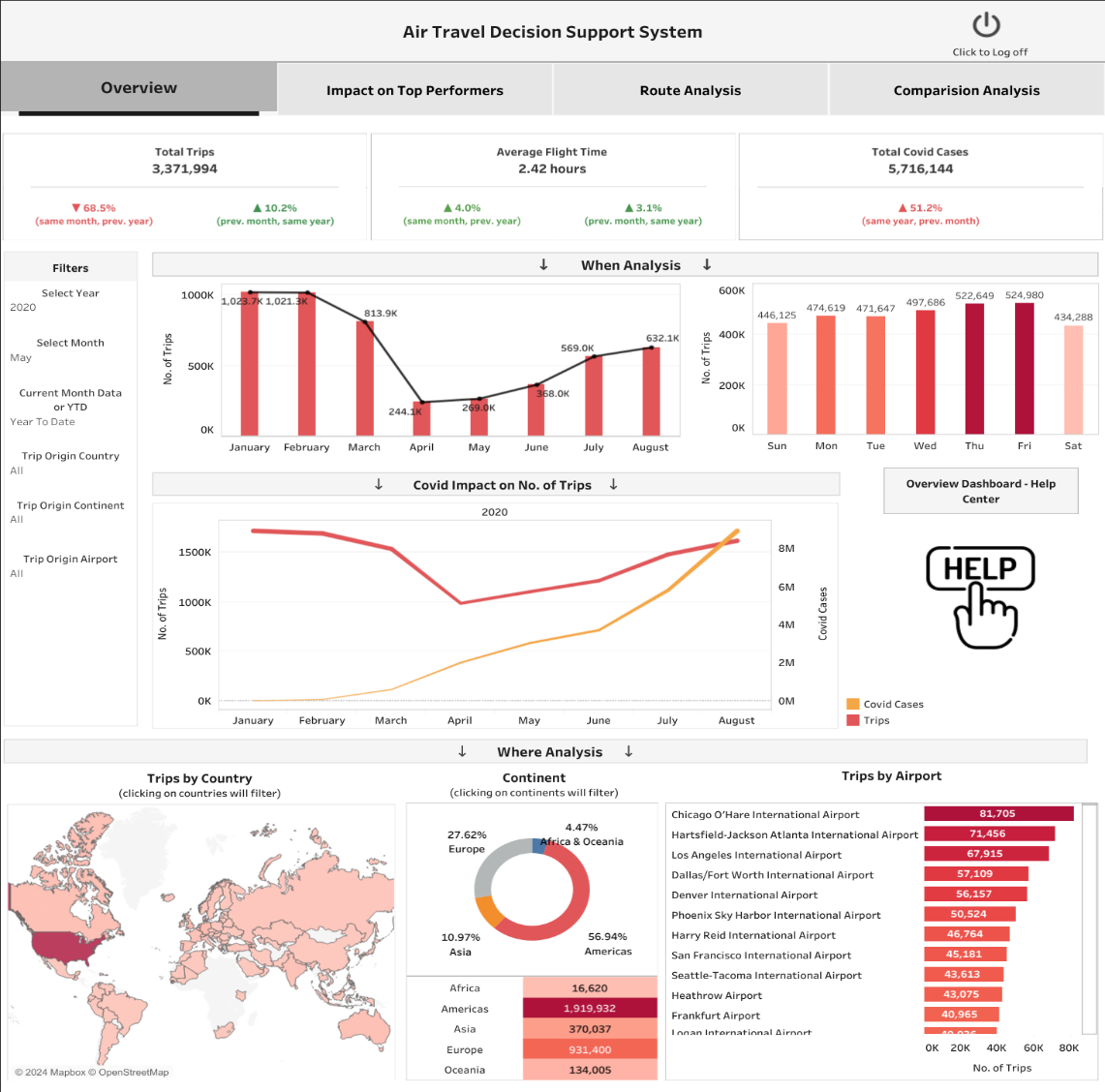
****

Figure 1 : Overview Dashboard Screenshot

The dashboard shows important measures that help understand how the airline industry is affected.

Using “when analysis” users can look at monthly and weekly trends of the number of flights. The users can relate it to the number of covid cases from the KPIs and see how the rise in cases is affecting the number of flights using the trend chart. It shows a direct impact of the pandemic on air travel, with a sharp decline in trips as cases rise and that fewer people are flying because of the pandemic. The dashboard also tells us how long flights are on average and how this might be changing.

Through the “where analysis”, users can visualize which countries, continents and airports were more active for the selected timeframe, giving a picture of where people are flying from. All this information helps users who make decisions for airlines to figure out what to do next and how the pandemic has affected the trips.

**KPIs:**

Total Trips: This metric represents the total number of trips according to the selected filters. The percentage change indicates the variation from the same month in the previous year and from the previous month in the current year.

Average Flight Time: This represents the average duration of flights based on the selected filters. The percentage change compares this average to the same month in the previous year and to the previous month in the current year.

Total COVID Cases: This metric indicates the cumulative number of COVID cases based on selected criteria. It includes a percentage change compared to the previous month of the same year.

**Interactive Dashboard Features:**

The filters enable users to narrow down the data displayed on the dashboard. The "Current Month Data or YTD" filter allows users to select data for a specific month or for the year-to-date (YTD), providing flexibility in the time frame of the data presented.

Users have the flexibility to choose any month and year they need to explore. However, please note that some key performance indicators (KPIs) might not display data for 2019 due to its unavailability, such as COVID-19 information and comparative data from prior years needed for percentage change calculations. The primary goal of this Decision Support System (DSS) is to facilitate the analysis of 2020 data, using 2019 as a reference point for calculations and comparisons.

The Trip Origin Country, Continent and Airports filters lets users select specific regions or airports, so they can focus on data relevant to those areas. The user can apply filters only in regions where it operates to analyze the local impact of COVID-19 on its flights. By manipulating these filters, users will observe the direct impact of covid on air travel, identify potential recovery periods, and make more informed decisions regarding flight scheduling.

The Decision Support System (DSS) includes a helpful guide in the help section of the overview page, making it easier to navigate through the system. Users can easily access the other three analysis tabs with a simple click on the respective buttons. Additionally, for those who need to exit, a logoff option is conveniently located at the top right corner of the page.

**Impact on Top Performers Dashboard**

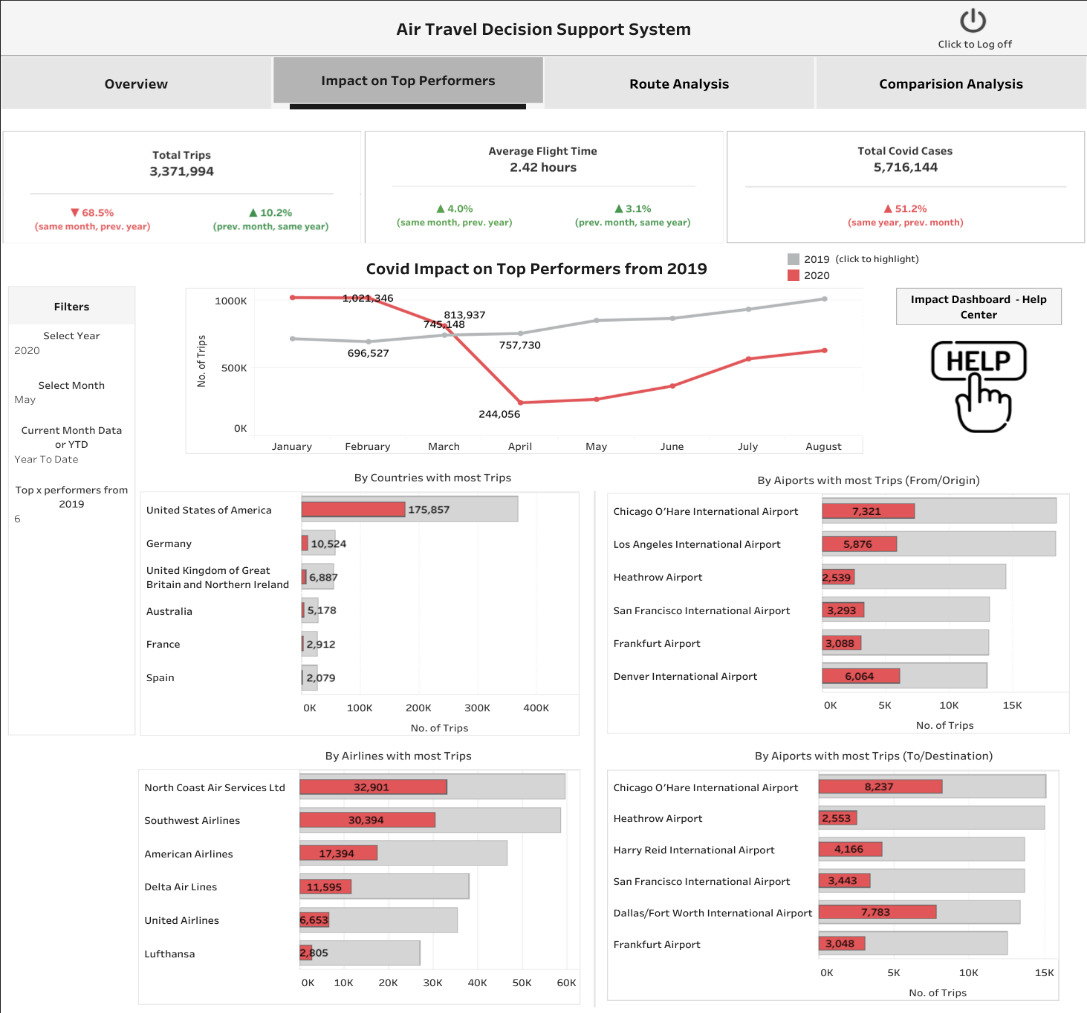
****

Figure 2: Impact on Top Performers

This dashboard provides a detailed view of the aviation industry, focusing on how global events like the COVID-19 pandemic impact major players—airlines, countries, and airports. It highlights their resilience and vulnerabilities, offering valuable insights for airline executives and government officials to identify strengths and areas for improvement.

The choice to focus on these top performers allows us to deeply analyze their responses to crises, aiding strategic planning and preparation for future disruptions. The dashboard features interactive charts and graphs that enhance user engagement and data understanding, enabling users to filter by specific years and months for a tailored analysis. This functionality is crucial for making informed, timely decisions.

By visualizing complex data effectively, the dashboard supports dynamic strategic decision-making, helping industry stakeholders predict trends and devise robust strategies. It’s an essential tool for navigating the complexities of the aviation sector.

#### KPIs:

Total Trips, Average Journey Time and Covid Cases along with the percentage changes remain the same as in the overview dashboard.

**Visuals:**

The charts use grey bars to represent data from 2019 and red bars for 2020, making it easy to spot the differences between the two years. Note that changing the year filter won’t affect these visuals, although the month filter will.

By Countries with Most Trips: This visual monitors air travel activities in countries with the highest number of flights in 2019. It aims to assess the impact of external factors, such as pandemics, on travel patterns and economic interactions within these nations. For the top performing countries in 2019, users can check the impact of COVID on looking at the 2020 data.

By Airports with Most Trips: This chart measures the flow of air traffic from the busiest airports from 2019. Users can analyse both origin and destination airports. The data on fluctuations in airport traffic serves as an indicator to help manage safety protocols.

By Airlines with Most Trips: This tracks the air traffic handled by the busiest airlines in 2019. It explores how external influences, like pandemics, affect flight patterns and airline operations. Users can assess the impact of COVID-19 by comparing the 2020 data for these leading airlines.

**Interactive Dashboard Features:**

Years, months, and current month/year-to-date data remain consistent. Additionally, there is a new filter that allows users to select the top 'n' performers across various categories such as countries, airports, and airlines based on user input.

The Decision Support System (DSS) includes a helpful guide in the help section of the overview page, making it easier to navigate through the system. Users can easily access the other three analysis tabs with a simple click on the respective buttons. Additionally, for those who need to exit, a logoff option is conveniently located at the top right corner of the page.

**Route Analysis Dashboard**



Figure 3: Route Analysis

This dashboard provides comprehensive insights into the intricacies of air travel dynamics. With a spotlight on route analysis, this dashboard offers an in-depth examination of air traffic flows, enabling stakeholders to discern patterns, assess the connectivity of airports, and understand the interplay between different locations. It's a tool tailored to empower airlines, airport authorities, and related agencies with data-driven intelligence to make strategic decisions, optimize operations, and navigate the ever-evolving landscape of air travel.

**KPIs:**

Total Trips: This metric captures the heartbeat of the aviation landscape by quantifying the volume of journeys embarked from a selected airport. It's a direct measure of the airport's activity, providing a clear picture of its traffic and, by extension, its significance in the network during a given time frame.

Average Flying Time: This metric offers a window into the operational aspect of air travel, presenting the mean duration of flights originating from the airport. This information is vital for understanding the nature of the routes serviced, whether they lean towards shorter domestic hops or longer international voyages.

**Visuals:**

Source -> Destination: This displays the flight routes emanating from the airport, highlighting its international network and the extent of its connectivity.

Top Carriers for the Selected Airport: The tree map highlights the busiest airlines operating from the airport, showcasing which carriers dominate the traffic there.

Covid Stats: This visual provides a snapshot of the pandemic's impact, with key statistics like case numbers, helping users gauge the COVID-19 situation relevant to travel and operations. A dynamic feature of this graphic is that selecting an origin or destination airport automatically updates the display with the relevant country's metrics.

Total Trips by Weekdays: This area plot breaks down the airport's traffic by days of the week, offering insights into peak and off-peak travel patterns.

**Interactive Dashboard Features:**

The filters enable users to refine their view based on specific criteria such as year, with 2020 set as the default for route analysis, and the ability to select a particular month for focused insights. Users can also filter by the airport of origin, destination airport, the day of the week, and the airline, allowing for a thorough and targeted analysis of air travel patterns and metrics.

To streamline the selection process and minimize confusion, choosing an origin airport automatically narrows down the destination airport options in the filters to those that are relevant.

Just like other dashboards, users can navigate to other parts of the system, go to the help section or logoff from the route analysis tab.

**Comparison Analysis**

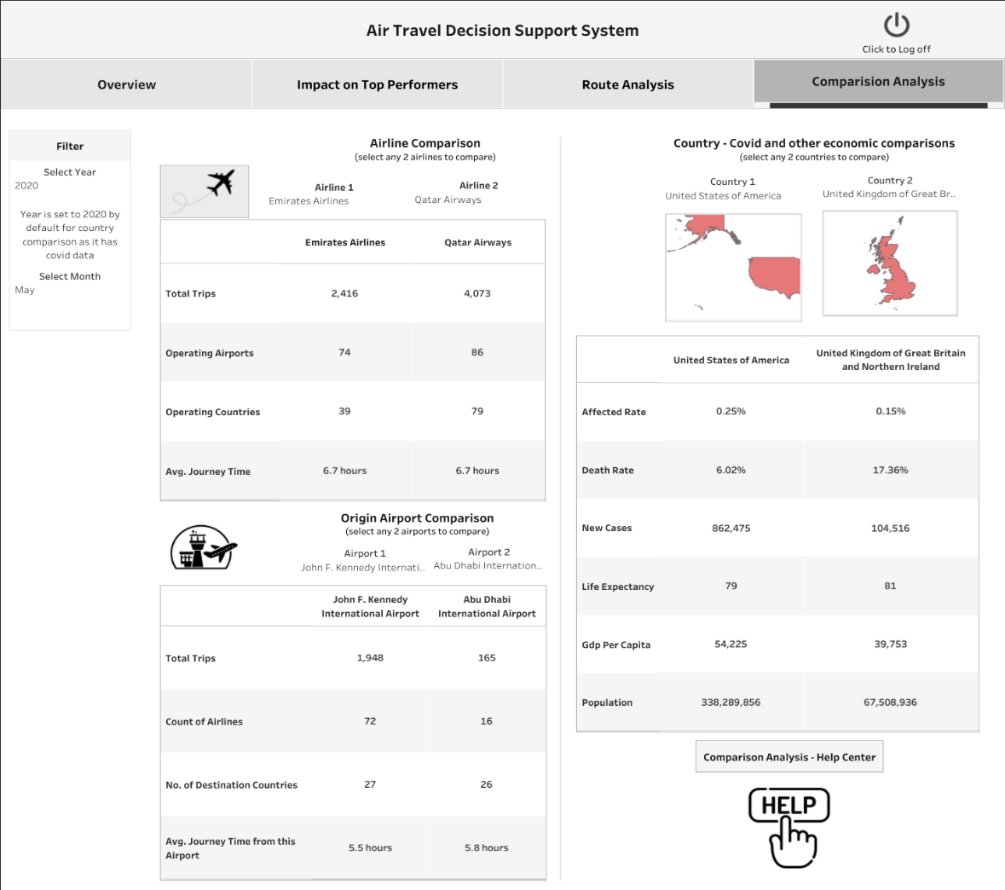


Figure 4: Comparison Analysis

This dashboard provides the user to compare any two airlines, origin airports or between any two countries. For the Airline Comparison, it gives a comprehensive view of the total trips, total operating airports and the number of countries at which it operates, along with average journey time. At the same time, the user can do a quick analysis and comparative study of the chosen airports. It highlights the total trips, the different airlines traveling from the chosen airport, number of destinations and average time from this airport.

For the country comparisons, covid and other economic can be compared. The affected rate is calculated as ratio of the number of new cases for the selected month and the population. The death rate is calculated as the number of deaths due to covid for the selected month divided by the number of people who are affected by it. It also shows the GDP per capita, population data for the countries.

Users can filter out airlines, airports or countries from the filters over the respective charts. Additional interactive elements include the navigations to other tabs, a logoff button and a help button.

**Conclusion:**

In conclusion, our decision support system provides comprehensive insights into the aviation industry's dynamics, offering tools for strategic decision-making, optimization of operations, and understanding the impact of global events like the COVID-19 pandemic. By leveraging data from multiple sources and employing advanced analytics, our system enables stakeholders to navigate the complexities of air travel, identify trends, and make informed decisions. With interactive dashboards tailored to various analysis needs, users can explore key metrics, visualize data trends, and gain actionable insights to address challenges and capitalize on opportunities in the ever-evolving landscape of aviation.

Link: <https://public.tableau.com/app/profile/rufus.m.g/viz/MIS41040Team3_17128417773800/Landing>

**References:**

[1] OpenSky Network (2020) ‘The OpenSky Network 2020 highlights air travel during the COVID-19 epidemic’, Zenodo. Available at: <https://zenodo.org/records/5815448> (Accessed: 29th March 2024)

[2] Ritchie, H., Mathieu, E., Rodés-Guirao, L., Appel, C., Giattino, C., Ortiz-Ospina, E., Hasell, J., Macdonald, B., Beltekian, D. and Roser, M. (2020) ‘Coronavirus Pandemic (COVID-19)’, Our World in Data. Available at: <https://ourworldindata.org/covid-cases> (Accessed: 1st April 2024)

[3] IP2Location (2023) ‘IATA/ICAO List’, GitHub. Available at: <https://github.com/ip2location/ip2location-iata-icao> (Accessed: 1st April 2024)

[4] Wikipedia. (2020) ‘List of airline codes’. Available at: <https://en.wikipedia.org/wiki/List_of_airline_codes> (Accessed: 2nd April 2024)

‌