UCD Micheal Smurfit Graduate Business School

Batch 2023/2024

MSc. Business Analytics

Module: MIS41040 – Business Decision Support System

Lecturer: Assoc Professor Peter Keenan

Submitted by

Members	Student ID	Contribution
Kesav Senguttuvel Jaikalarani	23206949	34%
Jerin Thomas Samson	23202264	34%
Caroline Jenita Jelastin	23202395	34%

Submitted on 17 th April 2024

Topic: MIS41040 Tableau Assignment 2024

UCD Micheal Smurfit Graduate Business School

Assessment Project: Declaration of Authorship

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: **MIS41040 Tableau Assignment 2024**

We have read and understood the **Guidelines for the Preparation of Assessment Projects** issued by the UCD Micheal 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**: KESAV SENGUTTUVEL JAIKALARANI , CAROLINE JENITA JELASTIN , JERIN THOMAS SAMSON
Name in Block Letters: KESAV SENGUTTUVEL JAIKALARANI , CAROLINE JENITA JELASTIN , JERIN THOMAS SAMSON
Student Numbers: 23206949, 23202395 , 23202264
Date: 17 /04/2023

Table of Contents

Introduction.....**4**
Access Group 39 Covid Air Traffic DSS.....**4**
Data Set Creation **4**
16 Laws of UX Design.....**6**
Covid Air-Traffic DSS.....**6**
Conclusion **7**
References **7**

Introduction

This DSS aims to demonstrate the extent of airline disruptions, changes in aircraft utilization, and the effect on airports, regions, and countries during the early months of 2019 and 2020. Challenges include data cleaning due to missing, invalid, or outlier data. We added supplementary public data to convey better insights and various user interaction features, and quantitative analysis methods to enhance the DSS's utility for stakeholders. Keeping the objective in mind, the DSS we developed aims to support the user to aid in better decision-making.

Access Group 39 Covid Air Traffic DSS
https://public.tableau.com/app/profile/kesav.s.j8420/viz/MIS41040Team39_17133931107820/Dashboard3?publish=yes

Data Set Creation

The data provided in the original files limited our exploration in finding insight full patterns. To overcome this we have included data from various public sites, the following is a reflection on the decisions made at arriving at our final cleaned dataset.

Approach

To create a complete and informative analysis of the impact of the COVID-19 pandemic on air travel, we went through two years of data provided by the OpenSky Network dataset, after conducting a lot of feasibility studies and debating our decisions as a group on what to show through thoughtful consideration of what information would yield the most insightful patterns that are vital and easily understandable by the end user. The raw dataset presented a set of varied columns, but not all were directly connected to our objective or the insights we wanted to show through our analysis. We started to add relevant fields to enrich the dataset with additional context that would allow us to answer more questions about the changes in air travel during the pandemic.

Data Used

From OpenSky Network Dataset, to compare the impact of COVID-19 on air traffic we took two years into accord, one was 2019 the year before COVID-19 and followed by 2020. Considering the huge data and the limitations of our computing devices we had to stick with the months from January to May across these two years.

Cleaning Tools

We used a combination of cleaning tools but the majority of cleaning was done with the help of Python, and due to repeated crashing of VS code citing poor memory space, a final few Left Joins were made through Tableau Prep. After this assignment, we can firmly say we now have a good understanding of Python pandas and other parallel computing libraries like python Dask and a basic understanding of Tableau Prep .

Core Columns Retained

Essential for our analysis were columns that offered direct insights into flight patterns, airline operations and their geographical impacts. Thus, we preserved data points like **callsign**, **origin**, **destination**, and **day**, which are fundamental in tracking and analyzing flights across years.

Enhancing the Dataset with Additional Information

To compensate for the limitations of the original dataset and to deepen our analysis, we integrated data from several public sources:

Adding Airline Names

We noticed that the dataset used codes to refer to airlines, which could be confusing. To make our analysis more accessible, we added the actual names of the airlines. We matched the codes in our dataset with a list of airline names and codes we found in a public dataset (Airlines.csv).

By matching the callsign from the original file with ICAO from airlines.csv, we added the airline names to the final cleaned data set.

Including Airport Names

Similar to airlines, airports were identified by codes, which are not widely recognized. By incorporating the real names of the origin and destination airports, we enhanced the dataset's readability and made our analysis more user-friendly. We wanted to make it easier to understand by adding the real names of the airports where flights were taking off and landing. We found another public dataset (Airports.csv) that helped us match airport codes to their names.

By matching the origin and destination codes from the original file with the ident from airports.csv , we added both the origin and destination airport names .

Understanding the Geographical Impact

To understand the impact of the pandemic on different countries, we added information about the countries associated with each airport. We used a country codes dataset to match the airport codes in our dataset with their respective countries and created the flight origin and destination country (Country Codes.csv).

Why did we drop those columns

Upon reviewing the original dataset, certain columns such as **number**, **aircraft_uid**, **type code**, **first seen**, **last seen**, **altitude_1**, and **altitude_2** were removed. These decisions were not made lightly; each column was assessed for its direct contribution to our analysis, we created multiple variations of the data set and ran them through tableau, to see the actual impact these columns had.

For instance, columns like **aircraft_uid** and **type code**, while useful for detailed aircraft tracking, did not contribute significantly to our broader analysis of trends in airline and airport activity across country and in now way were related to COVID impact. Similarly, **firstseen** and **lastseen** timestamps, along with the altitudes, were deemed less relevant for our macro-level examination of the pandemic's impact on air travel patterns.

Why did we drop those rows

This final dataset had data from airports.csv, airlines.csv. After creating the complete dataset and joining all those publicly available data, our total row count was at 25398024. To remove inconsistent rows, we dropped the rows where ICAO, airline name is null; this reduced the number of rows from 25398024 --> 18211053

After cleaning the first data set and doing the visualization we came across various null values that impacted the results we were trying to achieve. We found that the origin and destination columns from the original files from OpenSky Network had null values at random places, we couldn't ignore them as having null on any of those fields, would impact the cumulative final insights, then we tried to fix those, but couldn't find any field that would point us to the missing country data either in origin or destination. Hence, we decided to drop all the occurrences of the rows where either the origin or destination is null using the Python "dropna" method. This reduced the number of rows from 18211053 to 9182289, a massive reduction of 9,028,764 records

Later as we added more public data (Country_Codes.csv) to deliver more insights, this grew our file size grew from 9182289 to 9601023.

With the next cleaning we found that the origin and destination were the same in some records, and on further debugging found they both had the same code (same Airport) we decided there was no pointing in ignoring them as they would severely impact the final insights, dropped all the records where these were the same.

This reduced the number of rows from 9601023 to 9401719

Then we found origin and destination codes were wrong or had random words, and later dropped them to reduce the number of rows from 9401719 to 9316365

Final Dataset Composition

The cleaned dataset now comprises a rich tapestry of information, including not just the basic parameters of each flight, but also the airline identity, airport names, and geographical context(the flight origin and destination country). We retained essential columns like **callsign**, **origin**, **destination**, and **day** understanding their importance after rounds of different cleaning, later enriched the data set with detailed airport information (**ident_origin_Airport**, **Airport_name_origin**, **iso_country_origin**, and their destination counterparts) and airline information (**Callsign_Alpha**, **ICAO**, **Airline Name**).

Furthermore, we added **origin_Airport_country_**, **Destination_Airport_country_**, **origin_Airport_alpha3**, **Destination_Airport_alpha3** to facilitate country-level analysis. The additional time-related fields (**DATE_TIME**, **year**, **month_name**) were extracted to enable temporal analysis, crucial for understanding how air travel trends evolved during the initial stages of the pandemic and beyond.

16 Laws of UX Design

The roles in creating a better UI , that conveys the message with the least possible text have been followed throughout the DSS development. The principles of the Aesthetic Usability Effect, Law of Uniform Connectedness, Miller's Law, and the Pareto Principle have been followed to the core. Taking user behaviour into conversation, the buttons and heading have been placed at the point of most focus.

Covid Air-Traffic DSS

DSS

The DSS starts with the "home" screen that provides two options (Buttons) to the user, the first option takes you to the " Overview " Dashboard, which explains the overall COVID impact on airlines and countries on a higher level. The second button "Comparative Analysis", takes the user to the dashboard with additional user controls to get comparative insights, these provide the impact on countries and airlines where the user could drill down and choose the specific countries he/she desires to get more in-depth insights. Additionally, year and month filters are provided where the user could switch between 2019 and 2020 as well as select months from January to May, to get information on those periods. This aligns with the goal of the DSS to provide a Control Mechanism giving the power back to the user.

Overview

This dashboard screen provides an overall view of the COVID data across the world at a glance. The dashboard has five sheets that provide insights across various perspectives. The control mechanisms provided are Country, Year, and Month, even one change in value in these has an impact on all the sheets in this dashboard, achieving the goal of providing an interactive DSS. The default filtering option across all these filters is "ALL".

Top Destination Countries (heading) data is represented using a bubble chart, where the country with the most airline trips has a bigger size, and the size of the bubble decreases as the count of trips to the country declines. The reason for choosing this chart as it conveys the information with appropriate bubble sizes. The filter that affects this heading are "Year" and "Month", the bubble chart changes in real-time as per the filter chosen by the user.

Most Frequent Flight Paths (heading) data is shown using a Treemap chart, where the highest frequency flight path countries are shown starting with the darkest colour variant and the colour dims across the map as the frequency of the flight path decreases. As the user hovers around the box, we provide a pop-up showing, the count of trips between those countries. The "Year" and "Month" filters provide the user the power to influence

the information portrayed on the map.

Top Airlines by Trips shows the airline with the most trips made ie the airline that was the most active during that period. The main reason to show this bar chart is to make sure the user knows the most active airline during the covid period. The user has the control to choose the 'Months" and "Years" between 2019,2020 and the months between January and May, the airline list updates as per the user preference of filters.

AirTraffic across the Week, the bar chart shows the no of flights that flew on the days of the week. The choice of using bar charts conveys the dip across the days faster.

Comparative Analysis

The purpose of having a second screen (dashboard) is to provide the user with the capability to compare across airlines, countries, and days on a deeper level by providing figures, and percentages that show the decline or increase in traffic. This screen portrays the direct impact of COVID on air traffic comparing it with the pre covid year 2019 with 2020. Respecting the DSS principle of providing user control into consideration, filters have been provided that assist the DSS user to further narrow down their research.

Flight Path map shows the trajectory of flights from the origin country across different destinations, to provide a visual display of air traffic to the user. The user can toggle between the County, Year and Month and that has a direct impact on the path trajectory.

Change of Air traffic across 2019 to 2020, shows the comparison of Air traffic across January to May during pre-covid and covid years. We went with a bar chart as it provides a sharp difference across the months. We compared the months of 2020 with 2019 and provided the change in percentage over the bars. Hovering around the bars provides a pop-up with the total number of trips in that month and in that year.

Most Affected Countries in 2020, this chart shows the countries that witnessed the most decline both with respect to arrivals and departures in air traffic. The percentage at the tip of the bar shows the dip in air traffic in the country comparing it with 2019.

Airlines affected In 2020, similar to most affected countries, this chart shows the decline in the operation of those airlines in 2020 taking the base as 2019.

Impact of Covid on Airlines, this shows the direct impact of Covid on these airlines taking the numbers of 2019 into consideration. The user has been given the control to choose the airline he/she desires.

Conclusion

Through the development of this DSS, we were able to deploy the principle taught in class and through this practical approach, our understanding of the concepts is now firm. From the development of the Data set through various iterations of data cleaning and collection to developing visualizations in Tableau by following the 16 laws of UI Design, we now have a holistic experience in the actual implementation of concepts.

References

1. Olive, Strohmeier and Lubbe (2022), Crowdsourced Air Traffic Data from the OpenSky Network 2020. Available at: https://zenodo.org/records/5815448. [Accessed 20 March 2024]

2. OurAirports , Open Data Downloads. Available at: https://ourairports.com/data/ [Accessed 20 March 2024]

3. OpenFlights. Airport, airline and route data. Available at: https://openflights.org/data.php [Accessed 20 March 2024]

4. Our World in Data . Covid-19-data. Available at: https://github.com/owid/covid-19-data/tree/master/public/data/cases_deaths. [Accessed 20 March 2024]

5. Tadast : Country_codes_and_coordinates. Available at: https://gist.github.com/tadast/8827699 [Accessed 20 March 2024]