SCHOOL OF COMPUTING, ENGINEERING AND BUILT ENVIRONMENT COURSEWORK COVER SHEET	

FOR THE ATTENTION OF LECTURER: **(PLEASE PRINT) Dr Zoë Tieges**	
Student Name: Umunna Auslar Chimuanya	
Matriculation Number: S2275522	**Year of Study: 2023/24**
Programme Title: APPLIED DATA SCIENCE IN ENGINEERING	
Assignment Title: Data Visualisation Coursework	
Module: MMI2268820	
Student Declaration: **This piece of work is not plagiarised. It is my own original work and has not been submitted elsewhere in fulfilment of the requirements of this or any other award.**	
Signature:	**Date: 26/04/2023**
Your mark could have been higher if:	
Additional Marker Comments:	
Marker Signature:	**Total Mark**
Marker Name: Amit Kumar Jain	
Date:	

DATA VISUALIZATION COURSEWORK 2

EUROPEAN FLIGHT DATASET

by
Umunna Auslar Chimuanya

Student ID
S2275522

Coursework submitted in partial fulfilment of the requirement of Glasgow Caledonian University
for Data Visualisation Coursework 2

April 2023

DATA VISUALIZATION COURSEWORK 2 - EUROPEAN FLIGHT DATASET

UMUNNA AUSLAR CHIMUANYA

2023-03-20

Table of Contents

ABSTRACT

This coursework provides a comprehensive analysis of the European flight dataset, which was sourced from Eurocontrol and the official website of the European Union. The report starts by giving an overview of the dataset and then poses three research questions to guide the analysis. To answer the research questions, the report employs exploratory data analysis techniques and various visualisation techniques to identify patterns and relationships in the dataset. The report's visualisations provide clear and concise answers to the research questions and reveal interesting trends such as the monthly variation in air traffic during COVID-19, yearly flight movements, the percentage change in flight movement per year, seasonal trends, and the most popular airports before and after COVID-19. The report concludes that April was the most affected month during COVID-19, 2020 had the lowest flight movement, and air traffic has a clear seasonal pattern. The report uses R Studio and R Markdown to generate the final report and includes detailed explanations and code for each visualisation. The report provides valuable insights into the patterns of air traffic in Europe that can be used by researchers to analyze the traffic in major European airports, identify patterns in-flight data, assess the performance of airports, inform policy decisions, and improve the safety and efficiency of air traffic control.

1.0 INTRODUCTION

For several years, Eurocontrol and Ec.europa.eu have been gathering data on European flights. Ec.europa.eu is the official website of the European Union, which provides information on a range of topics related to the EU, including policies, news, and events. It also hosts several databases and portals, including the European Statistical Data Portal and the European Data Portal, which provide access to a wide range of statistical data and information related to the EU. The European Statistical Data Portal includes data on transport, including aviation, while the European Data Portal provides access to datasets collected by various EU institutions and agencies, including Eurocontrol. Their data includes passenger numbers, cargo volumes, and aircraft movements, obtained from national authorities and other sources across the EU. Eurocontrol, established in 1960, has been collecting data on air traffic in Europe since its inception. Their data includes metrics such as the number of flights, delays, and route patterns and is collected from air traffic control centres throughout Europe. Throughout the COVID-19 pandemic, both organizations have continued to collect data, which has been essential in shaping policy decisions and strategies in the aviation industry and providing insights into the pandemic's impact on air travel in Europe.

The European Union (EU) flight dataset contains vast amounts of information on commercial flights operating within the EU and flights departing from or arriving at EU airports, airline destinations, monthly flight movement, and airport traffic volumes, making it a valuable resource for understanding the European aviation industry's dynamics and tracking trends and patterns in air travel. Before the pandemic, the EU flight dataset reflected the industry's long-term growth, showing a steady and consistent increase in air traffic.

1.1 BACKGROUND

Air travel has become more affordable and accessible in recent decades, leading to a rise in passenger numbers globally, including in Europe. The Burghouwt and Hakfoort (2001) paper titled "The evolution of the European aviation network, 1990–1998" explores the changes in air travel demand, airline competition, and route structures during this period. The paper finds that competition among airlines rose due to a significant increase in air travel demand, leading to a shift in airline strategies from point-to-point routes to hub-and-spoke systems. However, the COVID-19 pandemic had an unprecedented impact on the aviation industry, drastically altering the patterns and trends of the EU flight dataset. Governments worldwide imposed travel restrictions and closed borders, leading to a sharp decline in air travel. Comparing the EU flight dataset before and after the COVID-19 pandemic provides insights into the impact of the crisis on the aviation industry. The Suzumura et al. (2020) paper titled "The Impact of COVID-19 on Flight Networks" examines the effects of the COVID-19 pandemic on global flight networks using a dataset of global flight information. The authors found a significant reduction in flight volume and airline connectivity globally, with more pronounced effects in regions with higher infection rates and stricter travel restrictions. According to the paper, air traffic in Europe was heavily impacted, with a significant reduction in flight volume and airline connectivity. Airlines have also shifted their focus from international to domestic routes, particularly evident in Europe, where intra-European routes have become more critical.

As air travel begins to recover from the COVID-19 pandemic, the European flight dataset is expected to continue to be an invaluable resource for understanding the changing patterns and trends in air travel. Researchers can use the dataset to monitor the recovery of the industry, identify areas of growth and opportunity, and assess the effectiveness of policy decisions and strategies in the aviation sector. With the continued expansion of the European aviation market, the dataset will likely continue to grow in size and complexity, presenting new opportunities for analysis and research.

This report mainly focuses on the European flight dataset and provides data on the volume and movement of flights that occurred pre-COVID-19, during the COVID-19 pandemic, and post-COVID-19. The report highlights significant airports across different European countries and showcases proficiency in data visualization by creating compelling and precise visuals that convey a meaningful narrative. The dataset can help researchers understand how airlines and airports have responded to the pandemic, the effects of the pandemic on passenger behavior, airline performance, and airport traffic volumes, and the long-term prospects of the aviation industry as it faces ongoing challenges from the pandemic.

1.2 AIM

This coursework aims to showcase expertise in the techniques taught in the class and beyond and apply them in a meaningful manner to a new dataset by creating visually appealing, authentic, and coherent narrative visualizations. The task involves investigating the European flight dataset, crafting a story around it, and presenting it through four distinct types of visualizations (or graphs).

1.3 OBJECTIVE

The objective of this report's European flight dataset is to provide information on the number of flights that occurred monthly, yearly, and in major airports across various European countries. This dataset can be used to analyze the traffic in seasons and airports, identify patterns in the flights, and assess the performance of airlines and airports. This coursework will further demonstrate proficiency in data visualization by creating compelling and accurate visuals that effectively communicate a story. The final report for this assignment should utilize a minimum of three distinct visualization techniques and employ R Studio and R Markdown. The objective is to create a report that is both informative and engaging.

2.0 METHODS USED TO ACCESS EUROPEAN FLIGHT DATASET

Two methods are presented to access and load the “flights” dataset, which is part of the tidytuesdayR project.

The first approach involves using the tidytuesdayR package, which can be installed from CRAN via the command install.packages(“tidytuesdayR”). After installation, the tt_load() function can be utilized to load the dataset for a particular week of interest, taking either an ISO-8601 date or year/week format as input. The loaded dataset is then assigned to a variable, such as flights.

The second method is to manually read the data using the read_csv() function from the readr package. The data is available at a specific URL, which is passed as an argument to read_csv(). The resulting dataset is then assigned to a variable, flights.

Both techniques accomplish the same goal, which is to load the “flights” dataset into the R environment for further analysis. The selection of a method is dependent on individual preference, ease of use, and factors like internet connectivity and package availability. This report uses a manual method to access and load the flight dataset.

3.0 RESEARCH QUESTIONS

For a clear and focused inquiry that seeks to address a specific issue or problem, in the context of the European flight dataset before and after COVID-19, a research question can help guide an investigation into the impact of the pandemic on the European aviation industry. These research questions can help guide the selection of variables to analyze, the statistical methods to use, and the conclusions to draw in the EU flight dataset.

1. How has the total number of flights in the European Union changed over the years?

One approach could be to obtain data from official sources such as Eurostat, which provides statistics on air transport in the European Union. The data can also be obtained from individual airports or airlines that operate in the EU. Then analysation of how to determine how the total number of flights has changed over the years by creating visualizations such as line charts or bar graphs to illustrate the changes in the total number of flights over time. In interpreting the data, it is important to consider any external factors that may have affected the total number of flights. For example, changes in the economy, pandemic, or government policies could all impact the number of flights.

2. Were there specific months or seasons during the COVID-19 pandemic where flights were more heavily impacted in Europe?

This would require analyzing data on flight schedules and travel restrictions for different months and seasons during the pandemic. To go about this, data from relevant sources will be collected, such as airport authorities, airlines, or official government statistics. The data should cover a period starting from the beginning of the pandemic till the end of the pandemic and should include information on the number of flights and any travel restrictions or regulations that were in place. After data collection, then there will be analysis to identify any trends or patterns in flight volumes across different months and seasons. This might involve looking at factors such as the number of COVID-19 cases in different countries, changes in travel restrictions or regulations, or fluctuations in demand for air travel. After analyzing the data, then there will be conclusions about which specific months or seasons were more heavily impacted by the pandemic in terms of air travel. For example, you might find that air travel was significantly reduced during the winter months due to increased COVID-19 cases and restrictions on non-essential travel.

3. How has the COVID-19 pandemic affected the busiest airports in Europe in terms of the number of flights before and after the pandemic?

This research question would involve collecting data on flight schedules, passenger traffic, and any travel restrictions or regulations that were in place for the busiest airports in Europe. To go about this, there will be identification of the busiest airports in Europe, and then collect data from relevant sources, such as Network Manager, airport authorities, airlines, or official government statistics. The data should cover a period starting from before the pandemic till the present and should include information on the number of flights, passenger traffic, and any travel restrictions or regulations that were in place. Once the data has been collected, then it can be analyzed to identify any changes in flight volumes at the busiest airports in Europe before and after the pandemic. This might involve comparing data from the same period in different years, looking at changes in the number of flights and passenger traffic, or examining any trends or patterns in the data. After analyzing the data, conclusions will be drawn about how the pandemic has affected the busiest airports in Europe in terms of the number of flights. For example, it might find that the number of flights at these airports dropped significantly during the early months of the pandemic due to travel restrictions and reduced demand for air travel and that there has been a gradual recovery in flight volumes in more recent months.

4.0 EUROPEAN FLIGHT DATASET SOURCES

The “Network Manager” and “Airport Operator” in the European flight dataset are important data sources that provide valuable information for analyzing air traffic patterns in Europe. The Network Manager data source refers to the pan-European air traffic control organization, Eurocontrol, which manages and coordinates air traffic across European airspace. This source provides a complete dataset of European flights, making it a valuable resource for researchers, policymakers, and industry stakeholders. On the other hand, the Airport Operator data source in the given European flight dataset refers to the organization or company that operates the airport from which the flight departed or arrived. While this column also provides important information, it often contains a large number of missing values (NA values). This is because not all airport operators have the resources to collect and provide detailed flight data. As a result, the dataset provided by airport operators may not be as comprehensive as the one provided by the Network Manager.

5.0 CHARACTERISTICS OF VARIABLES OF INTEREST

In R, a variable is a name that represents a value. It can be a numeric value, a character string, a logical value (TRUE or FALSE), or any other data type. Variables are used to store and manipulate data in R programs. To assign a value to a variable in R, you can use the assignment operator <-. The European Flight dataset consists of various variables. These variables are used to clearly define the flight operation processes. They include:

1. YEAR

This variable in the European Flights dataset indicates the year in which the data for each observation was collected. This is important information because air traffic patterns and statistics can vary significantly from year to year, depending on factors such as economic conditions, industry trends, and regulatory changes. The year variable can be used to analyze trends in air traffic over time and identify changes or patterns in air travel. For example, changes in the number of flights or passengers from one year to the next can provide insights into the overall growth or decline of the air travel industry and can be used to inform policy decisions related to air transport. In addition, the year variable can be used to compare air traffic patterns across different years, which can provide valuable information for forecasting future air traffic and planning for infrastructure and resource needs.

2. FLT_DATE

This is used to represent the date, month, and year of flight operations. It is a date-time variable that captures the exact time of takeoff of the flight. The variable is represented in the format of "YYYY-MM-DD", where YYYY refers to the year, MM refers to the month, and DD refers to the day of the month. This variable is useful for analyzing seasonal patterns in air travel and identifying trends in air traffic throughout the year. For example, researchers may use this variable to identify peak travel periods and compare them to other variables in the dataset, such as passenger volume or cargo weight. It can also be used to analyze the impact of weather patterns on air travel, as severe weather conditions may cause flight cancellations or delays.

3. APT_ICAO

This variable refers to the International Civil Aviation Organization (ICAO) code for the airport. The ICAO assigns unique 4-letter codes to airports worldwide, which are used for communication between pilots and air traffic controllers, as well as for tracking flights and air traffic management purposes. The ICAO code is distinct from the International Air Transport Association (IATA) code, which is a 3-letter code used for passenger and baggage handling and ticketing purposes.

The APT_ICAO variable in the European flight dataset is a critical identifier for airports in Europe and is used to track flight operations, airport traffic, and passenger flow. This information is crucial for air traffic management, aviation safety, and policy decision-making. The APT_ICAO code can also provide information about the location and size of the airport, its infrastructure, and its operational capabilities.

4. APT_NAME

This refers to the name of the airport where the flight departed from or arrived at. It is a variable that provides information on the location of the airport and can be used to identify specific airports in the European Union. The name of the airport can be important in understanding the flight data and can provide insights into the airline industry, such as identifying popular or busy airports or understanding flight routes and connections. This variable can be useful for analyzing flight data in terms of airline traffic, route planning, and passenger behavior, among other things.

5. STATE_NAME

This represents the name of the country where the airport is located. In the context of the European Flights dataset, STATE_NAME refers to the country within the European Union where the airport is located. This variable is useful for analyzing air transport trends in different countries and for making comparisons between different countries in the European Union.

6. FLT_DEP_1

This variable in the European flight dataset represents the number of flight departures from an airport. This variable is an important indicator of air traffic activity and is used to analyze trends in air traffic at specific airports and in different countries. The data for this variable is sourced from the Network Manager, which is responsible for the European Air Traffic Management network. The Network Manager collects and manages data on air traffic movements in real time, including flight departures. This data is used to support air traffic management and planning, as well as for research and analysis purposes. The FLT_DEP_1 variable in the European flight dataset provides valuable insights into the volume and frequency of flight departures from different airports in Europe.

7. FLT_ARR_1

This is a variable in the European flight dataset that indicates the number of flights that arrived at the airport on a given day. This data is sourced from the Network Manager, which is responsible for the operation and development of the European air traffic management network. The variable is useful for analyzing airport traffic patterns, assessing airport performance, and planning airport infrastructure development."

8. FLT_TOT_1

This represents a variable in the European Flight Dataset that represents the total number of flight movements at an airport. Flight movements include both departures and arrivals, and FLT_TOT_1 provides a measure of the overall level of air traffic at a given airport. The data for this variable is sourced from Network Manager, which collects and manages air traffic data in the European Union. It can be used to analyze trends in air traffic at different airports, compare levels of air traffic between different regions, and identify airports that are experiencing particularly high or low levels of traffic. This information can be useful for airport management, airlines, and policymakers in making decisions related to airport capacity, infrastructure investment, and air traffic control.

9. FLT_DEP_IFR_2

This European flight dataset represents the number of Instrument Flight Rules (IFR) flight departures from the airport. IFR refers to a set of rules and regulations that pilots must follow when flying in weather conditions where visibility is limited. The variable is sourced from the airport operator and provides insight into the number of flights that departed the airport using IFR during the specified period. This information is important for analyzing flight patterns and airport operations, as well as for assessing the impact of weather conditions on flight departures.

10. FLT_ARR_IFR_2

This variable represents the number of Instrument Flight Rules (IFR) flight arrivals at an airport. IFR refers to the set of rules and regulations that govern the operation of aircraft when flying in conditions where visibility is limited, such as in clouds or poor weather. These rules require pilots to use instrument-based navigation and communication systems to maintain safe and accurate flight paths. The data for this variable is typically collected and reported by the airport operator, who is responsible for tracking flight arrivals and departures. This information is often used to monitor airport activity, track flight delays, and inform airport capacity planning.

11. FLT_TOT_IFR_2

This is a variable that illustrates the total number of flight movements that involve Instrument Flight Rules (IFR) at an airport. IFR is a set of rules and procedures that govern the operation of aircraft when flying in conditions where visibility is reduced, such as in clouds or fog. In these conditions, pilots must rely on their instruments and air traffic control to navigate and avoid other aircraft. The data for FLT_TOT_IFR_2 is obtained from the Airport Operator, who collects data on all flights that use their airport, including those operating under IFR. This variable can provide insight into the overall air traffic at an airport and can be useful for analyzing trends in air traffic volume, identifying busy periods, and informing airport operations and planning.

Column name	Data Source	Label	Description	Example
YEAR	Network Manager	YEAR	Reference year	2014
MONTH_NUM	Network Manager	MONTH	Month (numeric)	1
MONTH_MON	Network Manager	MONTH_MON	Month (3-letter code)	JAN
FLT_DATE	Network Manager	DATE_FLT	Date of flight	01-Jan-2014
APT_ICAO	Network Manager	APT_ICAO	ICAO 4-letter airport designator	EDDM
APT_NAME	PRU	APT_NAME	Airport name	Munich
STATE_NAME	PRU	STATE_NAME	Name of the country in which the airport is located	Germany
FLT_DEP_1	Network Manager	Departures - (NM)	Number of IFR departures	278
FLT_ARR_1	Network Manager	IFR arrivals - (NM)	Number of IFR arrivals	241
FLT_TOT_1	Network Manager	IFR flights (arr + dep) - (NM)	Number total IFR movements	519
FLT_DEP_IFR_2	Airport Operator	IFR departures - (APT)	Number of IFR departures	278
FLT_ARR_IFR_2	Airport Operator	IFR arrivals - (APT)	Number of IFR arrivals	241
FLT_TOT_IFR_2	Airport Operator	IFR flights (arr + dep) - (APT)	Number total IFR movements	519

Figure 1-European flights Dictionary

6.0 LOADING AND INSTALLING LIBRARY

Loading a library in R means making the functions, data, and other features of the library available to be used in your R session. After installing a library using the install.packages() function, you need to load it into your R session using the library() function. When you load a library, all the functions and data sets defined in that library become available in your R session. You can then use these functions and data sets in your R code without having to define them yourself.

In this report, tidyverse package will be used in the collection of R packages designed to make data manipulation, exploration, and visualization easier and more intuitive. It is a popular choice for data analysis and is widely used in the R community.

RMarkdown is a powerful tool in R for creating reproducible reports, presentations, and documents that combine R code, text, and figures. In this report, read.csv will be used to import the dataset, this is because read.csv() is a base R function that reads comma-separated values (CSV) files into R as a data frame. It is a simple and easy-to-use function that is built into R.

```
## ── Attaching core tidyverse packages ──────────────────────── tidyverse 2.
0.0 ──
## ✔ dplyr     1.1.0     ✔ readr     2.1.4
## ✔ forcats   1.0.0     ✔ stringr   1.5.0
## ✔ ggplot2   3.4.1     ✔ tibble    3.1.8
## ✔ lubridate 1.9.2     ✔ tidyr     1.3.0
## ✔ purrr     1.0.1
## ── Conflicts ──────────────────────────────────────────── tidyverse_conflict
s() ──
## ✖ dplyr::filter() masks stats::filter()
## ✖ dplyr::lag()    masks stats::lag()
## ℹ Use the ]8;;http://conflicted.r-lib.org/conflicted package]8;; to force
all conflicts to become errors
```

7.0 IMPORTING DATASET AND LOADING LIBRARY

Importing a dataset in R means reading the data from an external file and storing it in R's memory as a data object. This allows you to manipulate and analyze the data using R functions and packages. Importation of a dataset is the first step in data analysis with R. R supports a wide variety of data formats, including CSV, Excel, SPSS, SAS, and more. Once you have imported the data into R, you can use various functions to explore, manipulate, and visualize the data. R provides several built-in functions and packages for importing different types of data files. The most commonly used functions for importing datasets in R are read.csv(), read.table(), read_excel() (from the readxl package), read_spss() (from the haven package), and read_sas() (from the haven package). The specific function you use will depend on the format of your data file. Overall, importing datasets in R is a critical first step in data analysis and requires careful attention to ensure accurate and reliable results.

7.1 Importing European flight dataset

This reads in data from a CSV file called "flights.csv" and stores it in a variable named **eu_flights dataset**. *eu_flights* is a new variable that has been assigned to the dataset that was imported as a tibble

```
## New names:
## Rows: 688099 Columns: 15
## ── Column specification
## ──────────────────────────────────────────────────────── Delimiter: "," ch
r
## (6): MONTH_NUM, MONTH_MON, APT_ICAO, APT_NAME, STATE_NAME, Pivot Label dbl
(8):
## ...1, YEAR, FLT_DEP_1, FLT_ARR_1, FLT_TOT_1, FLT_DEP_IFR_2, FLT_AR... date
(1):
## FLT_DATE
## ℹ Use `spec()` to retrieve the full column specification for this data. ℹ
## Specify the column types or set `show_col_types = FALSE` to quiet this mes
sage.
```

8.0 CLEANING/WRANGLING PROCESS

Data cleaning and data wrangling are two important stages in the data preparation process that are used to transform and organize raw data into a format that can be used for analysis. Data cleaning is the process of identifying and correcting or removing errors, inconsistencies, and inaccuracies in the data. This includes dealing with missing values, removing duplicate records, handling outliers, and ensuring that the data is in the correct format. Data wrangling, also known as data munging, is the process of transforming and reformatting the data into a more usable format. This involves extracting, merging, and restructuring data from multiple sources, converting data types, creating new variables, and reshaping data to meet the needs of the analysis. Together, data cleaning and data wrangling help ensure that the data is accurate, complete, and in a format that is appropriate for analysis. They are essential steps in the data preparation process and are often iterative, meaning that the process is repeated until the data is ready for analysis.

9.0 DATASET EXPLORATION

9.1 Confirming the data frame structure

This shows the number of rows in the *eu_flights dataset*

```
## [1] 688099
```

9.2 Confirming the data frame structure

This returns the number of columns in the *eu_flights dataset*

```
## [1] 15
```

9.3 Confirmation of the class variables in the dataset

This applies the class type function to each column of the *eu_flights dataset* and returns a vector of the class of each column.

```
##           ...1          YEAR     MONTH_NUM     MONTH_MON      FLT_DATE
##      "numeric"     "numeric"   "character"   "character"        "Date"
##       APT_ICAO      APT_NAME    STATE_NAME     FLT_DEP_1     FLT_ARR_1
##    "character"   "character"   "character"     "numeric"     "numeric"
##      FLT_TOT_1 FLT_DEP_IFR_2 FLT_ARR_IFR_2 FLT_TOT_IFR_2   Pivot Label
##      "numeric"     "numeric"     "numeric"     "numeric"   "character"
```

9.4 Confirmation of the missing values in each column of the *eu_flight dataset*

This calculates the number of missing values in each column of the *eu_flights dataset*. The output is a vector that shows the number of missing values in each column of the dataset. This information can be used to determine which variables have the most missing data and may need to be further investigated or imputed.

```
##           ...1          YEAR     MONTH_NUM     MONTH_MON      FLT_DATE
##              0             0             0             0             0
##       APT_ICAO      APT_NAME    STATE_NAME     FLT_DEP_1     FLT_ARR_1
##              0             0             0             0             0
##      FLT_TOT_1 FLT_DEP_IFR_2 FLT_ARR_IFR_2 FLT_TOT_IFR_2   Pivot Label
##              0        479785        479785        479785             0
```

9.5 Confirming how many missing values are in the European dataset.

This calculates the total number of missing values in the *eu_flights data frame* regardless of which column the missing values occur in.

```
## [1] 1439355
```

9.6 Confirming the structure of *eu_flights dataset*

This will display information about the data frame's structure, including the names and data types of each variable in the data frame, the number of observations, and any attributes associated with the data frame.

```
## spc_tbl_ [688,099 × 15] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
##  $ ...1         : num [1:688099] 1 2 3 4 5 6 7 8 9 10 ...
##  $ YEAR         : num [1:688099] 2016 2016 2016 2016 2016 ...
##  $ MONTH_NUM    : chr [1:688099] "01" "01" "01" "01" ...
##  $ MONTH_MON    : chr [1:688099] "JAN" "JAN" "JAN" "JAN" ...
##  $ FLT_DATE     : Date[1:688099], format: "2016-01-01" "2016-01-01" ...
##  $ APT_ICAO     : chr [1:688099] "EBAW" "EBBR" "EBCI" "EBLG" ...
##  $ APT_NAME     : chr [1:688099] "Antwerp" "Brussels" "Charleroi" "Liège"
...
##  $ STATE_NAME   : chr [1:688099] "Belgium" "Belgium" "Belgium" "Belgium" .
..
##  $ FLT_DEP_1    : num [1:688099] 4 174 45 6 7 98 18 1 401 3 ...
##  $ FLT_ARR_1    : num [1:688099] 3 171 47 7 7 99 21 1 341 4 ...
##  $ FLT_TOT_1    : num [1:688099] 7 345 92 13 14 197 39 2 742 7 ...
##  $ FLT_DEP_IFR_2: num [1:688099] NA 174 45 NA NA NA NA NA 401 NA ...
##  $ FLT_ARR_IFR_2: num [1:688099] NA 161 45 NA NA NA NA NA 306 NA ...
##  $ FLT_TOT_IFR_2: num [1:688099] NA 335 90 NA NA NA NA NA 707 NA ...
##  $ Pivot Label  : chr [1:688099] "Antwerp (EBAW)" "Brussels (EBBR)" "Charl
eroi (EBCI)" "Liège (EBLG)" ...
##  - attr(*, "spec")=
##   .. cols(
##   ..   ...1 = col_double(),
##   ..   YEAR = col_double(),
##   ..   MONTH_NUM = col_character(),
##   ..   MONTH_MON = col_character(),
##   ..   FLT_DATE = col_date(format = ""),
##   ..   APT_ICAO = col_character(),
##   ..   APT_NAME = col_character(),
##   ..   STATE_NAME = col_character(),
##   ..   FLT_DEP_1 = col_double(),
##   ..   FLT_ARR_1 = col_double(),
##   ..   FLT_TOT_1 = col_double(),
##   ..   FLT_DEP_IFR_2 = col_double(),
##   ..   FLT_ARR_IFR_2 = col_double(),
##   ..   FLT_TOT_IFR_2 = col_double(),
##   ..   `Pivot Label` = col_character()
##   .. )
##  - attr(*, "problems")=<externalptr>
```

9.7 To check for the 1st to 6th rows of the *eu_flights dataset*

This displays the first few rows of the dataset. Specifically, it shows the first 6 rows of the *eu_flights dataset*, allowing for a quick preview of the structure and content of the dataset.

This is a commonly used function in data analysis and is useful for getting an initial sense of what kind of data is being worked. *eu_flights dataset*

```
## # A tibble: 6 × 15
##     ...1  YEAR MONTH…¹ MONTH…² FLT_DATE   APT_I…³ APT_N…⁴ STATE…⁵ FLT_D…⁶ F
LT_A…⁷
##    <dbl> <dbl> <chr>   <chr>   <date>     <chr>   <chr>   <chr>     <dbl>
<dbl>
## 1      1  2016 01      JAN     2016-01-01 EBAW    Antwerp Belgium       4
3
## 2      2  2016 01      JAN     2016-01-01 EBBR    Brusse… Belgium     174
171
## 3      3  2016 01      JAN     2016-01-01 EBCI    Charle… Belgium      45
47
## 4      4  2016 01      JAN     2016-01-01 EBLG    Liège   Belgium       6
7
## 5      5  2016 01      JAN     2016-01-01 EBOS    Ostend… Belgium       7
7
## 6      6  2016 01      JAN     2016-01-01 EDDB    Berlin… Germany      98
99
## # … with 5 more variables: FLT_TOT_1 <dbl>, FLT_DEP_IFR_2 <dbl>,
## #   FLT_ARR_IFR_2 <dbl>, FLT_TOT_IFR_2 <dbl>, `Pivot Label` <chr>, and
## #   abbreviated variable names ¹MONTH_NUM, ²MONTH_MON, ³APT_ICAO, ⁴APT_NAM
E,
## #   ⁵STATE_NAME, ⁶FLT_DEP_1, ⁷FLT_ARR_1
```

9.8 Confirming the glimpse of the dataset

This is used to preview the structure and contents of a data frame, including variable names, data types, and a sample of the data itself

```
## Rows: 688,099
## Columns: 15
## $ ...1          <dbl> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 1
6, 1…
## $ YEAR          <dbl> 2016, 2016, 2016, 2016, 2016, 2016, 2016, 2016, 2016
, 20…
## $ MONTH_NUM     <chr> "01", "01", "01", "01", "01", "01", "01", "01", "01"
, "0…
## $ MONTH_MON     <chr> "JAN", "JAN", "JAN", "JAN", "JAN", "JAN", "JAN", "JA
N", …
## $ FLT_DATE      <date> 2016-01-01, 2016-01-01, 2016-01-01, 2016-01-01, 201
6-01…
## $ APT_ICAO      <chr> "EBAW", "EBBR", "EBCI", "EBLG", "EBOS", "EDDB", "EDD
C", …
## $ APT_NAME      <chr> "Antwerp", "Brussels", "Charleroi", "Liège", "Ostend
-Bru…
## $ STATE_NAME    <chr> "Belgium", "Belgium", "Belgium", "Belgium", "Belgium
", "…
## $ FLT_DEP_1     <dbl> 4, 174, 45, 6, 7, 98, 18, 1, 401, 3, 122, 92, 172, 2
```

```
76, …
## $ FLT_ARR_1     <dbl> 3, 171, 47, 7, 7, 99, 21, 1, 341, 4, 130, 90, 175, 2
36, …
## $ FLT_TOT_1     <dbl> 7, 345, 92, 13, 14, 197, 39, 2, 742, 7, 252, 182, 34
7, 5…
## $ FLT_DEP_IFR_2 <dbl> NA, 174, 45, NA, NA, NA, NA, NA, 401, NA, 125, NA, N
A, 2…
## $ FLT_ARR_IFR_2 <dbl> NA, 161, 45, NA, NA, NA, NA, NA, 306, NA, 129, NA, N
A, 2…
## $ FLT_TOT_IFR_2 <dbl> NA, 335, 90, NA, NA, NA, NA, NA, 707, NA, 254, NA, N
A, 5…
## $ `Pivot Label` <chr> "Antwerp (EBAW)", "Brussels (EBBR)", "Charleroi (EBC
I)",…
```

9.9 Confirming the summary of the dataset

This provides a summary of the *eu_flights dataset*. The summary() function provides basic statistical information on a given data frame or vector, including minimum and maximum values, median, mean, first quartile (25th percentile), the third quartile (75th percentile), and any missing values

```
##       ...1              YEAR       MONTH_NUM          MONTH_MON        
##  Min.   :     1   Min.   :2016   Length:688099      Length:688099     
##  1st Qu.:172026   1st Qu.:2017   Class :character   Class :character  
##  Median :344050   Median :2019   Mode  :character   Mode  :character  
##  Mean   :344050   Mean   :2019                                        
##  3rd Qu.:516075   3rd Qu.:2020                                        
##  Max.   :688099   Max.   :2022                                        
##                                                                       
##     FLT_DATE             APT_ICAO           APT_NAME          STATE_NAME       
##  Min.   :2016-01-01   Length:688099      Length:688099      Length:688099     
##  1st Qu.:2017-10-12   Class :character   Class :character   Class :charact
er  
##  Median :2019-05-01   Mode  :character   Mode  :character   Mode  :charact
er  
##  Mean   :2019-04-22                                                          
##  3rd Qu.:2020-11-22                                                          
##  Max.   :2022-05-31                                                          
##                                                                              
##    FLT_DEP_1         FLT_ARR_1         FLT_TOT_1       FLT_DEP_IFR_2   
##  Min.   :  0.00   Min.   :  0.00   Min.   :   0.0   Min.   :   0.0   
##  1st Qu.:  5.00   1st Qu.:  5.00   1st Qu.:  10.0   1st Qu.:  38.0   
##  Median : 17.00   Median : 17.00   Median :  35.0   Median :  91.0   
##  Mean   : 63.24   Mean   : 63.28   Mean   : 126.5   Mean   : 143.7   
##  3rd Qu.: 71.00   3rd Qu.: 71.00   3rd Qu.: 141.0   3rd Qu.: 195.0   
##  Max.   :847.00   Max.   :813.00   Max.   :1628.0   Max.   :1039.0   
##                                                     NA's   :479785   
##  FLT_ARR_IFR_2    FLT_TOT_IFR_2    Pivot Label       
##  Min.   :  0.0    Min.   :   0.0   Length:688099     
##  1st Qu.: 38.0    1st Qu.:  76.0   Class :character  
```

```
##  Median : 91.0    Median : 182.0   Mode  :character
##  Mean   :143.6    Mean   : 287.3
##  3rd Qu.:195.0    3rd Qu.: 390.0
##  Max.   :817.0    Max.   :1624.0
##  NA's   :479785   NA's   :479785
```

9.10 To confirm the dimensions of the dataset

This presents the dimensions of the *eu_flights dataset*, which is the number of rows and columns in the dataset. The output is a vector with two elements: the number of rows and the number of columns in the data frame. The first value indicates the number of rows in the data frame, and the second value indicates the number of columns.

```
## [1] 688099     15
```

10.0 CORRELATION COEFFICIENT

10.1 Confirming the correlation between FLT_TOT_1 and FLT_TOT_IFR_2

Description

This selects specific columns (YEAR, FLT_TOT_1, and FLT_TOT_IFR_2) from the eu_flights dataset, removes any missing values, and calculates the correlation between FLT_TOT_1 and FLT_TOT_IFR_2. It then creates a scatterplot with these variables on the x- and y-axis to show the strength and direction of their relationship. A correlation coefficient closer to 1 or -1 indicates a strong positive or negative relationship, while a value closer to 0 indicates a weak or no relationship. The output of this plot shows a correlation value of 0.95, which indicates a strong positive linear relationship between FLT_TOT_1 and FLT_TOT_IFR_2.

Reason for Choice of Variable

The total number of flight movements (total number of flight arrival and total number of flight departures) provided by Network Manager which is denoted by the variable FLT_TOT_1 has no missing value, while there are missing values in FLT_TOT_IFR_2 (total number of flight arrival and total number of flight departure) which is the total number of flight movement provided by Airport OPerator, this variable will be dropped, and FLT_TOT_IFR_2 will be used instead in this report.

Choice of Graph

This visualisation aims to plot a graph showing the correlation between FLT_TOT_1 and FLT_TOT_IFR_2 hence, a scatter plot and a line graph are used to visualize the relationship between the two variables.

```
## `geom_smooth()` using formula = 'y ~ x'
```

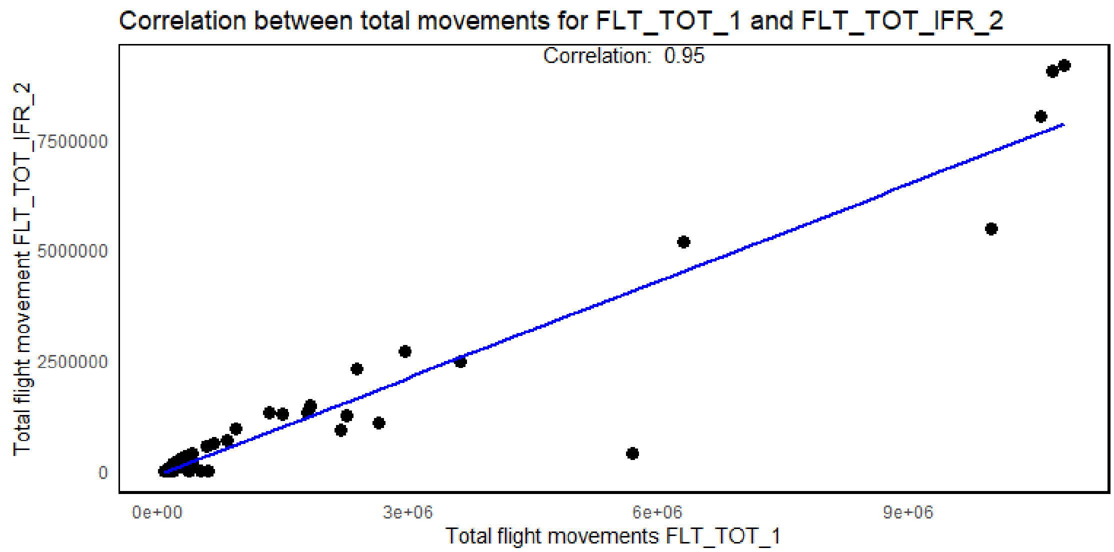

Figure 2 - Correlation between FLT_TOT_1 and FLT_TOT_IFR_2

Graph Analysis

As the value for the total number of flight movements provided by Network Manager data source increases, the total number of flight movements provided by the Airport Operator data source () also increases, and this shows a strong correlation between the two variables.

11.0 VISUALISING THE EUROPEAN FLIGHT DATASET BY YEARLY MOVEMENT

The initial research inquiry seeks to determine how the overall quantity of flights in the European Union has evolved throughout the years.: The following plot analysis answers this question

11.1 To check the trends of European flights from 2016 to 2022

Description

This takes the *eu_flights dataset*, groups it by year, calculates the percentage change in the total number of flights for each year and assigns the result to a new data frame named eu_flights_1. This new data frame contains two variables: YEAR and FLT_TOT_1. The resulting data frame can then be used to create visualizations

Reason for choice of graph

This visualisation aims to plot percentage change in the total number of flights per year, hence, a line graph and scatter plot were used. This is because they are effective in showing trends and patterns over time or across a continuous variable. Line graphs are particularly useful when visualizing how a single variable changes over time or as a function of another continuous variable. They are especially effective for showing trends and changes in data that occur gradually, and for identifying the direction and rate of change.

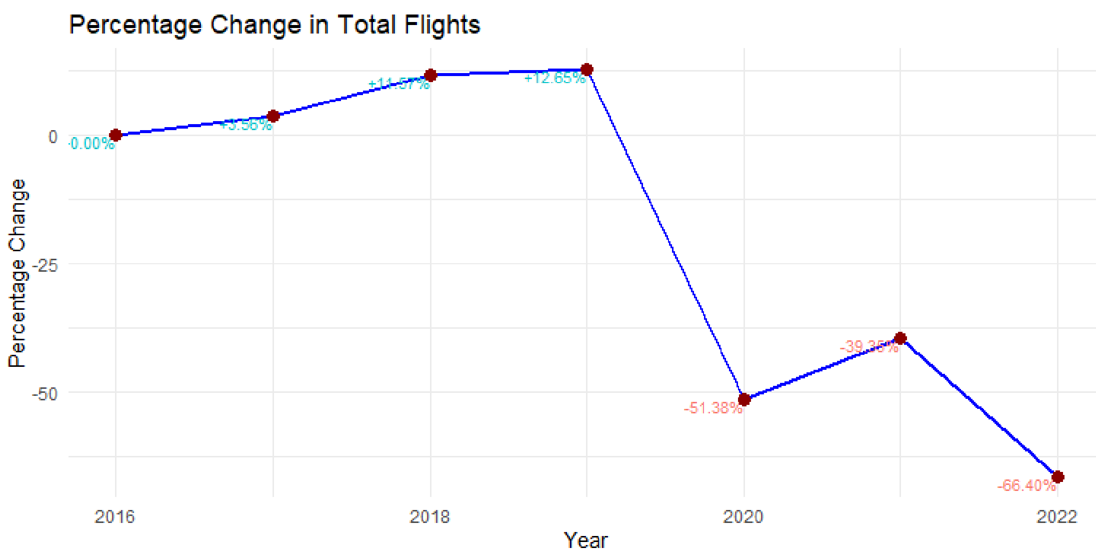

Figure 3- Percentage change in total flights yearly

Graph Analysis

Between 2016 and 2019, the total number of flights in Europe experienced a steady increase of 12.65% due to factors such as the growth of tourism, business activities, and improved ease of movement across European countries. However, the COVID-19 pandemic caused a significant decline in flight movements due to movement restrictions put in place by governments, resulting in a sharp reduction of flights by 58% in 2020. In 2021, as some of these restrictions were eased, there was a gradual increase in flight numbers by 39%. Unfortunately, this trend was short-lived as the early months of 2022 saw a sudden drop in the number of flights, potentially due to changes in the economy such as recessions caused by the Ukrainian war or shifts in consumer spending, impacting demand for air travel and the number of flights operated by airlines. This led to a significant decrease of 66.40% in-flight movement in the same year. Despite these challenges, the aviation industry is adapting to the changing circumstances and implementing measures to ensure the safety of passengers and staff. It is hoped that as the situation improves, the industry will continue to recover and eventually return to pre-pandemic levels of flight movements.

11.2 To check for the total number of flight movements per year using a bar chart

Description

This creates a bar plot of the FLT_TOT_1 variable for each year in the eu_flights data frame. Each bar represents the total number of flights for a given year, and the height of each bar corresponds directly to the value of the FLT_TOT_1 variable. This plot can be used to compare the total number of flights across different years.

Reason for choice of graph

The goal of this visualisation is to plot categorical variables, therefore a bar chart is used because geom_bar() is a good choice for plotting categorical variables and creating a visual representation of a count or frequency distribution.

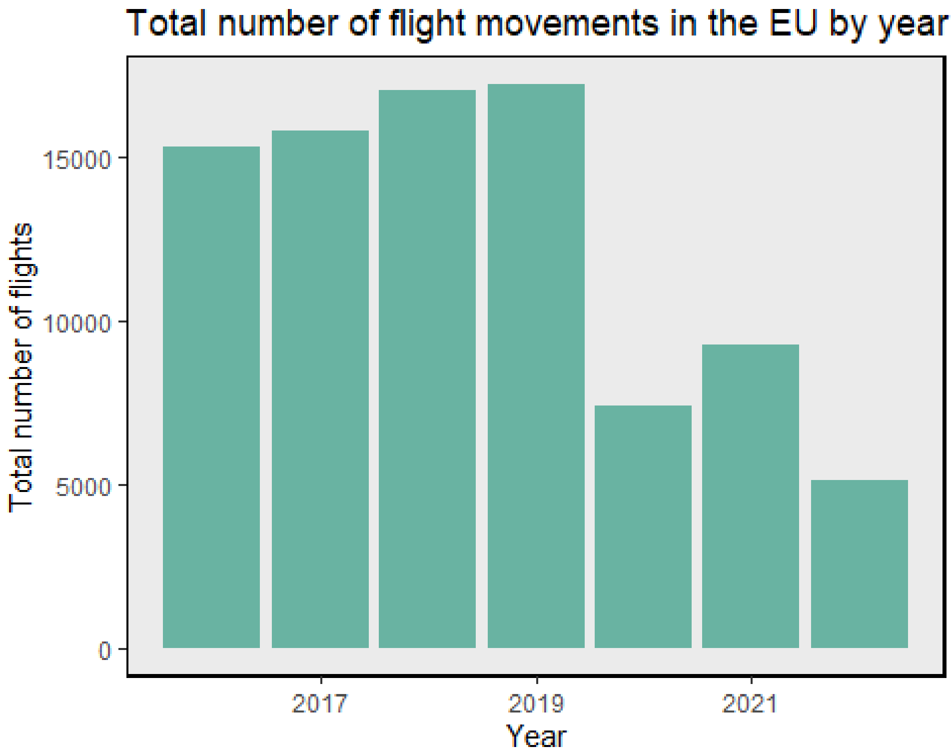

Figure 4 - Total number of flight movements in the EU yearly

Graph Analysis

The bar chart in the visualization represents the total number of flight movements across all airports in Europe for the years 2016 to 2022, based on data from the Network Manager. The chart indicates that 2019 had the highest number of flight movements, with a value close to 11 million. This was followed by 2018, and 2017, which had just over 10.7 million flights. In contrast, 2016 had the fourth-highest number of flight movements. However, the years 2020, 2021, and 2022 had significantly fewer flight movements compared to the previous years. This could be attributed to the Russian-Ukrainian war which had a profound impact on the aviation industry, leading to flight cancellations and reduced passenger demand. The decline in flight movements during these years is visible in the bar chart, with the values for each year dropping significantly compared to the earlier years. Overall, the bar chart provides a clear and concise representation of the trends in flight movements across Europe over the years, highlighting the significant impact of the COVID-19 pandemic on the aviation industry.

12.0 VISUALISING THE EUROPEAN FLIGHT DATASET BY MONTHLY MOVEMENT

The second research query pertains to whether there has been a noteworthy alteration in the number of monthly flight arrivals and departures at airports across various European nations *throughout the pandemic period.* : This question is addressed by the analysis of the plot that follows.

12.1 To check the trends of European flight movements during COVID-19 using line graph and scatter plot

Description

The aim is to generate a clear and visually appealing plot that effectively communicates the trends in monthly flight movement across Europe. The use of color and point markers makes it easy to distinguish between the line and data points, and the x-axis labels provide a clear indication of the time frame being represented.

Reason for Choice of Graph

The purpose of this visualization is to display the relationships and patterns between two variables: FLT_TOT_1 and MONTH_MON. The code employs a line graph to showcase the trend in the total number of flight departures and arrivals per month during the COVID-19 period, depicting the gradual increase or decrease in flight volume over time. Additionally, a scatter plot is used to plot individual data points for monthly flight arrivals.

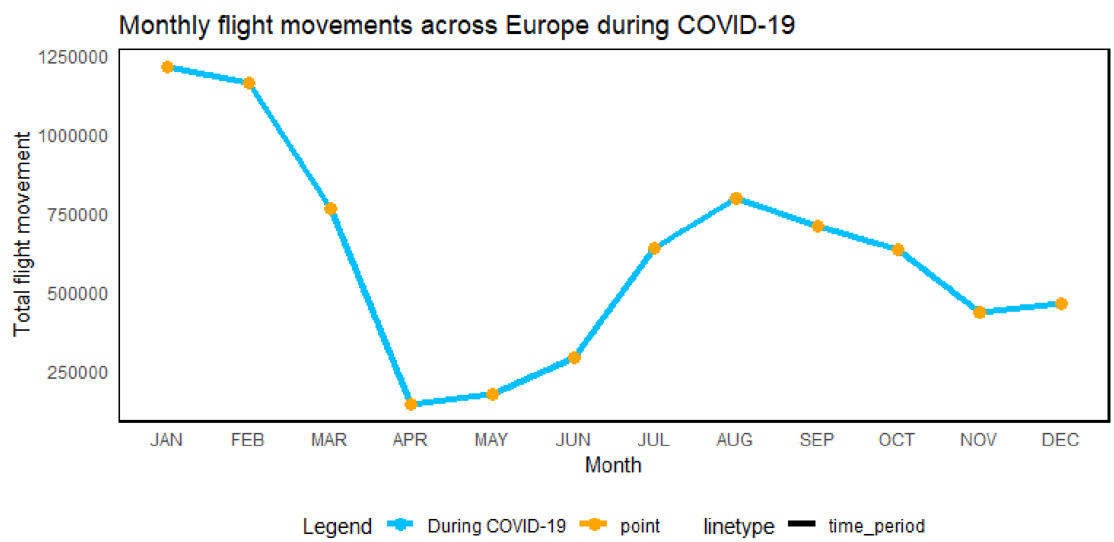

Figure 5 - Monthly flight movement across Europe during COVID-19

Graph Analysis

The data presented in the statement reveals some interesting insights about flight movements in Europe. The fact that August has the highest number of total flight movements and April has the lowest, followed by May, could be because of a variety of factors. For instance, August is the peak summer holiday season in Europe, and many people choose to travel abroad during this time. Conversely, April and May may see a dip in flight movements due to the end of the winter holiday season and the start of spring, which is not a peak travel season. The impact of the COVID-19 pandemic on flight movements is also evident from the statement. While there was a slight decrease in total flight movements from January to February, there was a sharp increase from February to March, likely due to the early stages of the pandemic and the need for people to travel back to their home countries. The pandemic has had a profound effect on the airline industry, with many airlines struggling to stay afloat because of reduced demand for air travel and restrictions on international travel. The line graph presented in the statement also shows a consistent pattern in total flight movements from April to August, indicating that this period is generally the busiest time for air travel in Europe. From September to November, there was a gradual decrease in the total number of flight movements, with a slight increase in December, which could be due to the Christmas and New Year holiday seasons.

In summary, the data presented in the statement highlights the seasonal nature of air travel in Europe and the impact of external factors such as the COVID-19 pandemic on flight movements. Understanding these trends and patterns is important for airlines, airport authorities, and policymakers to make informed decisions about flight scheduling, capacity planning, and travel policies.

12.2 Checking the European flight's departures per month during COVID-19

Description

This code filters and summarises the dataset to include only flights during the COVID period, and creates a bar chart to compare the total flights during COVID-19 for each departure monthly in the EU, with modified labels, a black and white theme, and a blank grid and black border.

Reason for choice of graph

This visualization aims to show the total number of flight departures during the COVID-19 period in the EU, with a focus on each month of the COVID-19 period, hence a bar chart is used to compare the total flight movements for each month.

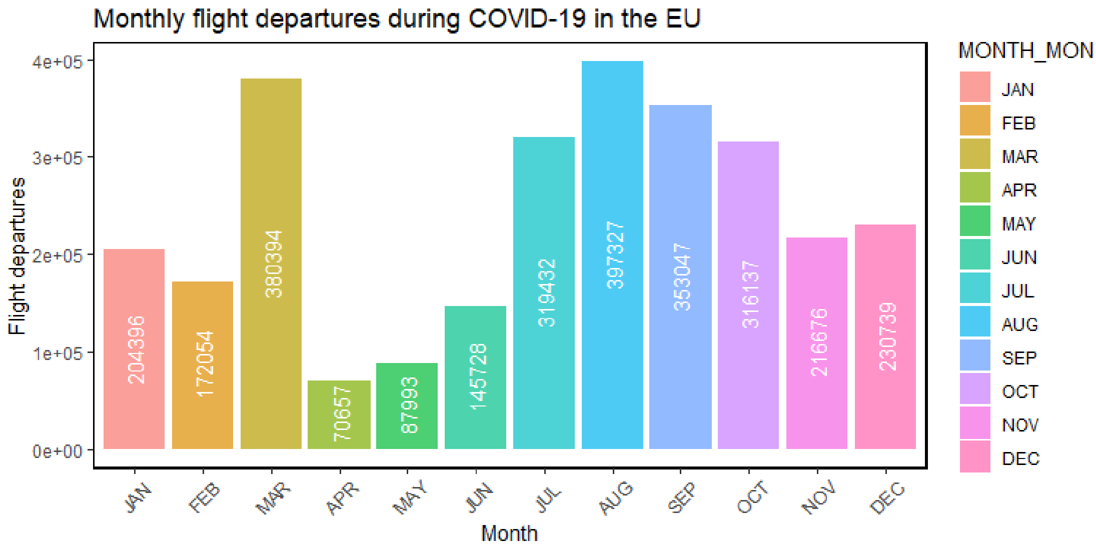

Figure 6 - Monthly flight departures during COVID-19 in the EU

Graph Analysis

The rise in flight activity during the summer months, particularly in August, is a result of the high number of people taking vacations during this time. As the weather gets warmer, individuals often take time off from work or school to explore new destinations. Airlines, therefore, respond to this demand by offering more flights and destinations during the summer months. Conversely, April records the lowest number of flight activities accounting for a total flight departure of 70657 because of COVID-19-related restrictions implemented by some European countries. However, there is a slight decline in flight activity from September to November from 353047 to 216676, which could be attributed to various factors. Firstly, this period is an off-peak season, with fewer individuals traveling due to the resumption of academic sessions in schools and universities. Additionally, the fall and winter seasons are not conducive to travel, particularly for leisure purposes. As a result, airlines often reduce their flight schedules during this period to cut costs and maintain profitability. Interestingly, despite the slight dip in flight activity during the fall and winter months, there is a slight increase in December totaling 230739 flight departures. This can be attributed to the holiday season, with many individuals traveling to be with loved ones during Christmas and New Year's Eve.

Overall, the bar chart highlights a consistent trend in flight activity in Europe, heavily influenced by seasonal factors. Peak travel periods coincide with the summer months and holiday seasons, while off-peak periods occur during the fall and winter seasons

12.3 Checking the European flight's arrivals monthly during COVID-19

Description

The purpose of this code is to extract and condense the *eu_flights dataset*, focusing on flights during the COVID period, and generate a bar chart comparing the total flights during COVID-19 for EU monthly arrivals.

Reason for choice of graph

The purpose of this visualization is to showcase the total count of flight arrivals in the European Union during the COVID-19 period, with a particular focus on individual months within that timeframe. A bar chart will be utilized to depict the data and enable a comparison of the total flights per month during COVID-19. The height of each bar will indicate the overall number of flight arrivals, while the use of color coding will help differentiate the months from each other.

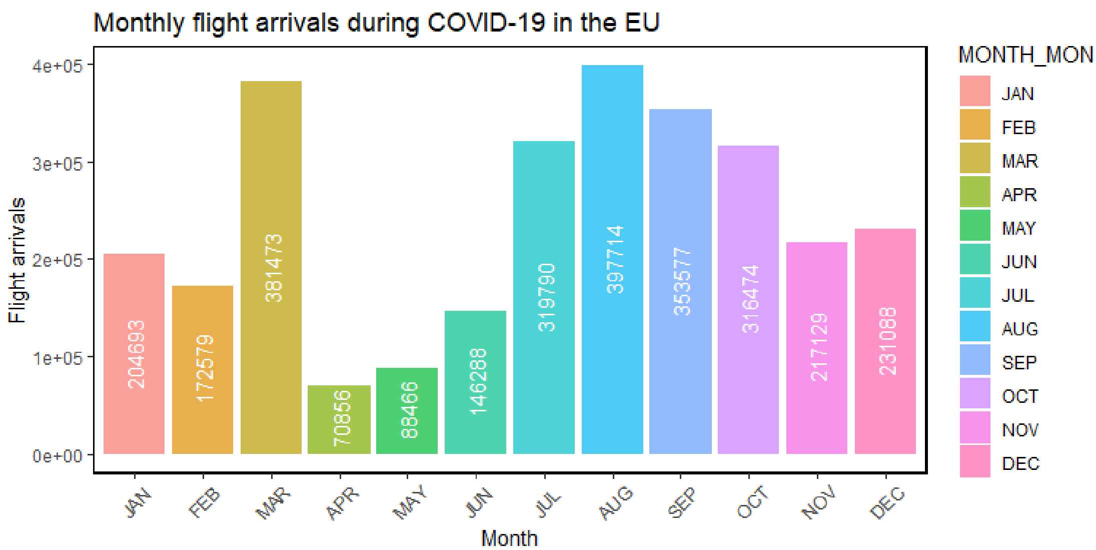

Figure 7 - Monthly flight arrivals during COVID-19 in the EU

Graph Analysis

Flight activities can be lower in January and February, especially after the holiday season, as people return to work and school. Airlines may reduce their flight schedules during this time. However, there was a drastic increase in the number of Flight arrivals as spring break approached, with more people traveling for vacations or to visit family and friends. Airlines may add more flights to popular destinations and offer discounts to attract travelers. April typically has the lowest flight activity due to COVID-19 restrictions dropping from 381473 in March to 70856 in April. Although there is a steady increase in Flight movements from May to August as summer approaches, with more people traveling for vacations and holidays such as Memorial Day weekend. Airlines may add more flights and routes to popular destinations to meet the demand. These months are typically the busiest for air travel, with the summer vacation season in full swing. Airlines add more flights and routes to popular destinations and may charge higher fares due to high demand. Although there was a gradual drop from September to November mainly because of the resumption of academic sessions and the end of the summer vacation season. Airlines may reduce their flight schedules during this time. Flight activity increased slightly to 231088 in December due to the holiday season, with many people traveling to be with family and friends. Airlines respond by offering more flights and routes to popular holiday destinations. Overall, flight activity varies throughout the year due to seasonal and holiday patterns, weather conditions, and other factors. Airlines adjust their flight schedules and fares to meet the demand, and travelers can take advantage of these changes to plan their trips accordingly. Additionally, the bar chart reveals a pattern of flight activity in Europe that is strongly influenced by seasonal factors. The high travel seasons align with the summer months and holiday periods, while the low travel seasons occur during the peak of COVID-19 movement restrictions.

13.0 VISUALISING THE EUROPEAN FLIGHT DATASET BY AIRPORT

The third research question examines the impact of the COVID-19 pandemic on the number of flights at the busiest airports in Europe, comparing the pre-pandemic and post-pandemic periods. : The following plot analysis answers this question

13.1 Confirming the 4 busiest airports by departure across Europe before and after COVID-19 using a bar chart

Description

The plot aims to extract and summarize the *eu_flights dataset* that highlights the busiest airports in Europe based on flight departures, both pre and post-COVID-19. It then generates a bar chart that compares the total number of flight departures across these busy airports, before and after the pandemic.

Reason for choice of graph

The purpose of this visualization is to present the total number of flight departures at the busiest airports in the European Union, both before and after the COVID-19 pandemic. A bar chart is an effective way to compare data across different categories or groups, making it a suitable choice for visualizing the frequency of flight departures from different locations. The bar graph is utilized to compare the overall number of departures for each airport, with the height of each bar representing the number of flights departing from major EU airports. The periods before and after the pandemic are color-coded to distinguish them from one another.

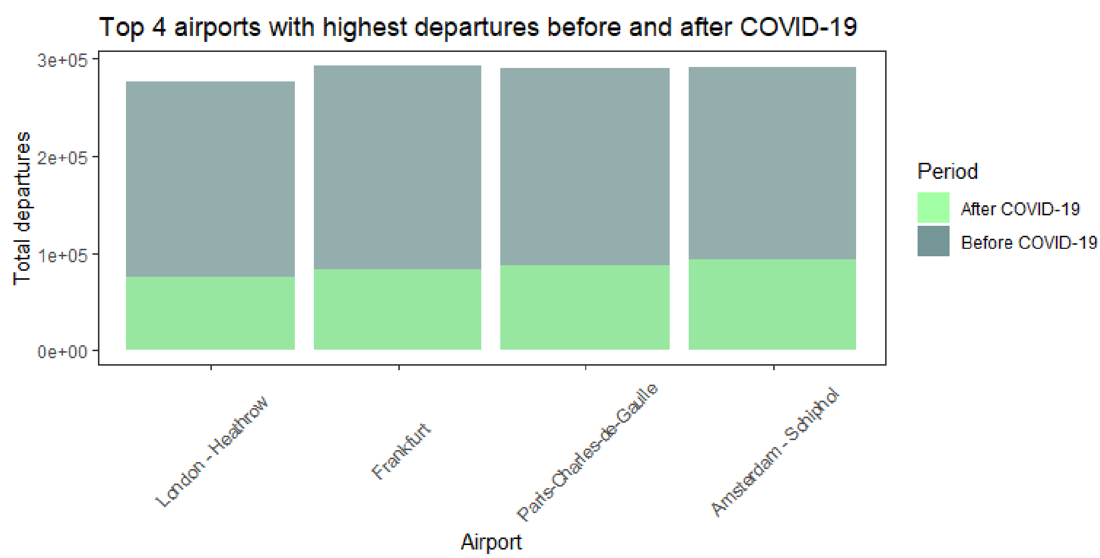

Figure 8 - Top 4 airports with the highest departures before and after COVID-19

Graph Analysis

The COVID-19 pandemic had a significant impact on the aviation industry, leading to a sharp decrease in air travel demand and resulting in reduced flight activity at airports around the world. Among Europe's busiest airports, Frankfurt was the top performer in terms of flight departures prior to the pandemic, with nearly 290,000 departures recorded. This is not surprising, considering Frankfurt is the hub for one of Europe's largest airlines, Lufthansa, and serves as an important connecting point for air travel within Europe and to other regions of the world. Amsterdam-Schiphol ranked second in terms of flight departures, closely following Frankfurt's lead. Amsterdam-Schiphol is a major airport in Europe and serves as the hub for KLM Royal Dutch Airlines. Like Frankfurt, Amsterdam-Schiphol connects travelers to numerous destinations within Europe and beyond. Paris-Charles-de-Gaulle airport ranked third in terms of flight departures, with slightly fewer departures than Amsterdam-Schiphol. Charles-de-Gaulle is the primary international airport serving Paris and is a hub for Air France, the country's flagship carrier. London-Heathrow had the fewest flight departures among Europe's four busiest airports before the pandemic. However, Heathrow is still one of the busiest airports in the world, serving as the primary hub for British Airways and connecting passengers to destinations across the globe. In the aftermath of the pandemic, all four airports saw reduced flight activity. London-Heathrow recorded the lowest number of departures, with many international travel restrictions still in place in the UK. Meanwhile, Amsterdam-Schiphol recorded the highest number of departures among the four airports, as the Netherlands gradually reopened for international travel. Frankfurt and Paris-Charles-de-Gaulle had similar levels of flight departures following the pandemic, as travel restrictions eased and air travel demand slowly began to recover.

13.2 Confirming the 4 busiest airports by arrivals across Europe before and after covid-19 using a bar chart

Description The objective of this plot is to extract and condense the *eu_flights dataset* that emphasizes the busiest airports in Europe based on flight arrivals, both before and following the COVID-19 pandemic. Subsequently, a bar chart is produced to compare the total number of flight arrivals across these airports before and after the pandemic.

Reason for choice of graph

The goal of this visualization is to display the total number of flight departures from the busiest airports in the European Union, both before and after the COVID-19 pandemic. A bar chart is an appropriate method to compare data among various categories or groups, making it a suitable option for visualizing the frequency of flight departures from different locations. The bar chart is utilized to compare the total number of arrivals for each airport, with the height of each bar indicating the number of flights departing from major EU airports. The periods before and after the pandemic are differentiated by color coding.

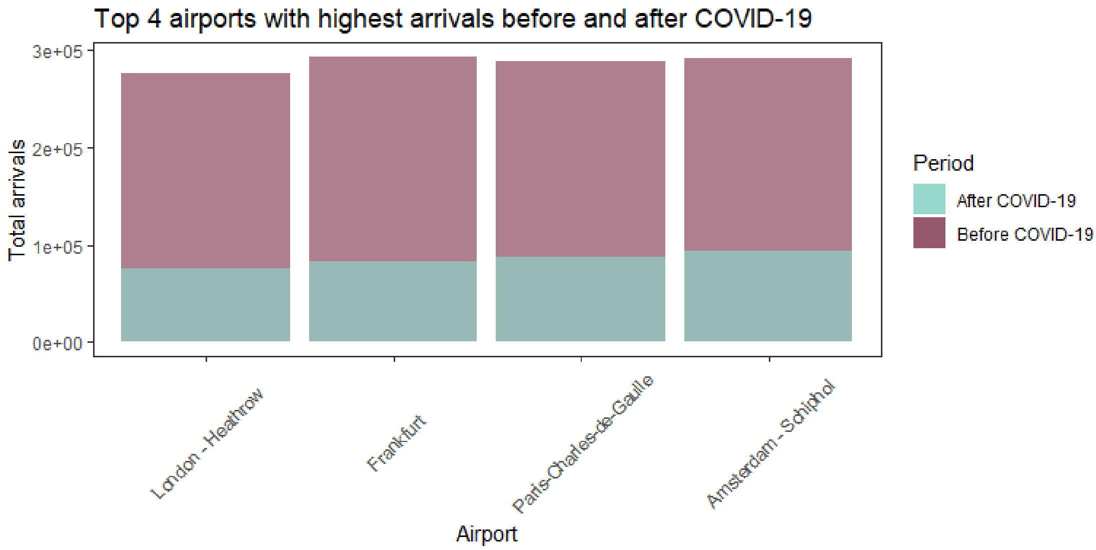

Figure 9 - Top airports with the highest arrivals before and after COVID-19

Graph Analysis

The COVID-19 pandemic had a profound impact on the aviation industry, leading to a decrease in air travel and resulting in the closure of many airports across the world. In Europe, the impact of the pandemic was felt heavily by the busiest airports, which saw a significant decrease in the number of flights arriving and departing. Before the pandemic, the four busiest airports in the European Union, namely Frankfurt, Amsterdam-Schiphol, Paris-Charles-de-Gaulle, and London-Heathrow, experienced an increase in the number of flight arrivals. Frankfurt and Amsterdam-Schiphol were the busiest airports, recording just below 300,000 arrivals, while Paris-Charles-de-Gaulle saw a slight decrease in arrivals compared to the other two airports. London-Heathrow had the fewest arrivals among the four airports.

However, after the pandemic, there was a sharp decline in the number of flights arriving at the four busiest airports in Europe. This was mainly due to the restrictions on international travel and the reduction in economic activities and tourism as a result of the pandemic. Amsterdam-Schiphol remained the busiest airport among the four, recording the highest number of arrivals, followed by Paris-Charles-de-Gaulle and Frankfurt, which both experienced a slight decrease in arrivals compared to Amsterdam-Schiphol. London-Heathrow continued to have the lowest number of total flight activity among the busiest European airports.

13.3 To check the trends of 4 busiest airports in Europe by flight departure from 2016 to 2022

Description

This visualization aims to identify the top four busiest airports in Europe in terms of flight departures between 2016 and 2022. It then presents a line graph that depicts the trends in the number of flight departures across these selected airports.

Reason for choice of graph

This visual representation aims to demonstrate the trend of flight arrivals from the busiest airports in the European Union between 2016 and 2022. A scatter plot is typically utilized to display the relationship between two continuous variables, which is useful for visualizing the distribution of data points and identifying patterns and for displaying the total number of flight arrivals from different airports during this period, a line graph is used to illustrate how data changes over time. The line graph emphasizes the trends in flight arrivals from the top 4 busiest airports in Europe, with each major EU airport being represented by a unique color.

```
## `summarise()` has grouped output by 'YEAR'. You can override using the
## `.groups` argument.
```

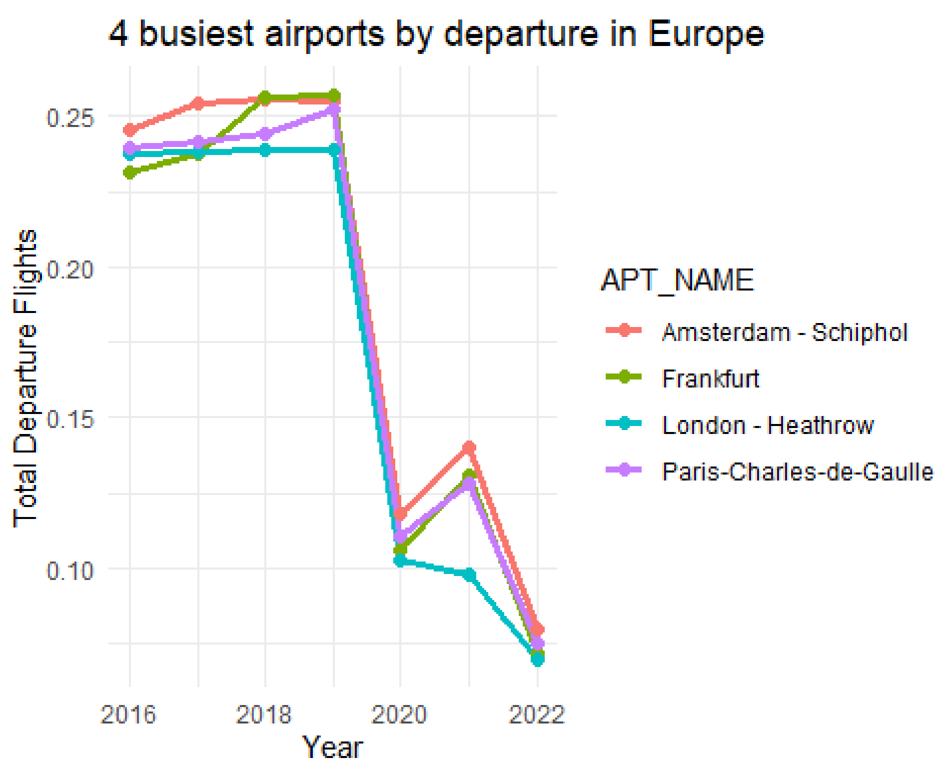

Figure 10 - 4 busiest airports by departure in Europe

Graph Analysis

Before the COVID-19 pandemic, the busiest airports in Europe experienced a steady increase in flight departures from 2016 to 2019, which was attributed to the growth of economic activities across European countries. Frankfurt emerged as the busiest airport during this period, with over 250,000 flight departures in 2019, closely followed by Amsterdam-Schiphol with a slightly smaller number of flight departures. Paris-Charles-de-Gaulle was the third busiest airport, while London-Heathrow remained the least patronized among the top four busiest airports in Europe, with a consistent trend throughout the period.

However, the COVID-19 pandemic had a significant impact on air travel and led to a drastic decline in flight departures from all major European airports in 2020 due to the various restrictions imposed by European countries. There was a slight recovery in 2021 as countries began to open up for movement, increasing the number of flight departures from these airports. However, the early part of 2022 saw a sharp decrease in the number of flight departures because of the Ukrainian war, with London-Heathrow being the most affected, while Amsterdam-Schiphol remained the busiest airport with the highest number of flight departures in 2022.

13.4 To check the trends of 4 busiest airports in Europe by flight arrival from 2016 to 2022

Description

This visualization depicts the trend in flight departures from the busiest airports in Europe from 2016 to 2022, utilizing the *eu_flights dataset*. A line graph is employed to illustrate the change in the frequency of flight departures across these airports over the given period.

Reason for choice of graph

The goal of this visualization is to display the trend in flight departures from the busiest airports in the European Union over 6 years, from 2016 to 2022. A line graph is an appropriate method to compare data among various categories or groups, making it a suitable option for visualizing the frequency of flight departures from different locations. The line graph is utilized to show the pattern followed by the 5 most patronized airports in Europe. with the height of each bar indicating the number of flights departing from major EU airports. Each European airport is differentiated by color coding.

```
## `summarise()` has grouped output by 'YEAR'. You can override using the
## `.groups` argument.
```

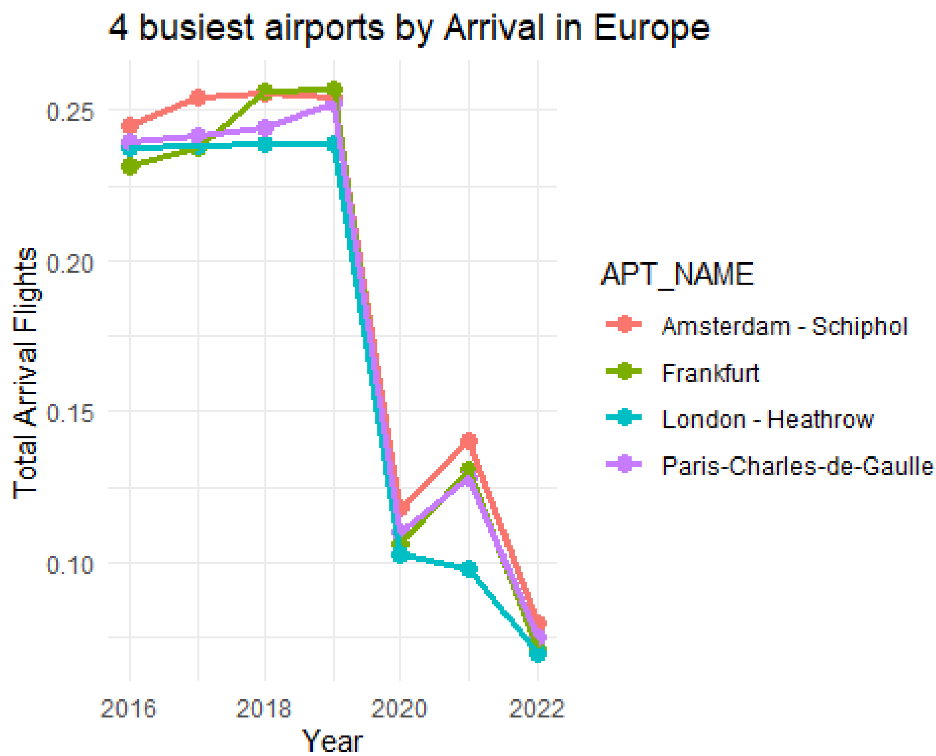

Figure 11 - 4 busiest airports by arrival in Europe

Graph Analysis

The information provided describes the trend in flight arrivals at the busiest airports in the European Union over 6 years from 2016 to 2022. Between 2016 and 2019, the airports experienced a gradual growth in the number of flight arrivals due to increased economic activities across European countries. Frankfurt had the highest peak in flight arrivals, accounting for over 250,000 flight departures in 2019, followed by Amsterdam-Schiphol with a small margin when compared to Frankfurt. Paris-Charles-de-Gaulle was the third busiest airport in the same year while London-Heathrow was the least patronized airport among the top 4 busiest airports in Europe. However, in 2020, these airports saw a significant decrease in the number of flight arrivals due to COVID-19 restrictions imposed by some European countries to curtail the spread of the pandemic. There was a little recovery in 2021 as countries began to open up for business activities and tourism, increasing the number of flight arrivals at these most patronized airports in Europe. In 2022, there was a sharp fall in the number of flights that arrived at the airports, with London-Heathrow being the most affected in terms of recorded flight arrivals, while Amsterdam-Schiphol Airport still recorded the highest number of flight arrivals in 2022. This suggests that the COVID-19 pandemic had a significant impact on the aviation industry in Europe, with many airlines reducing their flight schedules or completely suspending operations.

14.0 CONCLUSION

In recent years, air travel has become increasingly affordable and accessible, resulting in a surge of passengers globally, including in Europe. However, the outbreak of COVID-19 and the need to prevent its spread led to various restrictive measures being implemented by countries worldwide since the start of 2020, impacting the air transport industry. Nevertheless, recent statistics on commercial flights indicate a recovery.

Eurocontrol and Ec.europa.eu record of commercial flights in Europe has yielded valuable insights into the aviation industry's dynamics and the impact of the COVID-19 pandemic on air travel. As the industry adapts and recovers in the post-pandemic world, this dataset will continue to be an essential resource for policymakers, researchers, and industry stakeholders. Predictive models that factor in seasonal flights and pandemic periods can also help airlines and airports optimize their resources and schedules for more cost-effective and efficient operations and by analyzing the European flight dataset, researchers can uncover valuable patterns and insights that shed light on the aviation industry.

Overall, Future research on the European flight dataset should focus on analyzing the long-term effects of the COVID-19 pandemic on the aviation industry, including changes in airline and passenger behavior, airport infrastructure, and government policies. Researchers can also explore the potential impact of emerging technologies, such as electric and autonomous aircraft, on the industry. In addition, there is a need for research on sustainable aviation, including the development of alternative fuels and carbon capture technologies, as the industry seeks to reduce its environmental footprint. Finally, there is a need for continued data collection and analysis to monitor trends in the European aviation industry and inform policy decisions.

15.0 REFERENCE

https://ansperformance.eu/data/

https://ec.europa.eu/eurostat/web/products-eurostat-news/-/ddn-20210914-1

Suzumura, T., Kanezashi, H., Dholakia, M., Ishii, E., Napagao, S.A., Perez-Arnal, R. and Garcia-Gasulla, D. (Dec 10, 2020) The Impact of COVID-19 on Flight Networks. IEEE, pp. 2443.

Burghouwt, G. and Hakfoort, J. (2001) The evolution of the European aviation network, 1990–1998 Elsevier BV.

16. 0 APPENDIX

```
---
title: "DATA VISUALIZATION COURSEWORK 2 - EUROPEAN FLIGHT DATASET"
author: "UMUNNA AUSLAR CHIMUANYA"
date: "2023-03-20"
output:
  word_document:
    toc: yes # toc = table of contents
    toc_depth: 2 # refers to the number of "levels" of headings shown
  pdf_document:
    toc: yes
    toc_depth: 2
    fig_width: 7
    fig_height: 6
    fig_caption: yes
bibliography: Bibliography.bibtex
---
```

````
```{r setup, include=FALSE}
knitr::opts_chunk$set(echo = FALSE, warning = FALSE)
```
````

ABSTRACT

This coursework provides a comprehensive analysis of the European flight dataset, which was sourced from Eurocontrol and the official website of the European Union. The report starts by giving an overview of the dataset and then poses three research questions to guide the analysis. To answer the research questions, the report employs exploratory data analysis techniques and various visualisation techniques to identify patterns and relationships in the dataset. The report’s visualisations provide clear and concise answers to the research questions and reveal interesting trends such as the monthly variation in air traffic during COVID-19, yearly flight movements, the percentage change in flight movement per year, seasonal trends, and the most popular airports before and after COVID-19. The report concludes that April was the most affected month during COVID-19, 2020 had the lowest flight movement, and air traffic has a clear seasonal pattern. The report uses R Studio and R Markdown to generate the final report and includes detailed explanations and code for each visualisation. The report provides valuable insights into the patterns of air traffic in Europe that can be used by researchers to analyze the traffic in major European airports, identify patterns in flight data, assess the performance of airports, inform policy decisions, and improve the safety and efficiency of air traffic control.

1.0 INTRODUCTION

For several years, Eurocontrol and Ec.europa.eu have been gathering data on European flights. Ec.europa.eu is the official website of the European Union, which provides information on a range of topics related to the EU, including policies, news, and events. It also hosts several databases and portals, including the European Statistical Data Portal and the European Data Portal, which provide access to a wide range of statistical data and information related to the EU. The European Statistical Data Portal includes data on transport, including aviation, while the European Data Portal provides access to datasets collected by various EU institutions and agencies, including Eurocontrol. Their data includes passenger numbers, cargo volumes, and aircraft movements, obtained from national authorities and other sources across the EU. Eurocontrol, established in 1960, has been collecting data on air traffic in Europe since its inception. Their data includes metrics such as the number of flights, delays, and route patterns and is collected from air traffic control centres throughout Europe. Throughout the COVID-19 pandemic, both organizations have continued to collect data, which has been essential in shaping policy decisions and strategies in the aviation industry and providing insights into the pandemic's impact on air travel in Europe.

The European Union (EU) flight dataset contains vast amounts of information on commercial flights operating within the EU and flights departing from or arriving at EU airports, airline destinations, monthly flight movement, and airport traffic volumes, making it a valuable resource for understanding the European aviation industry's dynamics and tracking trends and patterns in air travel. Before the pandemic, the EU flight dataset reflected the industry's long-term growth, showing a steady and consistent increase in air traffic.

1.1 BACKGROUND

Air travel has become more affordable and accessible in recent decades, leading to a rise in passenger numbers globally, including Europe. The Burghouwt and Hakfoort (2001) paper titled "The evolution of the European aviation network, 1990–1998" explores the changes in air travel demand, airline competition, and route structures during this period. The paper finds that competition among airlines rose due to a significant increase in air travel demand, leading to a shift in airline strategies from point-to-point routes to hub-and-spoke systems. However, the COVID-19 pandemic had an unprecedented impact on the aviation industry, drastically altering the patterns and trends of the EU flight dataset. Governments worldwide imposed travel restrictions and closed borders, leading to a sharp decline in air travel. Comparing the EU flight dataset before and after the COVID-19 pandemic provides insights into the impact of the crisis on the aviation industry. The Suzumura et al. (2020) paper titled "The Impact of COVID-19 on Flight Networks" examines the effects of the COVID-19 pandemic on global flight networks using a dataset of global flight information. The authors found a significant reduction in flight volume and airline connectivity globally, with more pronounced effects in regions with higher infection rates and stricter travel restrictions. According to the paper, air traffic in Europe was heavily impacted, with a significant reduction in flight volume and airline connectivity. Airlines have also shifted their focus from international to domestic routes, particularly evident in Europe, where intra-European routes have become more critical.

As air travel begins to recover from the COVID-19 pandemic, the European flight dataset is expected to continue to be an invaluable resource for understanding the changing patterns and trends in air travel. Researchers can use the dataset to monitor the recovery of the industry, identify areas of growth and opportunity, and assess the effectiveness of policy decisions and strategies in the aviation sector. With the continued expansion of the European aviation market, it is likely that the dataset will continue to grow in size and complexity, presenting new opportunities for analysis and research.

This report mainly focuses on the European flight dataset and provides data on the volume and movement of flights that occurred pre-COVID-19, during the COVID-19 pandemic, and post-COVID-19. The report highlights significant airports across different European countries and showcases proficiency in data visualization by creating compelling and precise visuals that convey a meaningful narrative. The dataset can help researchers understand how airlines and airports have responded to the pandemic, the effects of the pandemic on passenger behavior, airline performance, and airport traffic volumes, and the long-term prospects of the aviation industry as it faces ongoing challenges from the pandemic.

1.2 AIM

The aim of this coursework is to showcase expertise in the techniques taught in the class and beyond, and apply them in a meaningful manner to a new dataset by creating visually appealing, authentic, and coherent narrative visualizations. The task involves investigating European flight dataset, crafting a story around it, and presenting it through at four distinct types of visualizations (or graphs).

1.3 OBJECTIVE

The objective of this report European flight dataset is to provide information on the number of flights that occurred monthly, yearly and in major airports across various European countries. This dataset can be used to analyze the traffic in seasons and airports, identify patterns in the flights, and assess the performance of airlines and airports. This coursework will further demonstrate the proficiency in data visualization by creating compelling and accurate visuals that effectively communicate a story. The final report for this assignment should utilize a minimum of three distinct visualization techniques and employ R Studio and R Markdown. The objective is to create a report that is both informative and engaging.

2.0 METHODS USED TO ACCESS EUROPEAN FLIGHT DATASET

Two methods are presented to access and load the "flights" dataset, which is part of the tidytuesdayR project.

The first approach involves using the tidytuesdayR package, which can be installed from CRAN via the command install.packages("tidytuesdayR"). After installation, the tt_load() function can be utilized to load the dataset for a particular week of interest, taking either an ISO-8601 date or year/week format as input. The loaded dataset is then assigned to a variable, such as flights.

The second method is to manually read in the data using the read_csv() function from the readr package. The data is available at a specific URL, which is passed as an argument to read_csv(). The resulting dataset is then assigned to a variable, flights.

Both techniques accomplish the same goal, which is to load the "flights" dataset into the R environment for further analysis. The selection of a method is dependent on individual preference, ease of use, and factors like internet connectivity and package availability. This report uses manual method to access and load the flight dataset.

3.0 RESEARCH QUESTIONS

For a clear and focused inquiry that seeks to address a specific issue or problem, in the context of the European flight dataset before and after COVID-19, a research question can help guide an investigation into the impact of the pandemic on the European aviation industry.
These research questions can help guide the selection of variables to analyze, the statistical methods to use, and the conclusions to draw in the EU flight dataset.

1. How has the total number of flights in the European Union changed over the years?

One approach could be to obtain data from official sources such as Eurostat, which provides statistics on air transport in the European Union. The data can also be obtained from individual airports or airlines that operate in the EU. Then analysation on how to determine how the total number of flights has changed over the years by creating visualizations such as line charts or bar graphs to illustrate the changes in the total number of flights over time. In interpreting the data, it is important to consider any external factors that may have affected the total number of flights. For example, changes in the economy, pandemic, or government policies could all impact the number of flights.

2. Were there specific months or seasons during the COVID-19 pandemic where flights were more heavily impacted in Europe?

This would require analyzing data on flight schedules and travel restrictions for different months and seasons during the pandemic.
To go about this, data from relevant sources will be collected, such as airport authorities, airlines, or official government statistics. The data should cover a period starting from the beginning of the pandemic till the end of the pandemic, and should include information on the number of flights and any travel restrictions or regulations that were in place.
After data collection, then there will be analysis to identify any trends or patterns in flight volumes across different months and seasons. This might involve looking at factors such as the number of COVID-19 cases in different countries, changes in travel restrictions or regulations, or fluctuations in demand for air travel. After analyzing the data, then there will be conclusions about which specific months or seasons were more heavily impacted by the pandemic in terms of air travel. For example, you might find that air travel was significantly reduced during the winter months due to increased COVID-19 cases and restrictions on non-essential travel.

3. How has the COVID-19 pandemic affected the busiest airports in Europe in terms of the number of flights before and after the pandemic?

This research question would involve collecting data on flight schedules, passenger traffic, and any travel restrictions or regulations that were in place for the busiest airports in Europe.
To go about this, there will be identification of the busiest airports in Europe, and then collect data from relevant sources, such as network Manager, airport authorities, airlines, or official government statistics. The data should cover a period starting from before the pandemic till the present, and should include information on the number of flights, passenger traffic, and any travel restrictions or regulations that were in place. Once the data has been collected, then it can be analyzed to identify any changes in flight volumes at the busiest airports in Europe before and after the pandemic. This might involve comparing data from the same time period in different years, looking at changes in the number of flights and passenger traffic, or examining any trends or patterns in the data. After analyzing the data, conclusions will be drwan about how the pandemic has affected the busiest airports in Europe in terms of the number of flights. For example, it might find that the number of flights at these airports dropped significantly during the early months of the pandemic due to travel restrictions and reduced demand for air travel, and that there has been a gradual recovery in flight volumes in more recent months.

4.0 EUROPEAN FLIGHT DATASET SOURCES

The "Network Manager" and "Airport Operator" in the European flight dataset are important data sources that provide valuable information for analyzing air traffic patterns in Europe. The Network Manager data source refers to the pan-European air traffic control organization, Eurocontrol, which manages and coordinates air traffic across European airspace. This source provides a complete dataset of European flights, making it a valuable resource for researchers, policymakers, and industry stakeholders. On the other hand, the Airport Operator data source in the given european flight dataset refers to the organization or company that operates the airport from which the flight departed or arrived. While this column also provides important information, it often contains a large number of missing values (NA values). This is because not all airport operators have the resources to collect and provide detailed flight data. As a result, the dataset provided by airport operators may not be as comprehensive as the one provided by the Network Manager.

5.0 CHARACTERISTICS OF VARIABLES OF INTEREST

In R, a variable is a name that represents a value. It can be a numeric value, a character string, a logical value (TRUE or FALSE), or any other data type. Variables are used to store and manipulate data in R programs. To assign a value to a variable in R, you can use the assignment operator <-. The European Flight dataset consists of various variables. These variables are used to clearly define the flight operation processes. They include:

1. YEAR

This variable in the European Flights dataset indicates the year in which the data for each observation was collected. This is important information because air traffic patterns and statistics can vary significantly from year to year, depending on factors such as economic conditions, industry trends, and regulatory changes. The year variable can be used to analyze trends in air traffic over time and identify changes or patterns in air travel. For example, changes in the number of flights or passengers from one year to the next can provide insights into the overall growth or decline of the air travel industry, and can be used to inform policy decisions related to air transport.
In addition, the year variable can be used to compare air traffic patterns across different years, which can provide valuable information for forecasting future air traffic and planning for infrastructure and resource needs.

2. FLT_DATE

This is used to represent the date, month and year of flight operations. It is a date-time variable that captures the exact time of takeoff of the flight. The variable is represented in the format of "YYYY-MM-DD", where YYYY refers to the year, MM refers to the month, and DD refers to the day of the month. This variable is useful for analyzing seasonal patterns in air travel and identifying trends in air traffic throughout the year. For example, researchers may use this variable to identify peak travel periods and compare them to other variables in the dataset, such as passenger volume or cargo weight. It can also be used to analyze the impact of weather patterns on air travel, as severe weather conditions may cause flight cancellations or delays.

3. APT_ICAO

This variable refers to the International Civil Aviation Organization (ICAO) code for the airport. The ICAO assigns unique 4-letter codes to airports worldwide, which are used for communication between pilots and air traffic controllers, as well as for tracking flights and air traffic management purposes. The ICAO code is distinct from the International Air Transport Association (IATA) code, which is a 3-letter code used for passenger and baggage handling and ticketing purposes.

The APT_ICAO variable in the European flight dataset is a critical identifier for airports in Europe and is used to track flight operations, airport traffic, and passenger flow. This information is crucial for air traffic management, aviation safety, and policy decision-making. The APT_ICAO code can also provide information about the location and size of the airport, its infrastructure, and its operational capabilities.

4. APT_NAME

This refers to the name of the airport where the flight departed from or arrived at. It is a variable that provides information on the location of the airport and can be used to identify specific airports in the European Union. The name of the airport can be important in understanding the flight data and can provide insights into the airline industry, such as identifying popular or busy airports, or understanding flight routes and connections. This variable can be useful for analyzing flight data in terms of airline traffic, route planning, and passenger behavior, among other things.

5. STATE_NAME

This represents the name of the country where the airport is located. In the context of the European Flights dataset, STATE_NAME refers to the country within the European Union where the airport is located. This variable is useful for analyzing air transport trends in different countries and for making comparisons between different countries in the European Union.

6. FLT_DEP_1

This variable in the European flight dataset represents the number of flight departures from an airport. This variable is an important indicator of air traffic activity and is used to analyze trends in air traffic at specific airports and in different countries. The data for this variable is sourced from the Network Manager, which is responsible for the European Air Traffic Management network. The Network Manager collects and manages data on air traffic movements in real-time, including flight departures. This data is used to support air traffic management and planning, as well as for research and analysis purposes. The FLT_DEP_1 variable in the European flight dataset provides valuable insights into the volume and frequency of flight departures from different airports in Europe.

7. FLT_ARR_1

This is a variable in the European flight dataset which indicates the number of flights that arrived at the airport on a given day. This data is sourced from the Network Manager, which is responsible for the operation and development of the European air traffic management network. The variable is useful for analyzing airport traffic patterns, assessing airport performance, and planning airport infrastructure development."

8. FLT_TOT_1

This represents a variable in the European Flight Dataset that represents the total number of flight movements at an airport. Flight movements include both departures and arrivals,

and FLT_TOT_1 provides a measure of the overall level of air traffic at a given airport. The data for this variable is sourced from Network Manager, which collects and manages air traffic data in the European Union.
It can be used to analyze trends in air traffic at different airports, compare levels of air traffic between different regions, and identify airports that are experiencing particularly high or low levels of traffic. This information can be useful for airport management, airlines, and policy makers in making decisions related to airport capacity, infrastructure investment, and air traffic control.

9. FLT_DEP_IFR_2

This European flight dataset represents the number of Instrument Flight Rules (IFR) flight departures from the airport. IFR refers to a set of rules and regulations that pilots must follow when flying in weather conditions where visibility is limited. The variable is sourced from the airport operator and provides insight into the number of flights that departed the airport using IFR during the specified time period. This information is important for analyzing flight patterns and airport operations, as well as for assessing the impact of weather conditions on flight departures.

10. FLT_ARR_IFR_2

This variable represents the number of Instrument Flight Rules (IFR) flight arrivals at an airport. IFR refers to the set of rules and regulations that govern the operation of aircraft when flying in conditions where visibility is limited, such as in clouds or poor weather. These rules require pilots to use instrument-based navigation and communication systems to maintain safe and accurate flight paths. The data for this variable is typically collected and reported by the airport operator, who is responsible for tracking flight arrivals and departures. This information is often used to monitor airport activity, track flight delays, and inform airport capacity planning.

11. FLT_TOT_IFR_2

This is a variable that illustrates the total number of flight movements that involve Instrument Flight Rules (IFR) at an airport. IFR is a set of rules and procedures that govern the operation of aircraft when flying in conditions where visibility is reduced, such as in clouds or fog. In these conditions, pilots must rely on their instruments and air traffic control to navigate and avoid other aircraft. The data for FLT_TOT_IFR_2 is obtained from the Airport Operator, who collects data on all flights that use their airport, including those operating under IFR. This variable can provide insight into the overall air traffic at an airport and can be useful for analyzing trends in air traffic volume, identifying busy periods, and informing airport operations and planning.

```
![Figure 1-European flights Dictionary](data.PNG)
```

6.0 LOADING AND INSTALLING LIBRARY

Loading a library in R means making the functions, data, and other features of the library available to be used in your R session. After installing a library using the install.packages() function, you need to load it into your R session using the library() function.
When you load a library, all the functions and data sets defined in that library become available in your R session. You can then use these functions and data sets in your R code without having to define them yourself.

In this report, tidyverse package will be used in collection of R packages designed to make data manipulation, exploration, and visualization easier and more intuitive. It is a popular choice for data analysis and is widely used in the R community.

RMarkdown is a powerful tool in R for creating reproducible reports, presentations, and documents that combine R code, text, and figures.
In this report, read.csv will be used to import dataset, this is because read.csv() is a base R function that reads comma-separated values (CSV) files into R as a data frame. It

is a simple and easy-to-use function that is built into R.

```{r, echo=FALSE}
library(tidyverse)
library(ggplot2)
library(rmarkdown)
library(ggridges)


```

7.0 IMPORTING DATASET AND LOADING LIBRARY

Importing a dataset in R means reading the data from an external file and storing it in R's memory as a data object. This allows you to manipulate and analyze the data using R functions and packages. Importation of a dataset is the first step in data analysis with R. R supports a wide variety of data formats, including CSV, Excel, SPSS, SAS, and more. Once you have imported the data into R, you can use various functions to explore, manipulate, and visualize the data. R provides a number of built-in functions and packages for importing different types of data files. The most commonly used functions for importing datasets in R are read.csv(), read.table(), read_excel() (from the readxl package), read_spss() (from the haven package), and read_sas() (from the haven package). The specific function you use will depend on the format of your data file.
Overall, importing datasets in R is a critical first step in data analysis and requires careful attention to ensure accurate and reliable results.

7.1 Importing European flight dataset

Thisreads in data from a CSV file called "flights.csv" and stores it in a variable named **eu_flights dataset**.
eu_flights is a new variable that has been assigned the dataset that was imported as a tibble

```{r}


eu_flights <- read_csv("flights.csv")


```

8.0 CLEANING/WRANGLING PROCESS

Data cleaning and data wrangling are two important stages in the data preparation process that are used to transform and organize raw data into a format that can be used for analysis.
Data cleaning is the process of identifying and correcting or removing errors, inconsistencies, and inaccuracies in the data. This includes dealing with missing values, removing duplicate records, handling outliers, and ensuring that the data is in the correct format.
Data wrangling, also known as data munging, is the process of transforming and reformatting the data into a more usable format. This involves extracting, merging, and restructuring data from multiple sources, converting data types, creating new variables, and reshaping data to meet the needs of the analysis. Together, data cleaning and data wrangling help ensure that the data is accurate, complete, and in a format that is appropriate for analysis. They are essential steps in the data preparation process and are often iterative, meaning that the process is repeated until the data is ready for analysis.

9.0 DATASET EXPLORATION

9.1 Confirming the dataframe structure

This shows the number of rows in the *eu_flights dataset*
```{r}
# calculate number of rows
  nrow(eu_flights)


```

9.2 Confirming the dataframe structure

This returns the number of columns in the *eu_flights dataset*
```{r}
# calculate number of columns
ncol(eu_flights)
```

9.3 Confirmation of the class variables in the dataset

This applies the class type function to each column of the *eu_flights dataset* and returns a vector of the class of each column.
```{r}

# apply class function to each column
  sapply(eu_flights, class)


```

9.4 Confirmation of the missing values in each column of the *eu_flight dataset*

This calculates the number of missing values in each column of the *eu_flights dataset*. The output is a vector that shows the number of missing values in each column of the dataset. This information can be used to determine which variables have the most missing data and may need to be further investigated or imputed.
```{r}
# Check missing values in each column
sapply(eu_flights, function(x) sum(is.na(x)))


```

9.5 Confirming how many missing values are in the European dataset.

This calculates the total number of missing values in the *eu_flights data frame* regardless of which column the missing values occur in.
```{r}
#  Check total number of missing values
sum(is.na(eu_flights))


```

9.6 Confirming the structure of *eu_flights dataset*

This will display information about the data frame's structure, including the names and data types of each variable in the data frame, the number of observations, and any attributes associated with the data frame.
```{r}

# Check the structure
str(eu_flights)


```

9.7 To check for the 1st to 6th rows of the *eu_flights dataset*

This displays the first few rows of the dataset. Specifically, it shows the first 6 rows of the *eu_flights dataset*, allowing for quick preview of the structure and content of the dataset. This is a commonly used function in data analysis and is useful for getting an initial sense of what kind of data being worked. *eu_flights dataset*
```{r}

# Check for
head(eu_flights)


```

9.8 Confirming the glimpse of the dataset

This is used to preview the structure and contents of a data frame, including variable names, data types, and a sample of the data itself

```{r}

# Check data overview
glimpse(eu_flights)

```

9.9 Confirming the summary of the dataset

This provides a summary of the *eu_flights dataset*. The summary() function provides basic statistical information on a given data frame or vector, including minimum and maximum values, median, mean, first quartile (25th percentile), the third quartile (75th

percentile), and any missing values
```{r}

# Check summary
summary(eu_flights)


```

9.10 To confirm the dimensions of the dataset

This presents the dimensions of the *eu_flights dataset*, which is the number of rows and columns in the dataset. The output is a vector with two elements: the number of rows and the number of columns in the data frame. The first value indicates the number of rows in the data frame, and the second value indicates the number of columns.

```{r}

# Check dimensions
dim(eu_flights)



```

10.0 CORRELATION COEFFICIENT

10.1 Confirming the correlation between FLT_TOT_1 and FLT_TOT_IFR_2

Description

This selects specific columns (YEAR, FLT_TOT_1, and FLT_TOT_IFR_2) from the eu_flights dataset, removes any missing values, and calculates the correlation between FLT_TOT_1 and FLT_TOT_IFR_2. It then creates a scatterplot with these variables on the x- and y-axis to show the strength and direction of their relationship. A correlation coefficient closer to 1 or -1 indicates a strong positive or negative relationship, while a value closer to 0 indicates a weak or no relationship. The output of this plot shows a correlation value of 0.95, which indicates a strong positive linear relationship between FLT_TOT_1 and FLT_TOT_IFR_2.

Reason for choice of Variable

The total number of flight movement (total number of flight arrival and total number of flight departure) provided by Network Manager which is denoted by the variable FLT_TOT_1 has no missing value, while there are missing values in FLT_TOT_IFR_2 (total number of flight arrival and total number of flight departure) which is the total number of flight movement provided by Airport OPerator, this variable will be dropped, and FLT_TOT_IFR_2 will be used instead in this report.

Choice of Graph

The aim of this visualisation is to plot a graph showing the correlation between FLT_TOT_1 and FLT_TOT_IFR_2 hence, a scatter plot and a line graph is used to visualize the relationship between the two variables.

```{r, fig.width=8}
# Calculate total movements for FLT_TOT_1 column, ignoring NA values
     flights_tot <-
```

```
    eu_flights %>%
    group_by(`STATE_NAME`) %>%
    summarize(total_movements =
    sum(`FLT_TOT_1`,
    na.rm = TRUE))

# Calculate total movements for FLT_TOT_IFR_2 column, ignoring NA values
    flights_tot_ifr <-
    eu_flights %>%
    group_by(`STATE_NAME`) %>%
    summarize(total_movements =
    sum(`FLT_TOT_IFR_2`,
    na.rm = TRUE))

# Calculate the correlation between FLT_TOT_1 and FLT_TOT_IFR_2
    correlation_value <-
    cor(flights_tot$total_movements,
    flights_tot_ifr$total_movements)

# Create the scatter plot
    ggplot(data =
    data.frame(flights_tot$total_movements,
    flights_tot_ifr$total_movements),
    aes(x = flights_tot$total_movements,
    y = flights_tot_ifr$total_movements)) +

# Add scatter point to the plot
    geom_point(size = 3,
    color = "black") +
    geom_smooth(method = "lm",
    se = FALSE,
    color = "blue") +

# Add the line graph
    labs(x = "Total flight movements FLT_TOT_1",
    y = "Total flight movement FLT_TOT_IFR_2",
    title = "Correlation between total movements for FLT_TOT_1 and FLT_TOT_IFR_2") +

# Add theme to the plot
    theme_minimal() +
    theme(panel.grid.major =
    element_blank(),
    panel.grid.minor =
    element_blank(),
    panel.border =
    element_rect(color = "black",
    fill = NA, size = 1)) +

# Set annotation to the plot
    annotate("text", x = Inf,
    y = Inf,
    hjust = 3.5,
    vjust = 1,
    label =
    paste("Correlation: ",
    round(correlation_value, 2)),
    size = 4)


```

Graph Analysis - As the value for total number of flight movements provided by Network Manager data source increases, total number of flight movements provided by Airport Operator data source () also increases, and this shows a strong correlation between the two variables.

11.0 VISUALISING THE EUROPEAN FLIGHT DATASET BY YEARLY MOVEMENT

The initial research inquiry seeks to determine the manner in which the overall quantity of flights in the European Union has evolved throughout the years.: The following plot analysis answers this question

11.1 To check the trends of European flights from 2016 to 2022

Description

This takes the *eu_flights dataset*, groups it by year, calculates the percentage change in the total number of flights for each year, and assigns the result to a new data frame named eu_flights_1. This new data frame contains two variables: YEAR and FLT_TOT_1. The resulting data frame can then be used to create visualizations

Reason for choice of graph

The aim of this visualisation is to plot percentage change in the total number of flights per year, hence, a line graph and scatter plot was used. This is because they are effective in showing trends and patterns over time or across a continuous variable. Line graphs are particularly useful when visualizing how a single variable changes over time or as a function of another continuous variable. They are especially effective for showing trends and changes in data that occur gradually, and for identifying the direction and rate of change.

````
```{r,fig.width=8}

# Assign the result of the code to a new variable named eu_flights_1.
   eu_flights_1 <-
   eu_flights %>%

# To group the data by the YEAR variable
   group_by(YEAR) %>%

# To summarise the grouped data by calculating the total number of flights for each year,
and assigns the result to a new variable named total_num_flight. sum() calculates the sum
of the FLT_TOT_1 variable, which contains the total number of flights for each
observation.
   summarise(total_num_flight =
   sum(FLT_TOT_1))

# Initialising a new ggplot object with eu_flights_1 as the data source and the YEAR and
total_num_flight variables as the x     and y aesthetics, respectively
# Calculate percentage change
   eu_flights_1$pct_change <-
   (eu_flights_1$total_num_flight -
   eu_flights_1$total_num_flight[eu_flights_1$YEAR == 2016]) /
   eu_flights_1$total_num_flight[eu_flights_1$YEAR == 2016] * 100

# Plot the line graph
   ggplot(eu_flights_1,
   aes(x = YEAR,
   y = pct_change)) +
   geom_line(color = "blue", size = 1) +
````
````

```
   geom_text(aes(label = sprintf("%+.2f%%", pct_change), color = ifelse(pct_change < 0,
"green", "red")), hjust = 1, vjust = 1,    size = 3) +
   geom_point(size = 3, color = "darkred") +

# Set plot title
   labs(title = "Percentage Change in Total Flights",
    x = "Year",
    y = "Percentage Change") +

# Set theme of the plot
    theme_minimal()+
    theme(legend.position = "none")
```

Graph Analysis

Between 2016 and 2019, the total number of flights in Europe experienced a steady increase of 12.65% due to factors such as the growth of tourism, business activities, and improved ease of movement across European countries. However, the COVID-19 pandemic caused a significant decline in flight movements due to movement restrictions put in place by governments, resulting in a sharp reduction of flights by 58% in 2020. In 2021, as some of these restrictions were eased, there was a gradual increase in flight numbers by 39%. Unfortunately, this trend was short-lived as the early months of 2022 saw a sudden drop in the number of flights, potentially due to changes in the economy such as recessions caused by the Ukrainian war or shifts in consumer spending, impacting demand for air travel and the number of flights operated by airlines. This led to a significant decrease of 66.40% in flight movement in the same year. Despite these challenges, the aviation industry is adapting to the changing circumstances and implementing measures to ensure the safety of passengers and staff. It is hoped that as the situation improves, the industry will continue to recover and eventually return to pre-pandemic levels of flight movements.

11.2 To check for the total number of flight movement per year using bar chart

Description

This creates a bar plot of the FLT_TOT_1 variable for each year in the eu_flights data frame. Each bar represents the total number of flights for a given year, and the height of each bar corresponds directly to the value of the FLT_TOT_1 variable. This plot can be used to compare the total number of flights across different years.

Reason for choice of graph

The goal of this visualisation is to plot categorical variables, therefore a bar chart is used because geom_bar() is a good choice for plotting categorical variables creating a visual representation of a count or frequency distribution.

```{r}
```

```
# Creates a ggplot object with the eu_flights data, mapping the x and y variables to the
YEAR and FLT_TOT_1 columns
   ggplot(eu_flights,
   aes(x = YEAR, y = FLT_TOT_1 / 1000)) +

# Add bar layer to the ggplot
   geom_bar(stat = "identity",
   fill = "#69b3a2") +

# Set the plot title and axis labels
   labs(title = "Total number of flight movements in the EU by year",
   x = "Year",
   y = "Total number of flights") +
   theme(plot.title =
   element_text(hjust = 0.5),
   panel.grid.major =
   element_blank(),
   panel.grid.minor =
   element_blank(),
   panel.border =
   element_rect(color = "black",
   fill = NA,
   size = 1))
```

Graph Analysis

The bar chart in the visualization represents the total number of flight movements across all airports in Europe for the years 2016 to 2022, based on data from the Network Manager. The chart indicates that 2019 had the highest number of flight movements, with a value close to 11 million. This was followed by 2018, and 2017, which had just over 10.7 million flights. In contrast, 2016 had the fourth highest number of flight movements. However, the years 2020, 2021, and 2022 had significantly fewer flight movements compared to the previous years. This could be attributed to the Russian-Ukrainian war which had a profound impact on the aviation industry, leading to flight cancellations, and reduced passenger demand. The decline in flight movements during these years is visible in the bar chart, with the values for each year dropping significantly compared to the earlier years. Overall, the bar chart provides a clear and concise representation of the trends in flight movements across Europe over the years, highlighting the significant impact of the COVID-19 pandemic on the aviation industry.

12.0 VISUALISING THE EUROPEAN FLIGHT DATASET BY MONTHLY MOVEMENT

The second research query pertains to whether there has been a noteworthy alteration in the number of monthly flight arrivals and departures at airports across various European nations throughout the pandemic period. : This question is addressed by the analysis of the plot that follows.

12.1 To check the trends of European flight movements during COVID-19 using line graph and scatter plot

Description - The aim is to generate a clear and visually appealing plot that effectively communicates the trends in monthly flight movement across Europe. The use of color and point markers makes it easy to distinguish between the line and data points, and the x-axis labels provide a clear indication of the time frame being represented.

Reason for choice of Graph The purpose of this visualization is to display the relationships and patterns between two variables: FLT_TOT_1 and MONTH_MON. The code employs a line graph to showcase the trend in the total number of flight departures and arrivals per month during COVID-19 period, depicting the gradual increase or decrease in flight volume over time. Additionally, a scatter plot is used to plot individual data points for monthly flight arrivals.

```
```{r, fig.width=8}

# Creates a new data frame
  covid_2020_period <- eu_flights %>%
  filter(YEAR == 2020)

# Creates a character vector with the three-letter abbreviations for the months of the
year in order from January to December
   month_order <- c("JAN","FEB","MAR","APR","MAY","JUN","JUL","AUG","SEP", "OCT", "NOV",
"DEC" )

# Creates a new data frame (total_movement_2020)
   total_movement_2020 <- covid_2020_period %>%
   group_by(MONTH_MON) %>%
   summarise(total_movement = sum(FLT_TOT_1)) %>%
   mutate(time_period = 'During COVID-19',MONTH_MON = factor(MONTH_MON, levels =
month_order )) %>%
   arrange(MONTH_MON)


# Plot the monthly arrivals with a legend
   ggplot(total_movement_2020,
   aes(x= MONTH_MON,
   y=total_movement, color = time_period, group = time_period)) +


# Add a line to the plot
   geom_line(aes(linetype = "time_period"),
   size=1.5) +

# Create scatter plot of points
   geom_point(aes(color="point"),
   size=3) +

# Set labels for the x-axis, y-axis, title and color legend of the plot.
   labs(x="Month",
   y="Total flight movement",
   title="Monthly flight movements across Europe during COVID-19", color="") +


# Set the colors of the legend
   scale_color_manual(values=c("deepskyblue",
   "orange")) +

# Set the appearance of the plot to a minimalistic theme
   theme_minimal() +

# Set the overall theme of the plot and add a legend to it
    theme(panel.border =
    element_rect(color = "black",
    fill = NA, size = 1),
    legend.position = "bottom",
    panel.grid.major =
    element_blank(),
    panel.grid.minor =
```
```

```
    element_blank()) +

# Customize the color legend in the plot and sets the title of the legend
    guides(color=
    guide_legend(title="Legend",
    ncol=2))

```

Graph Analysis

The data presented in the statement reveals some interesting insights about flight movements in Europe. The fact that August has the highest number of total flight movements and April has the lowest, followed by May, could be because of variety of factors. For instance, August is the peak summer holiday season in Europe, and many people choose to travel abroad during this time. Conversely, April and May may see a dip in flight movements due to the end of the winter holiday season and the start of spring, which is not a peak travel season. The impact of the COVID-19 pandemic on flight movements is also evident from the statement. While there was a slight decrease in total flight movements from January to February, there was a sharp increase from February to March, likely due to the early stages of the pandemic and the need for people to travel back to their home countries. The pandemic has had a profound effect on the airline industry, with many airlines struggling to stay afloat because of reduced demand for air travel and restrictions on international travel. The line graph presented in the statement also shows a consistent pattern in total flight movements from April to August, indicating that this period is generally the busiest time for air travel in Europe. From September to November, there was a gradual decrease in the total number of flight movements, with a slight increase in December, which could be due to the Christmas and New Year holiday season.

In summary, the data presented in the statement highlights the seasonal nature of air travel in Europe and the impact of external factors such as the COVID-19 pandemic on flight movements. Understanding these trends and patterns is important for airlines, airport authorities, and policymakers to make informed decisions about flight scheduling, capacity planning, and travel policies.

12.2 Checking the European flights departures per month during COVID-19

Description - This code filters and summarises the dataset to include only flights during the COVID period, and creates a bar chart to compare the total flights during COVID-19 for each departure monthly in the EU, with modified labels, a black and white theme, and a blank grid and black border.

Reason for choice of graph - The aim of this visualization is to show the total number of flight departure during the COVID-19 period in the EU, with a focus on each month of the COVID-19 period, hence a bar chart is used to compare the total flight movements for each month.

```{r,fig.width=8}

# Convert the "FLIGHT DATE" column to date format
    eu_flights$FLT_DATE <-
    as.Date(eu_flights$FLT_DATE)

# Define the periods COVID start date and COVID end date
    covid_start_date <-
    as.Date('2020-03-01')
    covid_end_date <-
    as.Date('2021-02-28')
```

```
# Filter the data to include only flights during COVID period
   during_covid_dep <-
   eu_flights %>%
   filter(FLT_DATE >=
   covid_start_date & FLT_DATE <=
   covid_end_date) %>%
   group_by(MONTH_MON) %>%
   summarise(total_departures =
   sum(FLT_DEP_1)) %>%
   mutate(MONTH_MON =
   factor(MONTH_MON, levels =
   c('JAN', 'FEB', 'MAR', 'APR', 'MAY', 'JUN', 'JUL', 'AUG', 'SEP', 'OCT', 'NOV', 'DEC')))

# Create a bar chart to plot the total flight during COVID-19 for each departure monthly
     ggplot(during_covid_dep,
     aes(x = MONTH_MON,
     y = total_departures,
     fill = MONTH_MON)) +

# Create a bar chart where the height of each bar is equal to the value of the data point,
and the bars are semi-transparent with an alpha value of 0.7
     geom_bar(stat = "identity",
     alpha = 0.7) +

# Modify the labels of the plot created using the ggplot2 package
      labs(x = 'Month',
      y = 'Flight departures',
      title = 'Monthly flight departures during COVID-19 in the EU') +

# Apply a pre-defined black and white theme to the plot
      theme_bw() +
      geom_text(aes(label = total_departures),
      position = position_stack(0.5),
      color = "white", angle = 90) +

# Set the grid lines of the plot to be blank and the panel border to be black and non-
filled
      theme(panel.grid =
      element_blank(),
      panel.border =
      element_rect(colour =
      "black",
      fill=NA,
      size=1),

# Rotate text angle and height
      axis.text.x = element_text(angle=45,
      vjust=0.5))


```

Graph Analysis - The rise in flight activity during the summer months, particularly in August, is a result of the high number of people taking vacations during this time. As the weather gets warmer, individuals often take time off from work or school to explore new destinations. Airlines therefore, respond to this demand by offering more flights and destinations during the summer months. Conversely, April records the lowest number of flight activity accounting for total flight departure of 70657 because of COVID-19 related restrictions implemented by some European countries. However, there is a slight decline in flight activity during the months of September to November from 353047 to 216676, which could be attributed to various factors. Firstly, this period is an off-peak season, with fewer individuals traveling due to the resumption of academic sessions in schools and universities. Additionally, the fall and winter seasons are not conducive to travel, particularly for leisure purposes. As a result, airlines often reduce their flight schedules during this period to cut costs and maintain profitability. Interestingly, despite the slight dip in flight activity during the fall and winter months, there is a slight increase in December totaling 230739 of flight departure. This can be attributed to the holiday season, with many individuals traveling to be with loved ones during Christmas and New Year's Eve.

Overall, the bar chart highlights a consistent trend in flight activity in Europe, heavily influenced by seasonal factors. Peak travel periods coincide with the summer months and holiday seasons, while off-peak periods occur during the fall and winter seasons

12.3 Checking the European flights arrivals monthly during COVID-19

Description

The purpose of this code is to extract and condense *eu_flights dataset*, focusing on flights during the COVID period, and generate a bar chart comparing the total flights during COVID-19 for EU monthly arrivals.

Reason for choice of graph

The purpose of this visualization is to showcase the total count of flight arrivals in the European Union during the COVID-19 period, with a particular focus on individual months within that timeframe. A bar chart will be utilized to depict the data and enable comparison of the total flights per month during COVID-19. The height of each bar will indicate the overall number of flight arrivals, while the use of color-coding will help differentiate the months from each other.

```{r,fig.width=8}

# This line of code converts the FLT_DATE column in the eu_flights dataset from character format to Date format
   eu_flights$FLT_DATE <-
   as.Date(eu_flights$FLT_DATE)

# Define the periods before and after COVID
   covid_start_date <-
   as.Date('2020-03-01')
   covid_end_date <-
   as.Date('2021-02-28')

# Filter the data to include only flights during COVID period
   during_covid_arr <-
   eu_flights %>%
   filter(FLT_DATE >=
   covid_start_date & FLT_DATE <=
   covid_end_date) %>%
   group_by(MONTH_MON) %>%
   summarise(total_arrivals =
   sum(FLT_ARR_1)) %>%
   mutate(MONTH_MON =
   factor(MONTH_MON,
   levels = c('JAN', 'FEB', 'MAR', 'APR', 'MAY', 'JUN', 'JUL', 'AUG', 'SEP', 'OCT', 'NOV',
'DEC')))

# Create a bar chart to plot the total flights during COVID-19 for monthly arrivals
   ggplot(during_covid_arr,
   aes(x = MONTH_MON,
   y = total_arrivals, fill = MONTH_MON)) +

# Create a bar chart where the height of each bar is equal to the value of the data point,
```

```
and the bars are semi-transparent with an alpha value of 0.7
   geom_bar(stat = "identity",
   alpha = 0.7) +

# Modify the labels of the plot created using the ggplot2 package
   labs(x = 'Month',
   y = 'Flight arrivals',
   title = 'Monthly flight arrivals during COVID-19 in the EU') +

# Apply a pre-defined black and white theme to the plot
   theme_bw() +

# This line of code adds text labels to each bar in the plot, with the label text set to
the total_arrivals variable
   geom_text(aes(label = total_arrivals),
   position =
   position_stack(0.5),
   color = "white",
   angle = 90) +

# Set the grid lines of the plot to be blank and the panel border to be black and non-
filled
   theme(panel.grid =
   element_blank(),
   panel.border =
   element_rect(colour = "black",
   fill=NA,
   size=1),
   axis.text.x = element_text(angle=45,
   vjust=0.5))
```

Graph Analysis

Flight activities can be lower in January and februARY, especially after the holiday season, as people return to work and school. Airlines may reduce their flight schedules during this time. However, there was a drastic increase in the number of Flight arrivals as spring break approaches, with more people traveling for vacations or to visit family and friends. Airlines may add more flights to popular destinations and offer discounts to attract travelers. April typically has the lowest flight activity due to COVID-19 restrictions dropping from 381473 in March to 70856 in April. Although there is a steady increase in Flight movements from May to August as summer approaches, with more people traveling for vacations and holidays such as Memorial Day weekend. Airlines may add more flights and routes to popular destinations to meet the demand. These months are typically the busiest for air travel, with the summer vacation season in full swing. Airlines add more flights and routes to popular destinations and may charge higher fares due to high demand. Although there was a gradual drop from September to November mainly because of resumption of academic sessions and the end of summer vacation season. Airlines may reduce their flight schedules during this time. Flight activity increased slightly to 231088 in December due to the holiday season, with many people traveling to be with family and friends. Airlines respond by offering more flights and routes to popular holiday destinations.

Overall, flight activity varies throughout the year due to seasonal and holiday patterns, weather conditions, and other factors. Airlines adjust their flight schedules and fares to meet the demand, and travelers can take advantage of these changes to plan their trips accordingly. Additionally, the bar chart reveals a pattern of flight activity in Europe that is strongly influenced by seasonal factors. The high travel seasons align with the summer months and holiday periods, while the low travel seasons occur during the peak of COVID-19 movement restrictions.

13.0 VISUALISING THE EUROPEAN FLIGHT DATASET BY AIRPORT

The third research question examines the impact of the COVID-19 pandemic on the number of flights at the busiest airports in Europe, comparing the pre-pandemic and post-pandemic periods. : The following plot analysis answers this question

13.1 Confirming the 4 busiest airport by departure across Europe before and after COVID-19 using bar chart

Description

The plot aims to extract and summarize *eu_flights dataset* that highlights the busiest airports in Europe based on flight departures, both pre and post COVID-19. It then generates a bar chart that compares the total number of flight departures across these busy airports, before and after the pandemic.

Reason for choice of graph

The purpose of this visualization is to present the total number of flight departures at the busiest airports in the European Union, both before and after the COVID-19 pandemic. A bar chart is an effective way to compare data across different categories or groups, making it a suitable choice for visualizing the frequency of flight departures from different locations. The bar graph is utilized to compare the overall number of departures for each airport, with the height of each bar representing the number of flights departing from major EU airports. The time periods before and after the pandemic are color-coded to distinguish them from one another.

```{r,fig.width=8}
# Convert the "FLT_DATE" column to date format
    eu_flights$FLT_DATE <-
    as.Date(eu_flights$FLT_DATE)

# Define the periods before and after COVID
    before_covid_start <-
    as.Date('2019-01-01')
    before_covid_end <-
    as.Date('2020-02-29')
    after_covid_start <-
    as.Date('2020-03-01')
    after_covid_end <-
    as.Date('2021-02-28')

# Calculate the total departures before and after COVID-19 for each airport
# Departure before COVID-19
    before_covid_departures <-
    eu_flights %>%
    filter(FLT_DATE >=
    before_covid_start & FLT_DATE <=
    before_covid_end) %>%
    group_by(APT_NAME) %>%
    summarise(total_departures =
    sum(FLT_DEP_1)) %>%
    rename(before_covid_departures =
    total_departures) %>%
    arrange(desc(before_covid_departures)) %>%


# Filter the top 4 airports before COVID-19
    head(4)

# Departure after COVID-19
```

```
    after_covid_departures <-
    eu_flights %>%
    filter(FLT_DATE >=
    after_covid_start & FLT_DATE <=
    after_covid_end) %>%
    group_by(APT_NAME) %>%
    summarise(total_departures =
    sum(FLT_DEP_1)) %>%
    rename(after_covid_departures =
    total_departures) %>%
    arrange(desc(after_covid_departures)) %>%

# Filter the top 4 airports after COVID-19
    head(4)

# Merge the two data frames to create a single data frame for plotting
    combined_airport_departures1 <-
    merge(before_covid_departures,
    after_covid_departures,
    by='APT_NAME', all=TRUE)
    combined_airport_departures1[is.na(combined_airport_departures1)] <- 0

#  Initialize ggplot object with data frame and reorder the airport names based on the
number of departures after COVID-19,       setting the maximum value as the reference
point
    ggplot(combined_airport_departures1,
    aes(x=reorder(APT_NAME,
    after_covid_departures, max))) +


# Add a column chart for the number of departures before COVID-19, using a fill color and
adjusting alpha and position
    geom_col(aes(y=before_covid_departures,
    fill='Before COVID-19'),
    alpha=0.7,
    position=position_dodge(width=0.9)) +

#  Add a column chart for the number of departures after COVID-19, using a fill color and
adjusting alpha and position
    geom_col(aes(y=after_covid_departures,
    fill='After COVID-19'),
    alpha=0.7,
    position=
    position_dodge(width=0.9)) +

# Add axis labels, chart title, and legend label
    labs(x='Airport',
    y='Total departures',
    title='Top 4 airports with highest departures before and after COVID-19',
    fill='Period') +

# Use a black and white theme for the chart
    theme_bw() +

# Add theme
    theme(axis.text.x = element_text(angle=45,
    vjust=0.5),
    panel.grid.major =
    element_blank(),
    panel.grid.minor =
    element_blank()) +
    scale_fill_manual(values=c('palegreen1', 'paleturquoise4'),
    labels=c('After COVID-19', 'Before COVID-19'))
```

````
```
````

Graph Analysis

The COVID-19 pandemic had a significant impact on the aviation industry, leading to a sharp decrease in air travel demand and resulting in reduced flight activity at airports around the world. Among Europe's busiest airports, Frankfurt was the top performer in terms of flight departures prior to the pandemic, with nearly 290,000 departures recorded. This is not surprising, considering Frankfurt is the hub for one of Europe's largest airlines, Lufthansa, and serves as an important connecting point for air travel within Europe and to other regions of the world. Amsterdam-Schiphol ranked second in terms of flight departures, closely following Frankfurt's lead. Amsterdam-Schiphol is a major airport in Europe and serves as the hub for KLM Royal Dutch Airlines. Like Frankfurt, Amsterdam-Schiphol connects travelers to numerous destinations within Europe and beyond. Paris-Charles-de-Gaulle airport ranked third in terms of flight departures, with slightly fewer departures than Amsterdam-Schiphol. Charles-de-Gaulle is the primary international airport serving Paris and is a hub for Air France, the country's flagship carrier. London-Heathrow had the fewest flight departures among Europe's four busiest airports prior to the pandemic. However, Heathrow is still one of the busiest airports in the world, serving as the primary hub for British Airways and connecting passengers to destinations across the globe. In the aftermath of the pandemic, all four airports saw reduced flight activity. London-Heathrow recorded the lowest number of departures, with many international travel restrictions still in place in the UK. Meanwhile, Amsterdam-Schiphol recorded the highest number of departures among the four airports, as the Netherlands gradually reopened for international travel. Frankfurt and Paris-Charles-de-Gaulle had similar levels of flight departures following the pandemic, as travel restrictions eased and air travel demand slowly began to recover.

13.2 Confirming the 4 busiest airport by arrivals across Europe before and after covid-19 using bar chart

Description
The objective of this plot is to extract and condense *eu_flights dataset* that emphasizes the busiest airports in Europe based on flight arrivals, both prior to and following the COVID-19 pandemic. Subsequently, a bar chart is produced to compare the total number of flight arrivals across these airports before and after the pandemic.

Reason for choice of graph

The goal of this visualization is to display the total number of flight departures from the busiest airports in the European Union, both before and after the COVID-19 pandemic. A bar chart is an appropriate method to compare data among various categories or groups, making it a suitable option for visualizing the frequency of flight departures from different locations. The bar chart is utilized to compare the total number of arrivals for each airport, with the height of each bar indicating the number of flights departing from major EU airports. The periods before and after the pandemic are differentiated by color-coding.

````
```{r,fig.width=8}


# Convert the "FLT_DATE" column to date format
  eu_flights$FLT_DATE <-
  as.Date(eu_flights$FLT_DATE)

# Define the periods before and after COVID
  before_covid_start <-
  as.Date('2019-01-01')
  before_covid_end <-
````

```
  as.Date('2020-02-29')
  after_covid_start <-
  as.Date('2020-03-01')
  after_covid_end <-
  as.Date('2021-02-28')

# Calculate the total arrivals before and after COVID for each airport
# Total arrivals after COVID-19
  before_covid_arrivals <-
  eu_flights %>%
  filter(FLT_DATE >=
  before_covid_start & FLT_DATE <=
  before_covid_end) %>%
  group_by(APT_NAME) %>%
  summarise(total_arrivals =
  sum(FLT_ARR_1)) %>%
  rename(before_covid_arrivals =
  total_arrivals) %>%
  arrange(desc(before_covid_arrivals)) %>%
  head(4)

# Total arrivals after COVID-19
  after_covid_arrivals <-
  eu_flights %>%
  filter(FLT_DATE >=
  after_covid_start & FLT_DATE <=
  after_covid_end) %>%
  group_by(APT_NAME) %>%
  summarise(total_arrivals =
  sum(FLT_ARR_1)) %>%
  rename(after_covid_arrivals =
  total_arrivals) %>%
  arrange(desc(after_covid_arrivals)) %>%
  head(4)


# Merge the two data frames to create a single data frame for plotting
  combined_airport_arrivals <-
  merge(before_covid_arrivals,
  after_covid_arrivals,
  by='APT_NAME', all=TRUE)
  combined_airport_arrivals[is.na(combined_airport_arrivals)] <- 0

# Create a grouped bar chart using ggplot2 to compare the total arrivals before and after
COVID-19 for the top 4 airportss
  ggplot(combined_airport_arrivals,
  aes(x=reorder(APT_NAME,
  after_covid_arrivals, max))) +

# Use geom_bar() to create bars before COVID-19 period, and customize them with color fill
and alpha value
  geom_bar(aes(y=before_covid_arrivals,
  fill='Before COVID-19'),
  alpha=0.7,
  position=position_dodge(width=0.9),
  stat="identity") +

# Use geom_bar() to create bars before COVID-19 period, and customize them with color fill
and alpha value
  geom_bar(aes(y=after_covid_arrivals,
  fill='After COVID-19'),
  alpha=0.7,
  position=
  position_dodge(width=0.9),
  stat="identity") +
```

```
# Customize the x and y-axis labels, as well as the chart title and legend
  labs(x='Airport',
  y='Total arrivals',
  title='Top 4 airports with highest arrivals before and after COVID-19', fill='Period') +

# Set plot theme and customize x-axis text
  theme_bw() +
  theme(axis.text.x =
  element_text(angle=45,
  vjust=0.5),

# Remove the major and minor grid lines of the plot to improve its clarity.
  panel.grid.major =
  element_blank(),
  panel.grid.minor =
  element_blank()) +

# Customize the colors and labels for the fill scale of the plot
  scale_fill_manual(values=c('#8dd3c7', 'palevioletred4'),
  labels=c('After COVID-19',
  'Before COVID-19'))
```

Graph Analysis

The COVID-19 pandemic had a profound impact on the aviation industry, leading to a decrease in air travel and resulting in the closure of many airports across the world. In Europe, the impact of the pandemic was felt heavily by the busiest airports, which saw a significant decrease in the number of flights arriving and departing. Before the pandemic, the four busiest airports in the European Union, namely Frankfurt, Amsterdam-Schiphol, Paris-Charles-de-Gaulle, and London-Heathrow, experienced an increase in the number of flight arrivals. Frankfurt and Amsterdam-Schiphol were the busiest airports, recording just below 300,000 arrivals, while Paris-Charles-de-Gaulle saw a slight decrease in arrivals compared to the other two airports. London-Heathrow had the fewest arrivals among the four airports.

However, after the pandemic, there was a sharp decline in the number of flights arriving at the four busiest airports in Europe. This was mainly due to the restrictions on international travel and the reduction in economic activities and tourism as a result of the pandemic. Amsterdam-Schiphol remained the busiest airport among the four, recording the highest number of arrivals, followed by Paris-Charles-de-Gaulle and Frankfurt, which both experienced a slight decrease in arrivals compared to Amsterdam-Schiphol. London-Heathrow continued to have the lowest number of total flight activity among the busiest European airports.

13.3 To check the trends of 4 busiest airports in Europe by flight departure from 2016 to 2022

Description

The aim of this visualization is to identify the top four busiest airports in Europe in terms of flight departures between 2016 and 2022. It then presents a line graph that depicts the trends in the number of flight departures across these selected airports.

Reason for choice of graph

This visual representation aims to demonstrate the trend of flight arrivals from the busiest airports in the European Union between 2016 and 2022. A scatter plot is typically utilized to display the relationship between two continuous variables, which is useful for visualizing the distribution of data points and identifying patterns and for displaying the total number of flight arrivals from different airports during this period, a line graph is used to illustrate how data changes over time. The line graph emphasizes the trends in flight arrivals from the top 4 busiest airports in Europe, with each major EU airport being represented by a unique color.

```{r}

# Filter for top 4 airports based on total departure flights
  top_countries_ab <-
  eu_flights %>%
  group_by(APT_NAME) %>%
  summarise(total_departure_flights =
            sum(FLT_DEP_1)) %>%
  arrange(desc(total_departure_flights)) %>%
  head(4) %>%
  pull(APT_NAME)

# Filter data for top 5 airports
  filtered_data_ab <-
  eu_flights %>%
  filter(APT_NAME %in% top_countries_ab)

# Aggregate total departure flights per year
  filtered_data_ab <-
  filtered_data_ab %>%
  group_by(YEAR, APT_NAME) %>%
  summarise(total_departure_flights =
  sum(FLT_DEP_1)/1000000)

# Create line graph
  ggplot(filtered_data_ab,
  aes(x = YEAR,
  y = total_departure_flights,
  group = APT_NAME,
  color = APT_NAME)) +
  geom_line(size = 1.2) +

# Create scatter plot
  geom_point(size = 2) +
  labs(title = "4 busiest airports by departure in Europe",
  x = "Year",
  y = "Total Departure Flights") +
  theme_minimal() +
  theme(legend.position = "right")


```

Graph Analysis

Prior to the COVID-19 pandemic, the busiest airports in Europe experienced a steady increase in flight departures from 2016 to 2019, which was attributed to the growth of economic activities across European countries. Frankfurt emerged as the busiest airport during this period, with over 250,000 flight departures in 2019, closely followed by Amsterdam-Schiphol with a slightly smaller number of flight departures. Paris-Charles-de-Gaulle was the third busiest airport, while London-Heathrow remained the least patronized among the top four busiest airports in Europe, with a consistent trend throughout the period.

However, the COVID-19 pandemic had a significant impact on air travel and led to a drastic decline in flight departures from all major European airports in 2020 due to the various restrictions imposed by European countries. There was a slight recovery in 2021 as countries began to open up for movement, resulting in an increase in the number of flight departures from these airports. However, the early part of 2022 saw a sharp decrease in the number of flight departures from the busiest airports in Europe, with London-Heathrow being the most affected, while Amsterdam-Schiphol remained the busiest airport with the highest number of flight departures in 2022.

13.4 To check the trends of 4 busiest airports in Europe by flight arrival from 2016 to 2022

Description

This visualization depicts the trend in flight departures from the busiest airports in Europe from 2016 to 2022, utilizing the *eu_flights dataset*. A line graph is employed to illustrate the change in the frequency of flight departures across these airports over the given period.

Reason for choice of graph

The goal of this visualization is to display the trend in flight departures from the busiest airports in the European Union over a 6 year period, from 2016 to 2022. A line graph is an appropriate method to compare data among various categories or groups, making it a suitable option for visualizing the frequency of flight departures from different locations. The line graph is utilized to show the patter followed by the 5 most patronised airports in Europe by. with the height of each bar indicating the number of flights departing from major EU airports. Each european airport is differentiated by color-coding.

```{r}

# Filter for top 4 countries based on total departure flights
  top_countries1 <-
  eu_flights %>%
  group_by(APT_NAME) %>%
  summarise(total_arrival_flights =
  sum(FLT_ARR_1)) %>%
  arrange(desc(total_arrival_flights)) %>%
  head(4) %>%
  pull(APT_NAME)

# Filter data for top 5 airports
  filtered_data1 <-
  eu_flights %>%
  filter(APT_NAME %in% top_countries1)

# Aggregate total departure flights per year
  filtered_data1 <-
  filtered_data1 %>%
  group_by(YEAR, APT_NAME) %>%
  summarise(total_arrival_flights =
  sum(FLT_ARR_1)/1000000)

# Create line graph
  ggplot(filtered_data1,
  aes(x = YEAR,
  y = total_arrival_flights,
  group = APT_NAME,
  color = APT_NAME)) +
  geom_point(size = 3) +
  geom_line(size = 1.2) +
  labs(title = "4 busiest airports by Arrival in Europe",
  x = "Year",
  y = "Total Arrival Flights") +
  theme_minimal() +
  theme(legend.position = "right")
```

```

*Graph Analysis*

The information provided describes the trend in flight arrivals at the busiest airports in the European Union over a 6-year period from 2016 to 2022. Between 2016 and 2019, the airports experienced a gradual growth in the number of flight arrivals due to increased economic activities across European countries. Frankfurt had the highest peak in flight arrivals, accounting for over 250,000 flight departures in 2019, followed by Amsterdam-Schiphol with a small margin when compared to Frankfurt. Paris-Charles-de-Gaulle was the third busiest airport in the same year while London-Heathrow was the least patronized airport among the top 4 busiest airports in Europe. However, in 2020, these airports saw a significant decrease in the number of flight arrivals due to COVID-19 restrictions imposed by some European countries to curtail the spread of the pandemic. There was a little recovery in 2021 as countries began to open up for business activities and tourism, resulting in an increase in the number of flight arrivals at these most patronized airports in Europe. In 2022, there was a sharp fall in the number of flights that arrived at the airports, with London-Heathrow being the most affected in terms of recorded flight arrivals, while Amsterdam-Schiphol airport still recorded the highest number of flight arrivals in 2022. This suggests that the COVID-19 pandemic had a significant impact on the aviation industry in Europe, with many airlines reducing their flight schedules or completely suspending operations.

# 14.0 CONCLUSION

In recent years, air travel has become increasingly affordable and accessible, resulting in a surge of passengers globally, including Europe. However, the outbreak of COVID-19 and the need to prevent its spread led to various restrictive measures being implemented by countries worldwide since the start of 2020, impacting the air transport industry. Nevertheless, recent statistics on commercial flights indicate a recovery.

Eurocontrol and Ec.europa.eu record of commercial flights in Europe has yielded valuable insights into the aviation industry's dynamics and the impact of the COVID-19 pandemic on air travel. As the industry adapts and recovers in the post-pandemic world, this dataset will continue to be an essential resource for policymakers, researchers, and industry stakeholders. Predictive models that factor in seasonal flights and pandemic periods can also help airlines and airports optimize their resources and schedules for more cost-effective and efficient operations and by analyzing the European flight dataset, researchers can uncover valuable patterns and insights that shed light on the aviation industry.

Overall, Future research on the European flight dataset should focus on analyzing the long-term effects of the COVID-19 pandemic on the aviation industry, including changes in airline and passenger behavior, airport infrastructure, and government policies. Researchers can also explore the potential impact of emerging technologies, such as electric and autonomous aircraft, on the industry. In addition, there is a need for research on sustainable aviation, including the development of alternative fuels and carbon capture technologies, as the industry seeks to reduce its environmental footprint. Finally, there is a need for continued data collection and analysis to monitor trends in the European aviation industry and inform policy decisions.

# 15.0 REFERENCE

https://ansperformance.eu/data/

https://ec.europa.eu/eurostat/web/products-eurostat-news/-/ddn-20210914-1

Suzumura, T., Kanezashi, H., Dholakia, M., Ishii, E., Napagao, S.A., Perez-Arnal, R. and Garcia-Gasulla, D. (Dec 10, 2020) The Impact of COVID-19 on Flight Networks. IEEE, pp. 2443.

Burghouwt, G. and Hakfoort, J. (2001) The evolution of the European aviation network,
```

16.0 APPENDIX