

TIL6022 Project Group 7

Members: Viresh Budhram, Iris den Ouden, Jelle Top, Ivaylo Iliev, Finn van Benthem

Student numbers: 5426359, 5879671, 5656982, 6294340, 6239730

1. Research Objectives and Research Questions:

Covid-19 had a major impact on the worldwide society. The air travel segment was among the worst hit sectors during this period. Hence, we are interested in how bad Covid-19 has impacted this sector, and how well the sector recovered once countries and regions welcomed international travellers again. Here, we decided to focus on the Dutch airport Schiphol. We aim to compare air travel passenger numbers in the period (2015 – 2019) before Covid-19 with the period after (2021 – 2024). We will look at load factor statistics, possible changes in air travel behaviour, and changes in destination popularity. Also, we wish to check if we could predict future passenger flows for Schiphol for the forthcoming years. Thus, the next research and sub research objectives could be formulated:

Main objective: How have air travel patterns, including passenger volumes and destination preferences, evolved for travelers to and from Dutch airports between the periods 2015-2019 and 2021-2024?

Sub objectives :

1. What's the difference between intercontinental and international flights in terms of number of passengers and number of flights?
2. What is the load factor (number of passenger per flight) and how it changed over time?
3. What are the changes in air travel behaviour of passengers per month before covid and after covid in terms of preferred destinations both for short- and long-haul destinations?
4. What are the forecasted air travel passengers in the years 2024 - 2030?
5. Which continent region did Dutch people chose as their favorite for their vacation abroad per year? Has covid affected it?"

2. Dataset & Motivation

The dataset provides information on flight movements, passengers, cargo, and mail at Dutch airports. The available data is presented on a yearly basis from 1997 onwards and on a monthly basis from 1999 to 2024. For this assignment, we will focus on the data related to flight movements and passengers at Dutch airport Schiphol. For flight

movements, the dataset shows differences between commercial air traffic flights and local flights. The dataset includes all the data for the period before COVID-19 (2015-2019) and after COVID-19 (2021-2024) that we need to process for the assignment.

The dataset we will use for our project is dataset_full.csv This dataset is extracted from: https://opendata.cbs.nl/statline/portal.html?_la=en&_catalog=CBS&tableId=37478eng&_theme=1185

3. Data Pipeline [Methodology]

Methodologies that will be used to answer the sub and main research questions

Every datapipeline consists of the three basic components:

1. Data sources: What does the data consists of? Which data do we need?
2. Transformation and processing: transformation can exists out of operations sucg as grouping, filtering, and sorting . In this step, it is important that the data is cleansed and prepared.
3. Destination: in this part, the data will be used for analysis and will be made more visual.

Step 1: Data sources

For this analysis, a dataset will be used which consists of the following information: - Airport ID number, periods - Flights (total, arrivals, departures) - Passengers (total, arrivals, departures, scheduled, non-scheduled) - Destinations (total, Europe, EU, intercontinental, every part of every continent) - Cargo (total, destinations) - Mail (total, destinations)

Not all information from this dataset will be relevant, the following data will be needed:

- Airport ID number, periods, flight information about passengers.

Step 2: Data Transformation and processing

To make sure the data will be used correctly, it is important to filter, group and sort the data. The following steps will be/are followed (???):

1. Filter out all unneccessary data:
 - 1.1 Data about mail and cargo flights.
 - 1.2 Remove data about local flights. Local flights are not common in the Netherlands as and are not relevant for this project. The local flights that did happen where not meant to transport people from one airport to the other as public transport is way faster for that.

2. Filter data about periods. This projects focusses on the years 2016-2019 and 2021-2024, the useful data during these periods of time will be extracted from the datasheet.
3. Replace Airport ID number with logical name. An ID number does not give direct information about which airport is being analyzed. By replacing the ID number with a logical name, this will be more clear when multiple airports will be analyzed. For example: AO3590 = Schiphol.
4. Split the data column 'periods'. This column is formulated as year-month (2001-mm01 = january 2001), this column is splitted in years and months which creates an extra column. This is done to remove the letters from this data column and to be able to analyze the data by year and to compare the data in time by month.
5. The outcome is an overview of the data related to passenger flights going abroad during the periods 2016-2019 and 2021-2024.

Step 3: Destination

The final step of the data pipeline focuses on visualizing the processed data to uncover trends and insights related to air travel patterns. To answer the main objective—how air travel patterns, a combination of visualizations will be used.

For comparing intercontinental and international flights in terms of passenger numbers and flight volumes, a grouped bar chart will display the data, allowing for a clear comparison across these categories and time periods. The load factor (passengers per flight) will be shown using a line chart, highlighting trends over time for different periods before and after COVID-19.

To visualize the shift in air travel behavior by month, especially for short- and long-haul destinations, a stacked area chart will display how these destination preferences evolved before and after the pandemic. Forecasted passenger volumes for 2024-2030 will be represented by a forecast line chart, with confidence intervals to show the range of possible future values. For Dutch vacation preferences by continent, a pie chart or stacked bar chart will show the proportion of passengers traveling to different regions each year, revealing any shifts caused by COVID-19.

Additional visualizations like heatmaps for seasonal travel patterns or choropleth maps for geographic distribution of passengers by country will complement the main graphs. These visualizations will use Python's `matplotlib`, `seaborn`, and `Plotly` to provide clear, well-labeled insights and make trends in air travel behavior easy to interpret across the periods in focus.