Analysis and Prediction of Air Traffic Volume and Pattern in the Netherlands Post-COVID-19

Members & Student numbers:

Minghao Li 6212999

Xinyu Yang 6302750

Yilin Shi 6140343

Yue Guo 6147275

Yumeng Pan 6134130

# 1. Introduction

The COVID-19 pandemic had profound impacts on global air traffic. In the Netherlands, travel restrictions and behavioral changes led to substantial shifts in air traffic volume. As the country gradually reopened, the recovery process revealed varying patterns across different types of flights and travel groups.

By analyzing historical traffic patterns, this project will explore the variations in recovery across different flight modes, while also visualizing the comparison between actual and predicted values. Understanding these changes and predicting future traffic volume in the face of possible new pandemics or other disruptions is crucial for air traffic management and policy-making.

# 2. Research Objectives

This project aims to analyze these changes by comparing traffic volumes before, during, and after the pandemic, focusing on different types of flights (e.g. domestic, international, passenger, cargo). Additionally, it will forecast the potential impact of future pandemics on air traffic based on historical data, providing insights for aviation management and policy decisions.

# 3. Research Questions (RQs)

## RQ1: What is the impact of the monthly number of new cases before, during, and after the COVID-19 pandemic on different flight types (e.g. cross-country and local flights) in the Netherlands?

## RQ2: What is the impact of the monthly number of new cases before, during, and after the COVID-19 pandemic on the numbers of passengers, cargo, and mail in the Netherlands?

## RQ3: The prediction of future air traffic volume and pattern in response to potential pandemic or disruptions based on existing data.

# 4. Data Pipeline and Methodology

## Data Sources:

### Our World in Data (OWID) COVID-19 Dataset:

<https://ourworldindata.org/coronavirus>

Pandemic-related data including new cases in each month will be used to correlate air traffic pattern variation.

### CBS (Statistics Netherlands) Open Data:

CBS Transport and Mobility Dataset:

<https://opendata.cbs.nl/statline/#/CBS/en/dataset/37478eng/table?ts=1728287180831>

This source is used to research the variation across different flight modes and the numbers of passengers, cargo, and mail through the COVID period.

## 

## Data Collection:

Retrieve the CBS traffic data using the CBS StatLine portal, focusing on air traffic data for different flight types (local flights, cross-country flights, passenger, cargo, etc.).

Extract COVID-19 new cases, death cases and vaccination counts from Our World in Data.

## Data Cleaning:

Handle missing values, smooth outliers in both air traffic and pandemic datasets.

Ensure consistency in time intervals and align air traffic data with pandemic measures.

## Data Merging and Integration:

Merge the datasets by time (e.g., weekly or monthly intervals), enabling analysis of how air traffic volume changes relate to existing pandemic data.

Change the air traffic dataset from a weekly format to a monthly format.

## Visualization:

Plot the existing data and various aviation-related variables over time.

Use line charts to visualize the trends and relationships between the variables.

## Data Analysis:

Using existing historical data, correlation analysis, linear regression, and time series analysis will be conducted, which will illustrate the relationships and trends through visual diagrams, enabling us to estimate future transportation volumes under comparable conditions. By analyzing these relationships, we aim to uncover patterns and dependencies that may inform our understanding of how various factors influence each other in the context of transportation and public health.

### Correlation Analysis:

Correlation analysis will be employed to determine whether a discernible relationship exists between infection cases and air traffic data, serving as the foundation for the subsequent prediction model. We will introduce the epidemic data and integrate it with three additional datasets to analyze the relationship between the dependent variable (Y) and the changing factors. These coefficients indicate the strength and direction of the linear relationship between each variable and existing data.

### Linear Regression:

Linear regression will be employed to establish a clear and interpretable relationship between epidemic data and transportation volume. We will divide the data into training and testing sets, fit a Multiple Linear Regression model, and assess its performance using metrics such as Mean Squared Error (MSE) and R-squared (R²) to visualize the actual versus predicted values for each variable to enhance understanding. To stabilize variance and achieve a more normal distribution, we will also apply a log transformation to the dependent variables. Additionally, random forest models will also be conducted to further explore the relationship.

1. Time Series Analysis:

A comprehensive time series analysis will be conducted on three distinct sets of data, employing Granger causality testing to explore the relationships between variables over time. This method will be conducted on 5 aviation data in air traffic and 3 covid-related data, providing valuable predictive insights into the dynamics of the data.

# 5. Deliverables

## Analytical Report:

A detailed report of the air traffic recovery pattern analysis.

## Visualizations:

Graphs and charts to illustrate air traffic recovery trends and predictive analysis results.