

# Comprehensive Summary of the Data selected

Thomas Bordes - Alexandre Brun

September 22, 2024

## 1 Introduction

In this document, we provide an in-depth overview of the datasets and external factors that we analyzed to understand cyclist safety and traffic patterns in Paris. Our analysis incorporates weather data, road accidents, strikes, the COVID-19 pandemic (including lockdowns, curfews, and holidays), hospital situations, and public holidays. Each dataset is linked to relevant variables.

## 2 Original Data (Kaggle Challenge)

The original dataset provided by the Kaggle challenge focuses on weather-related variables essential for predicting how weather conditions may affect cyclist traffic and accident risks. The key variables used from this dataset include:

- **number\_sta**: Weather observation station identifier.
- **date**: Timestamp of the weather observation.
- **pmer**: Sea-level atmospheric pressure.
- **tend**: Atmospheric pressure tendency over time.
- **cod\_tend**: Type of pressure tendency.
- **dd**: Wind direction in degrees.
- **ff**: Wind speed.
- **t**: Temperature.
- **td**: Dew point temperature.
- **u**: Relative humidity.
- **vv**: Visibility.
- **ww**: Present weather conditions.
- **w1, w2**: Past weather conditions.
- **n**: Cloud cover.
- **lon, lat**: Longitude and latitude of the weather station.

This dataset is crucial because weather conditions significantly impact both cyclist activity and accident likelihood. The original data was sourced directly from the Kaggle competition and contains station-specific weather data.

## 3 External Data Sources

We incorporated several external datasets to account for real-world events, including road accidents, strikes, hospital situations, COVID-19-related restrictions, and public holidays. Each dataset adds important variables to create a more complete understanding of cyclist safety in Paris.

### 3.1 Road Accidents Data (Kaggle)

The road accidents dataset provides detailed information about accidents involving cyclists, which is critical for understanding cyclist safety.

- **Max\_Grav\_accidents**: Maximum severity of cyclist accidents, categorized into:
  1. Uninjured
  2. Slightly injured
  3. Hospitalized
  4. Fatality
- **Count\_accidents**: The total number of cyclist accidents per hour in Paris.

**Source**: This dataset is publicly available on Kaggle under the title "Accidents in Paris." It includes detailed records from the official Etalab database on road traffic accidents. The dataset can be accessed at <https://www.kaggle.com/paris-road-accidents>.

### 3.2 Strikes Data

Paris is frequently impacted by strikes, which influence transportation patterns and consequently cyclist traffic. Strikes reduce public transport availability, leading to a spike in bicycle usage.

- **Strike\_date**: The date of each major strike event.
- **Affected\_transport**: The specific modes of public transportation impacted by the strike (e.g., metro, buses).
- **Cyclist\_increase**: The percentage increase in cyclist traffic during the strike.

**Source**: The strikes dataset was compiled from public records and transportation data published by the Paris transportation authorities. Official strike announcements were used to gather this data.

### 3.3 COVID-19 Data: Lockdowns, Curfews, and Holidays

The COVID-19 pandemic had a significant impact on road usage, including cyclist traffic. Lockdowns and curfews restricted movement, while COVID-19-related holiday restrictions altered normal traffic patterns.

- **Lockdown\_start, Lockdown\_end:** The start and end dates of each lockdown period.
- **Curfew\_start, Curfew\_end:** The start and end times of nightly curfews.
- **Holiday\_restriction:** Whether COVID-related travel restrictions were in effect during holidays.

**Source:** The COVID-19 dataset was compiled from government health reports, Paris municipal records, and media outlets tracking the pandemic and corresponding restrictions. Official lockdown and curfew data was used to identify specific timeframes.

### 3.4 Hospital Data: Situation During COVID-19

Hospital data was incorporated to track hospital capacity during the COVID-19 pandemic and analyze the effect of high hospitalization rates on public behavior and traffic patterns.

- **Hospital\_beds\_occupied:** Percentage of hospital beds occupied by COVID-19 patients.
- **ICU\_occupancy:** Percentage of Intensive Care Unit (ICU) beds occupied.
- **Covid\_cases:** Number of daily COVID-19 cases reported.
- **Public\_mobility:** The reduction in public mobility in response to hospital occupancy rates.

**Source:** This data was sourced from *Santé publique France*, which publishes hospital reports on bed occupancy and COVID-19 cases. The data is available at <https://www.santepubliquefrance.fr/>.

### 3.5 Public Holidays Data

Public holidays in France tend to influence traffic patterns, as many people shift from commuting to leisure activities, often increasing cyclist activity.

- **Holiday\_date:** Date of the public holiday.
- **Cyclist\_increase:** The percentage increase in cyclist activity on public holidays compared to normal weekdays.

**Source:** The public holiday data was sourced from official government records on national holidays in France.

## 4 Why These Datasets Were Chosen

We chose these datasets because each provides unique insights into the factors affecting cyclist traffic and safety in Paris:

- **Weather Data:** Offers environmental context that may influence traffic and accident risks.
- **Road Accidents Data:** Directly relates to cyclist safety, tracking accident frequency and severity.
- **Strikes Data:** Explains changes in cyclist traffic during public transportation strikes.
- **COVID-19 Data:** Essential for understanding the impact of pandemic-related restrictions on road usage.
- **Hospital Data:** Tracks public mobility in response to hospital capacity during the pandemic.
- **Public Holidays Data:** Accounts for increases in cyclist activity on holidays.

## 5 Conclusion

By integrating weather data, road accident records, and external datasets—such as strikes, COVID-19 lockdowns, hospital situations, and public holidays—our analysis provides a comprehensive view of the factors influencing cyclist traffic and safety in Paris. Each dataset was carefully selected to ensure a robust, multifaceted model for predicting cyclist activity and accident risks.