Traffic Accident Data Analysis (Global Dataset)

Name: Trần Phương Nam

Student ID: 2202085

# 1. Introduction

Traffic accidents are a major cause of death and injury worldwide. This report utilizes a global traffic accident dataset to analyze trends, causes, and high-risk conditions, with the goal of improving road safety awareness and suggesting preventive measures.

# 2. Objectives

- Analyze whether traffic accident frequency is increasing or decreasing.  
- Identify the most common weather and road conditions related to accidents.  
- Determine leading causes of accidents.  
- Evaluate severity based on number of vehicles involved and casualties.

# 3. Data Source

The data is sourced from Kaggle's 'Global Traffic Accidents Dataset'. It includes over 10,000 traffic accident records worldwide with details such as date, time, location, weather condition, road condition, vehicles involved, casualties, and cause.

# 4. Data Analysis Process

Step 1: Data Collection and Preprocessing

- Cleaned missing or invalid values.  
- Normalized fields like weather condition, road condition, and cause.  
- Converted date/time formats as needed for temporal analysis.

Step 2: Exploratory Data Analysis (EDA)

- Analyzed accident frequency over time.  
- Classified accidents by weather and road condition.  
- Calculated average casualties and vehicles per accident.  
- Grouped and counted accidents by cause.

Step 3: Visualization

- Line chart: Number of accidents over time.  
- Bar chart: Most frequent causes of accidents.  
- Pie chart: Distribution of weather conditions.  
- Heatmap: Accidents by hour of day.  
- Geo-map: Plotting accident density by coordinates using Folium (interactive map).

# 5. Findings and Observations

- The dataset indicates variability in accident rates by weather condition, with higher rates under rain or storm.  
- Leading causes include reckless driving and drunk driving.  
- Peak accident times are often early morning and evening hours.  
- Accidents involving more vehicles tend to result in more casualties.

# 6. Conclusion and Recommendations

Conclusion:

- The global dataset reveals critical insights about conditions and behaviors leading to traffic accidents.  
- Patterns observed can help in formulating safety strategies worldwide.

Recommendations:

- Promote driver education, especially regarding reckless and drunk driving.  
- Enhance infrastructure in high-risk areas.  
- Consider weather and road condition alerts integrated with driver navigation systems.

# 7. Appendix

- Dataset Source: https://www.kaggle.com/datasets/adilshamim8/global-traffic-accidents-dataset  
- Python libraries used: pandas, matplotlib, seaborn, plotly, folium (for interactive maps).  
- Folium interactive map is used to visualize accident locations by plotting latitude and longitude on a world map. Each marker on the map represents a traffic accident with a popup showing the cause and casualties.