TASK 5

Title: Traffic Accident Analysis – Patterns & Hotspots Detection

Objective

To analyze **traffic accident data** to identify patterns related to **road conditions**, **weather**, and **time of day**. Also, to **visualize accident hotspots** and understand key **contributing factors** that lead to accidents.

Dataset Information

- **Dataset Name:** US Traffic Accident Data
- **Source:** Kaggle
- **URL:** https://www.kaggle.com/code/harshalbhamare/us-accident-eda

- **Size:** Over 2.5 million accident records from 2016 to 2020

- **Key Features:**

- Start_Time, End_Time, Start_Lat, Start_Lng

- Severity, Weather_Condition, Visibility, Temperature

- Street, City, State

- Humidity, Wind_Speed, Precipitation

Steps Performed

1. Data Preprocessing

- Checked for null values and handled missing weather data.

- Converted datetime columns for time-based filtering.

- Removed outliers from Wind_Speed, Temperature, and Visibility.

2. Exploratory Data Analysis (EDA)

Time-Based Patterns:

- Found peaks in accidents between **7–9 AM** and **4–6 PM** (commute hours).
- Accidents more frequent during **weekdays** than weekends.

Weather and Road Conditions:

- High number of accidents during **fog**, **rain**, and **low visibility**.
- Severity 4 (most severe) mostly occurred in poor weather or late night.

Location-Based Analysis:

- Identified top 10 accident-prone **cities** and **states**.

- Used **heatmaps** to visualize **geographical hotspots** using Latitude and Longitude.

Key Insights

- **Weather** and **visibility** play a major role in accident severity.

- Accidents are concentrated during **peak traffic hours**.

- **Urban areas** like Los Angeles, Houston, and Miami showed more accident hotspots.

- Accident severity was higher at **night and early morning hours**.

Tools & Technologies Used

- **Programming Language:** Python
- **Libraries Used:** Pandas, NumPy, Matplotlib, Seaborn, Plotly, Folium
- **IDE/Notebook:** Jupyter Notebook / Google Colab
- **Map Visualization:** Folium (for heatmaps)

Conclusion

This analysis helped identify important patterns like **accident-prone times**, **weather influences**, and **geographic hotspots**. Authorities can use this insight to improve **road safety**, install **warning systems**, and **optimize patrol** during high-risk periods.

Link to Dataset

🔗 Kaggle - US Accident EDA