**Road Accident Analysis and Prediction in the UK 🚗📊**

**📌 Overview**

This project explores road accident data in the UK from 2017 to 2020, focusing on uncovering temporal, spatial, and behavioural patterns that contribute to road incidents. Leveraging SQL, data visualization, machine learning, clustering, and time series forecasting, the project provides actionable insights into accident trends and offers predictive modelling to assist with prevention and planning.

**📁 Datasets Used**

The analysis utilizes a structured dataset containing four main tables:

* **Accident**: When, where, and how accidents occurred.
* **Casualty**: Info about individuals involved, their role, injury severity, etc.
* **Vehicle**: Vehicle type, engine size, and driver demographics.
* **LSOA**: Location-level geographic mapping of accidents.

**🧪 Analytical Components**

**1. Temporal Analysis**

* **Hourly & Weekly Patterns**: Accidents peak during rush hours (8–9 AM & 4–6 PM) and Fridays are the most accident-prone days.
* **Motorcycle Accidents**: Smaller engines dominate weekday accidents; larger engines spike on weekends.
* **Pedestrian Accidents**: Concentrated during school and work commute hours.

**2. Accident Severity Analysis**

* **Method**: Association Rule Mining using the Apriori algorithm.
* **Findings**:
  + Slight accidents are more frequent in daylight, dry conditions, and at lower speed zones.
  + Gender, time of day, and weather influence severity.
  + No strong link found between specific variables and severe/fatal accidents.

**3. Spatial Clustering (DBSCAN)**

* **Regions Analyzed**: Kingston upon Hull, Humberside, East Riding of Yorkshire.
* **Output**: Clusters of accident hotspots and outliers.
* **Insights**:
  + Hull has concentrated hotspots.
  + Humberside shows a dispersed pattern.
  + East Riding falls in between.

**4. Social Network Analysis**

* **Dataset**: Facebook social network of 4,039 nodes and 88,234 edges.
* **Methods Used**:
  + Community Detection: Louvain and Label Propagation Algorithm (LPA)
  + Centrality Measures: Betweenness, Flow, Strength, Degree
* **Results**:
  + Louvain detects finer-grained communities.
  + LPA produces broader, fewer communities.

**5. Time Series Forecasting**

**a. Weekly Predictions by Police Force**

* **Models Compared**: ARIMA, Holt-Winters, ETS
* **Best Performers**: Holt-Winters & ETS (with grid search) for most accurate results.
* **Regions**: Metropolitan Police, West Midlands Police, Kent Police

**b. Daily Predictions for Hull LSOAs**

* **Top LSOAs Analyzed**: Based on highest accident counts in early 2020.
* **Models Used**: ARIMA, SARIMA, ETS
* **Findings**:
  + ARIMA fits well for LSOA E01012817.
  + SARIMA and ETS more stable for other two LSOAs.

**✅ Recommendations**

* **Targeted Enforcement**: Increase patrols during rush hours and weekends.
* **Safety Campaigns**: Focus on motorcyclists and pedestrian awareness.
* **Data-Driven Planning**: Deploy resources based on hotspot clustering and accident forecasts.
* **Infrastructure Enhancements**: Prioritize improvements in high-density clusters.

**📊 Tools & Technologies**

* **Languages**: Python, SQL
* **Libraries**: Pandas, Matplotlib, Seaborn, Scikit-learn, Statsmodels, NetworkX
* **Algorithms**: Apriori, DBSCAN, ARIMA, SARIMA, Holt-Winters, ETS
* **Platform**: Jupyter Notebook