**Problem Statement**

Road accidents remain a critical public safety concern, with multiple factors such as weather conditions, road surface conditions, speed limits, and traffic controls influencing accident severity.

However, raw accident data often exists in unstructured formats, making It difficult for stakeholders to derive actionable insights. Hence, you have been tasked as a Data Scientist to perform statistical analysis and Create dashboard for better understanding of the data.

**Aim of Analysis**

Is to help decision-makers in the:

* Analysis of Road Traffic Accidents
* Make Data-Driven Decisions
* Address Key Road Safety Challenges
* Enhance Road Safety Policies
* Optimized Resource Allocation

**Data Description**

*Accident\_Index:* Unique Identifier for each accident record

*Accident Date:* The date the accident occurred

*Month:* Month of the Accident

*Year:* Year of the accident

*Day\_of\_week:* The day of the week

*Time:* Time of the accident

*Accident\_Severity:* Category of the accident

*Latitude:* Coordinates of the accident location

*Local\_Authority:* The district where the accident happened

*Urban\_or\_Rural\_Area:* The area where the accident occurred

*Weather\_conditions:* The weather at the time of the accident

*Light\_Conditions:* The Light condition when the accident happened

*Road\_surface\_Conditions:* The state of the road at the time of the accident

*Junction\_Control:* Indicates the state of junction control at that point

*Junction\_Detail:* Detail on the junction type

*Road\_type:* Classifies the road

*Speed\_Limit:* The speed limit of the road where the accident occurred

*Carriageway\_Hazards:* Hazards like roadworks, oil spills or debris

*Number\_of\_Casualties:* The total number of people injured or killed in the accident

*Number\_of\_Vehicles*: The number of vehicles involved in the accident

*Vehicle\_Type:* Types of vehicles involved

*Police\_Force:* The police department in that area

**Data Cleaning Workflow**

1. Conversion of dataset to table
2. Day\_of\_Week: Format Change from General to Time, Correction of output using Text(Value, Format\_Text) Function
3. Speed\_Limit: Format Change From General to Number
4. Number\_of\_Casualties: Format Change From General to Number
5. Number\_of\_Vehicles: Format Change From General to Number
6. Time: Format Change From Custom to Time
7. Junction\_Control: Correction of by using Unique and Countif functions, to replace "Data missing or out of range" with "Give way or uncontrolled" which is our modal value in this column

**Descriptive Analysis**

Performed descriptive Statistics and correlation on the continuous variables ( Number\_of\_casualties, Number\_of\_Vehicles and Speed\_Limit) in our dataset using the data analysis command. Below are my observations

**Number of Casualties**

Mean (1.36): On average, each accident results in about 1.36 casualties.

Median & Mode (1): The most common and middle value is 1, meaning that most accidents result in only one casualty.

Standard Deviation (0.82): The spread of casualties per accident is relatively low, meaning the majority of accidents have similar casualty numbers.

Range (47): Some extreme accidents have up to 48 casualties, showing that severe accidents, although rare, do happen.

Skewness (5.69) & Kurtosis (129.86): A high positive skew means that most accidents have low casualties, but a few have extremely high casualties. The high kurtosis tells us that these extreme values are outliers.

**Number of Vehicles Involved**

Mean (1.83): On average, each accident involves about 1.83 vehicles, meaning many accidents are single-vehicle crashes.

Median & Mode (2): The most common number of vehicles involved is two, meaning that most accidents are collisions between two vehicles.

Standard Deviation (0.71): The spread of vehicle involvement per accident is relatively low, meaning most accidents involve similar vehicle numbers.

Range (31): The maximum number of vehicles in a single accident is 32, showing rare cases of multi-vehicle pile-ups.

Skewness (1.85) & Kurtosis (21.23): A moderate skew suggests that most accidents involve fewer vehicles, but there are rare multi-vehicle accidents.

**Speed Limit**

Mean (38.87): The average speed limit where accidents occur is about 39 mph.

Median & Mode (30): The most common speed limit where accidents happen is 30 mph, suggesting that many accidents occur in urban areas where this is the standard limit.

Standard Deviation (14.03): The speed limits vary quite a bit, which makes sense since accidents happen on both slow city roads and fast highways.

Range (60): Speed limits in accident locations range from 10 mph (very low, possibly school zones) to 70 mph (highways).

Skewness (1.14) & Kurtosis (-0.31): The skewness shows a slight right tail, meaning more accidents happen in lower-speed areas. The negative kurtosis suggests that the distribution is relatively flat, meaning accidents occur at a fairly even spread of speed limits.

**Key Takeaways from the Correlation Matrix**

**Number of Vehicles & Number of Casualties (0.23) →** A weak positive correlation means that accidents with more vehicles tend to have more casualties, but it's not a strong relationship. This suggests that multi-vehicle crashes are more dangerous but not always fatal.

**Speed Limit & Number of Casualties (0.14)** → A very weak positive correlation means higher speed limits slightly increase the number of casualties, but it's not a strong factor. This could mean that accidents in low-speed areas (urban roads) are more frequent but less severe, while high-speed areas (highways) have fewer but deadlier crashes.

**Speed Limit & Number of Vehicles (0.08) →** Almost no correlation. This means that accidents happen at all speed limits regardless of the number of vehicles involved.

**Key findings from the Dashboard**

Most accidents involve 1 or 2 vehicles and only 1 casualty.

Severe accidents with high casualties are rare, but they do exist.

Speed limits where accidents occur vary, but 30 mph is the most common.

Multi-vehicle accidents tend to have more casualties, but not always.

Higher speed limits slightly increase the number of casualties, but not significantly.

Speed limits do not strongly impact the number of vehicles involved in an accident.

Junction-Related Accidents: The highest number of accidents occur at "Give Way" or uncontrolled junctions, indicating a need for better traffic management at these points.

Vehicle Involvement: Two-vehicle collisions are the most common type of accident.

Accident Locations: The majority of accidents happen away from junctions or within 20 miles of them (123,000 accidents). T or staggered junctions account for 96,700 accidents. Together, these locations make up over 70% of all recorded accidents.

Accident Severity: 86% of accidents are classified as "Slight", 13% are "Serious", while only 1% result in fatalities.

Daylight vs. Nighttime Accidents: More than 73% of accidents occur during daylight hours, emphasizing the importance of daytime road safety measures.

Vehicle Types Involved: Cars are involved in over 77% of accidents, making them the most common vehicle type in collisions.

Road Hazards: Over 98% of accidents happen when there is no obstacle (such as debris, roadworks, or pedestrians) on the road, suggesting driver behavior is a key factor.

High-Accident Cities: Any pilot safety initiatives should first be tested in the top 7 cities with the highest accident rates: Birmingham, Leeds, Manchester, Bradford, Westminster, Sheffield, and Liverpool.

Time Distribution of Accidents: Accidents are evenly distributed across all months and days of the week, indicating no specific seasonal trend.

Peak accident hours are around 9 AM and 5 PM, coinciding with rush hours.

Weather Conditions: 79% of accidents occur in fine weather with no high winds, debunking the common belief that bad weather is the primary cause of accidents.

Road Types: 74% of accidents happen on single carriageways.16% occur on dual carriageways, while the remaining 10% happen on other road types.

Urban vs. Rural Accidents: The accident ratio between rural and urban areas is 16:9, indicating a higher number of accidents in urban settings.

Speed Limits: 65% of accidents occur on roads with a 30 mph speed limit, highlighting urban roads as accident hotspots.

Road Surface Conditions: 68% of accidents happen on dry roads, followed by 26% on wet or damp roads, reinforcing the idea that accidents are more linked to driver behavior than road conditions.

Police Jurisdictions with the Highest Accidents: The Metropolitan Police recorded the highest number of accidents (15%), far exceeding other regions.

The next four regions with the highest accident counts are: West Midlands (4.4%), West Yorkshire (3.9%), Greater Manchester (3.9%), Thames Valley (3.7%). This highlights that London (Metropolitan Police jurisdiction) has a significantly higher accident rate compared to other areas.

**Recommendations**

**1. Accident Hotspots Identified**

**A. High-Accident Cities:** Birmingham, Leeds, Manchester, Bradford, Westminster, Sheffield, and Liverpool. These cities should be prioritized for traffic safety interventions.

**B. High-Risk Road Types:** Single carriageways (74% of accidents) have the highest number of crashes.

* T or staggered junctions (96,700 accidents) and accidents occurring within 20 miles of a junction (123,000 accidents) account for more than 70% of all recorded accidents.

**C. Speed Limit Zones with High Accidents**

* 30 mph roads account for 65% of accidents, indicating urban areas as major accident hotspots.
* Speeding may not be the biggest contributor, as accidents occur across various speed limits.

**D. Peak Accident Times**

* 9 AM and 5 PM are the most accident-prone hours, coinciding with rush-hour traffic.
* Accidents are evenly distributed across all months and days of the week, meaning safety measures must be in place year-round.

**2. Proposed Traffic Safety Interventions**

**A. Infrastructure Improvements**

* + Install traffic lights and roundabouts at high-risk junctions to replace uncontrolled “Give Way” junctions.
  + Implement smart traffic signals that adjust based on real-time congestion levels.
  + Increase the number of speed cameras and enforcement officers on roads with frequent high-speed crashes.
  + Introduce variable speed limits during peak hours to manage congestion and reduce accident risks.
  + Enhance street lighting in accident-prone areas to reduce nighttime crashes.
  + Improve signage and road markings to ensure better visibility, especially at high-risk junctions.
  + Regular maintenance to fix potholes and road defects that may contribute to crashes.

**B. Policy Recommendations**

* + Increase penalties for reckless driving and distracted driving (e.g., using mobile phones while driving).
  + Implement a zero-tolerance policy on drunk driving, with random breath tests in high-accident areas.
  + Mandatory speed limit monitoring systems in commercial and public transport vehicles.
  + Designate pedestrian-friendly zones in busy urban centers to minimize vehicle-pedestrian accidents.
  + Expand public transport networks to reduce congestion and accidents during peak hours.
  + Equip police forces with AI-driven accident prediction systems to improve response time.
  + Deploy more patrol officers in high-accident zones based on historical accident data.

**C. Public Awareness Campaigns**

* + Young driver safety programs: Provide additional training for new drivers, as they are more prone to accidents.
  + Elderly driver assessments: Encourage routine safety evaluations for older drivers to ensure they can still drive safely.
  + Run media campaigns (TV, radio, social media) to highlight the dangers of speeding, distracted driving, and not using seat belts.
  + Encourage employers to adopt road safety policies for employees who drive for work.
  + Set up neighborhood road safety forums where residents can report dangerous road conditions or suggest safety improvements.
  + Organize school-based road safety programs to educate children on traffic rules and safe pedestrian practices.