Data Cleaning and Analysis Process

Data Cleaning

The dataset underwent a detailed cleaning process to ensure accuracy and completeness. Below are the steps taken:

1. Duplicate Removal:

   - I Identified and removed 1 duplicate row from the dataset.

2. Handling Errors and Missing Data:

   - Junction Control Field: I corrected 93 instances of "Auto traffic sigl" to "Auto traffic signal."

   - Accident Severity Field: I changed 49 occurrences of "fetal" to "fatal."

   - Road Surface Conditions Field: I labeled 317 blank rows as "Unknown."

   - Road Type Field: I changed 1,534 blank rows to "Others."

   - Weather Conditions Field: I filled 6,057 blank rows with "Unknown."

3. Creating Additional Columns:

   A secondary KPI required showing a monthly trend, hence the need to create new columns; ‘Month’ and ‘Year’:

   - Month: Extracted using the formula =TEXT(B2, "MMM").

   - Year: Extracted using the formula =TEXT(B2, "YYYY").

Data Analysis

Pivot tables were created to calculate key metrics and support visualization. The analysis was divided into primary and secondary KPIs.

Primary KPIs:

1. **Total Casualties:**

   - A pivot table revealed the total casualties across all accidents: 417,882.

2. **Casualties by Severity:**

   - I created a pivot table to calculate casualties for each severity level and their percentage of the total.

   - The data was then visulaized using donut charts.

3. **Casualties by Vehicle Type:**

   - I grouped vehicle types into broader categories using a helper column, ‘Vehicle Category’.

Secondary KPIs:

1. **Casualties by Vehicle Category:**

   - I displayed the sum of casualties by vehicle category on the dashboard.

2. **Monthly Trends:**

   - I created pivot tables to show monthly casualty trends for 2021 and 2022.

   - Line charts were used to compare trends side by side.

3. **Casualties by Road Type:**

   - I analyzed casualties for each road type and visualized them using a bar chart.

4. **Distribution by Road Surface:**

   - I classified road surface conditions into four categories: Dry, Wet, Snow/Ice, and Unknown.

   - I then created a custom treemap-like visualization due to Excel's version limitations.

5. **Casualties by Location**:

- I categorized locations into Urban and Rural using a helper column and visualized this with a donut chart.

6. **Casualties by Time of Day**:

- I created another helper column to classify "Light Conditions" into "Day" and "Night."
- Visualized the results using a donut chart.

Dashboard Creation

The cleaned and analyzed data was used to create an interactive dashboard with the following features:

- **Primary and Secondary KPIs**: Visualized using bar charts, line charts, donut charts, and custom treemap-like charts.
- **Slicer Functionality**: This was created to allow filtering by geographic location for a tailored view of the data.
- **Data Sheet**: Included all pivot tables for detailed reference by stakeholders.

Enhancements for User Interaction

1. **Sidebar Navigation:**

- I linked the dashboard and data sheets for seamless switching.
- I alos added clickable icons for email and external contact channels.

2. **Formatting**:

- I applied a clean and minimalistic color scheme for better readability and aesthetics.