Digital Egypt Pioneers Initiative (DEPI)

Traffic Accident Analysis

Group 1
Submitted by:
Ahmed Atef
Marvy Ayman
Yomna Shahin
Yasmeen El said

Supervised by:
Dr. Eman Ghanem

Acknowledgements

---

1. Problem Statement

Traffic accidents are a significant public safety concern, leading to injuries and fatalities. Understanding the patterns and factors contributing to these incidents is crucial for effective prevention and policy-making.

2. Data-Driven Questions

- What are the trends and contributing factors associated with traffic accidents over time?
- How do different time periods affect accident frequency?
- What environmental and situational conditions contribute to traffic accidents?
- How can data insights be used to improve public safety measures?

---

3. Modeling of the Data Tables

3.1 Date Table (DimDate_Table)

A Date Table was created to facilitate time-based analysis. It contains the following attributes:

- **Date**: The specific date of the incident.
- **Year**: The year of the incident.
- **Quarter**: The quarter of the year (Q1, Q2, Q3, Q4).
- **QuarterNo**: Numeric representation of the quarter (1-4).
- **Month**: The month in which the accident occurred (e.g., Jan, Feb).
- **MonthNo**: Numeric representation of the month (1-12).
- **WeekDay**: The abbreviated name of the day (e.g., Mon, Tue).
- **WeekDayNo**: The numeric representation of the weekday (1-7).

3.2 Date Periods Table (DimDate_Periods)

This table allows dynamic filtering and analysis based on different time frames. It includes:

- **QTD (Quarter-to-Date)**: Filters data from the beginning of the current quarter to the selected date.

- **YTD (Year-to-Date)**: Filters data from the start of the current year to the selected date.

```
DimDate_Table = ADDCOLUMNS(
    CALENDARAUTO(),
    "Year", year([Date]),
    "Quarter", FORMAT([Date],"\QQ"),
    "QuarterNo", Quarter([Date]),
    "Month", format([Date],"MMM"),
    "MonthNo", MONTH([Date]),
    "WeekDay",format([Date],"DDD"),
    "WeekDayNo",DAY([Date])
    )
```

```
DimDate Periods =
VAR Dates_YTD = DATESYTD(FactTraffic_accidents[Correct_crash_day])
VAR Dates_QTD = DATESQTD(FactTraffic_accidents[Correct_crash_day])
RETURN
UNION(
    SELECTCOLUMNS(
        DimDate_Table,
        "Date", DimDate_Table[Date],
        "Level", "All Year",
        "Level Sort", 1
    ),
    SELECTCOLUMNS(
        FILTER(DimDate_Table, DimDate_Table[Date] IN Dates_YTD),
        "Date", DimDate_Table[Date],
        "Level", "YTD",
        "Level Sort", 2
    ),
    SELECTCOLUMNS(
        FILTER(DimDate_Table, DimDate_Table[Date] IN Dates_QTD),
        "Date", DimDate_Table[Date],
        "Level", "QTD",
        "Level Sort", 3
    )
)
```

3.3 Relationships

- **Date Table to Date Periods Table**: A Many-to-One relationship enabling cross-filtering between both tables for better trend analysis.

- **Date Table to Traffic Accidents Table**: A Single relationship with Many-to-One cardinality, allowing accident data to be filtered based on date selections.


---
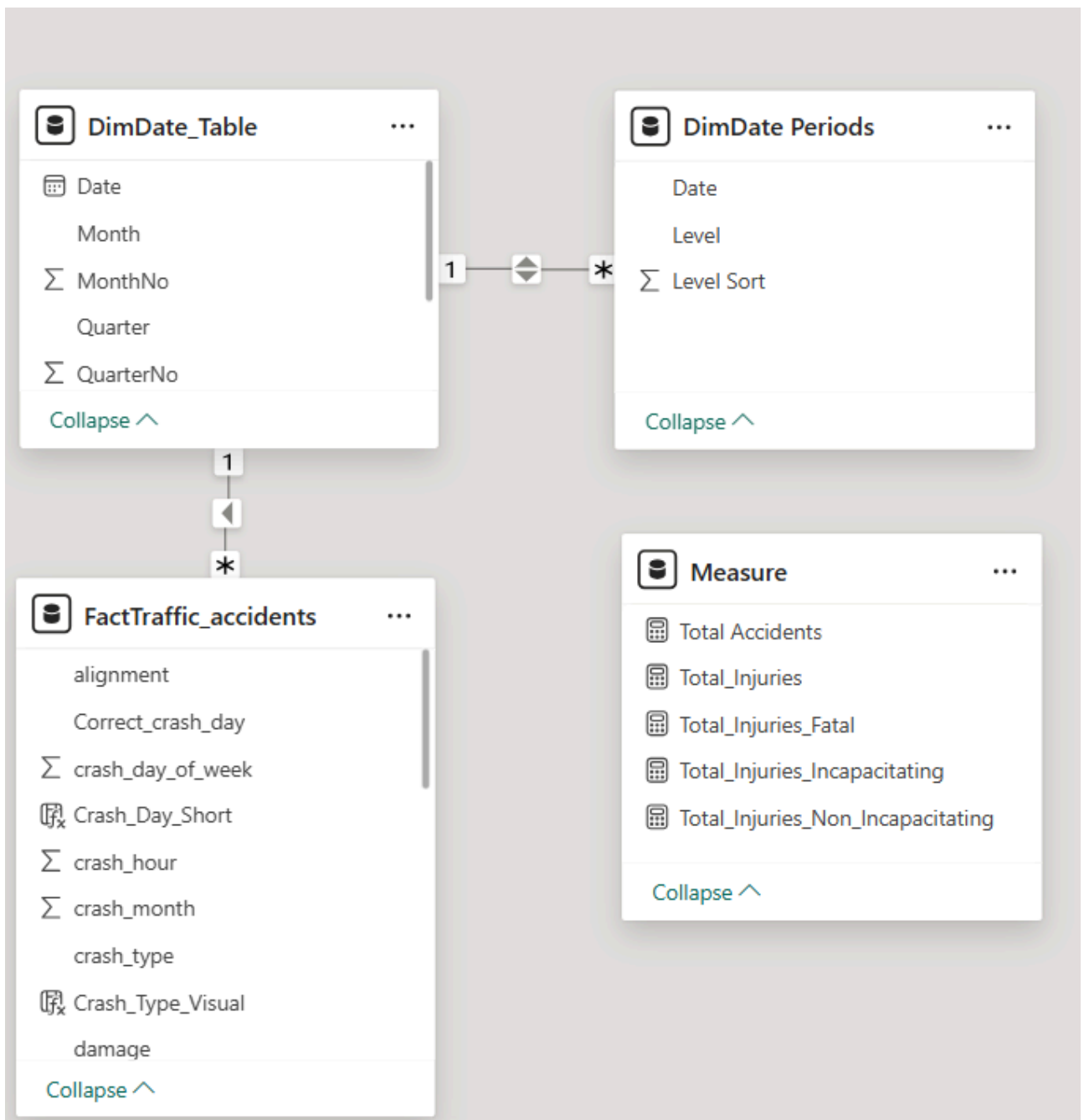
4. Dashboard Overview

The dashboard is designed to provide a comprehensive overview of traffic accidents using various slicers, filters, and visualizations.

4.1 Interactive Features

- **Date Slicers**: Allow selection of specific time frames.

- **Popup Filters**: Provide additional filtering options without cluttering the main interface.
- **Dynamic Charts**: Visualizations update based on user selections.
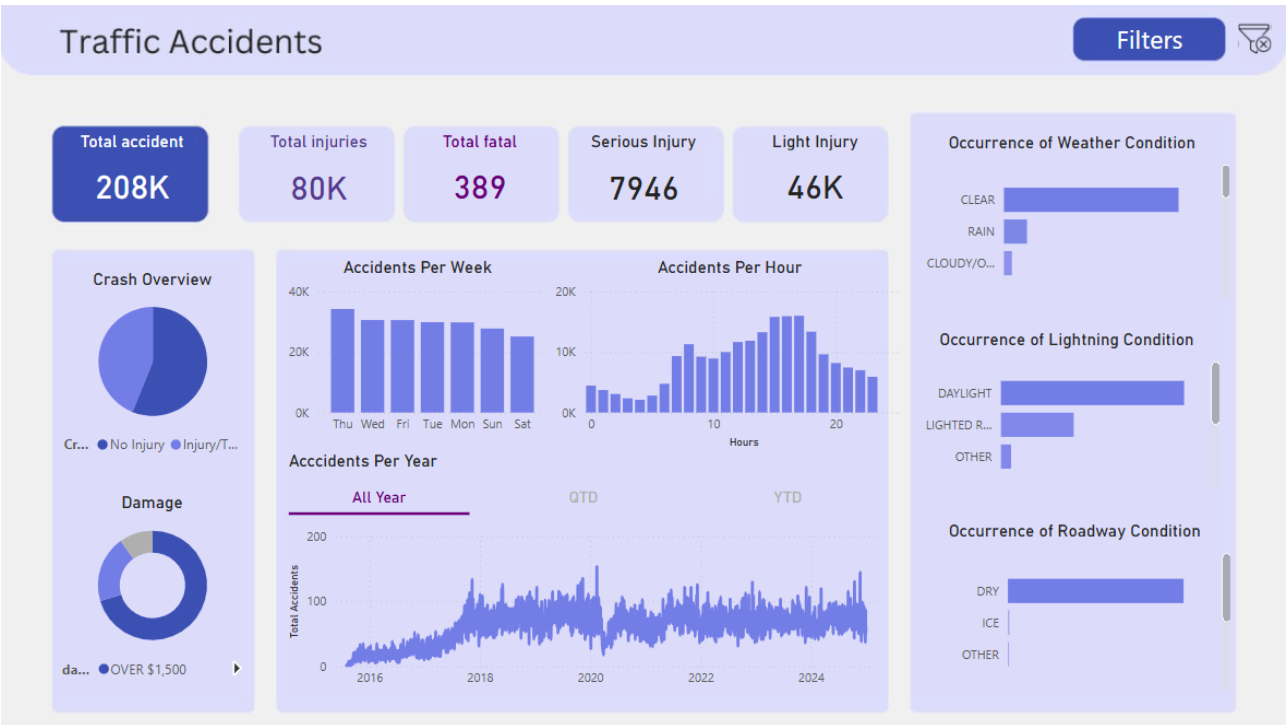
---

5. Chart Descriptions

5.1 Traffic Accident Overview

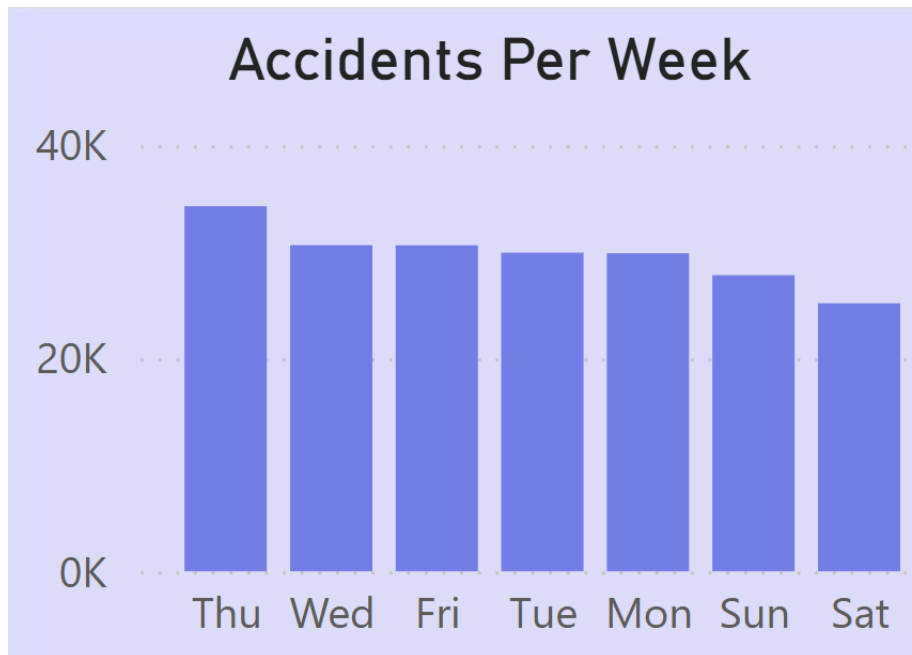| Total accident	Total injuries	Total fatal	Serious Injury	Light Injury
208K	80K	389	7946	46K

- Displays the total number of accidents (**208K**), categorized into:
  - Total injuries: **80K**
  - Serious injuries: **389**
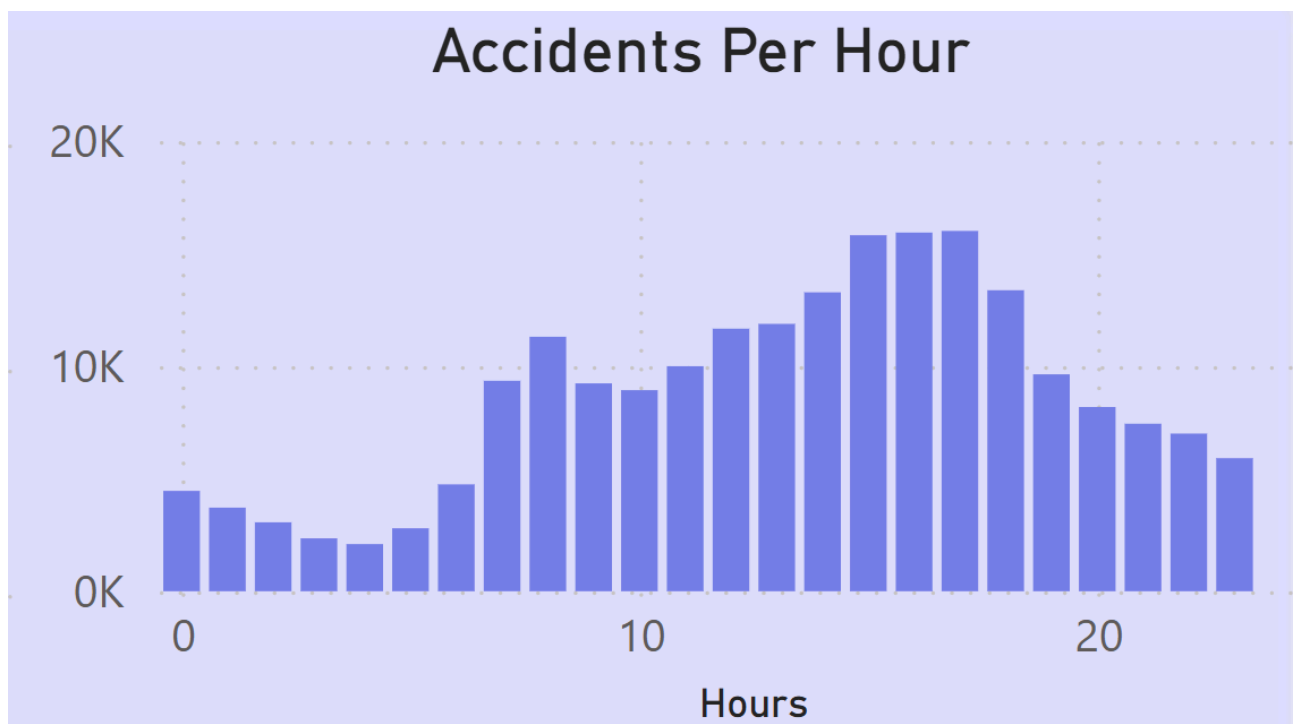 - Light injuries: **46K**

- This provides a high-level summary of accident frequency and severity.

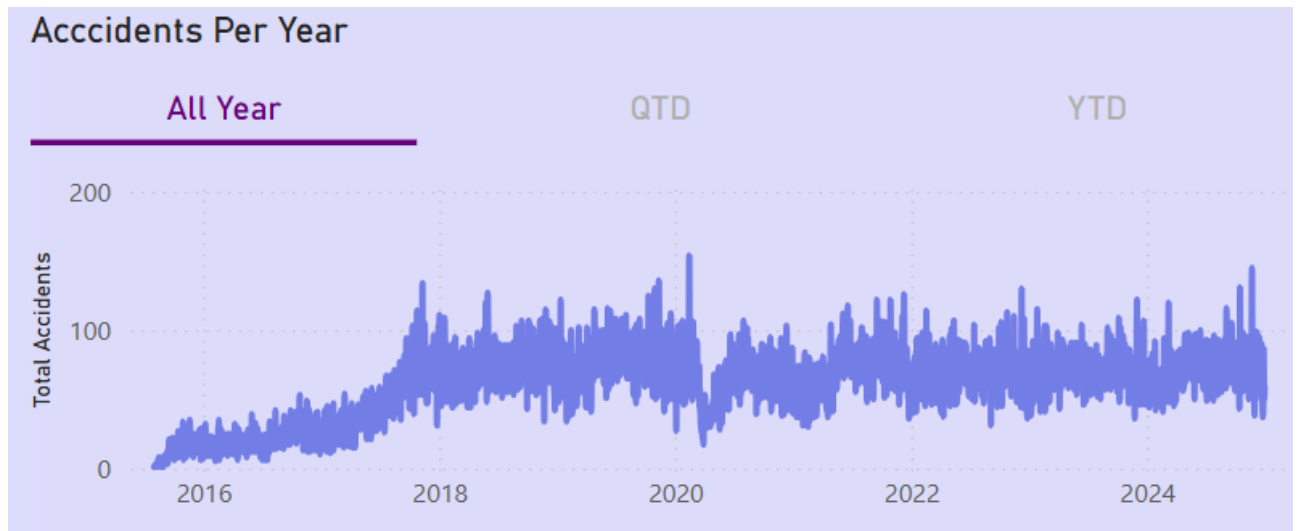## 5.2 Accidents Per Week

Accidents Per Week

- Shows the number of accidents recorded each week.
- Identifies peak accident periods and helps analyze seasonal variations.

5.3 Accidents Per Hour
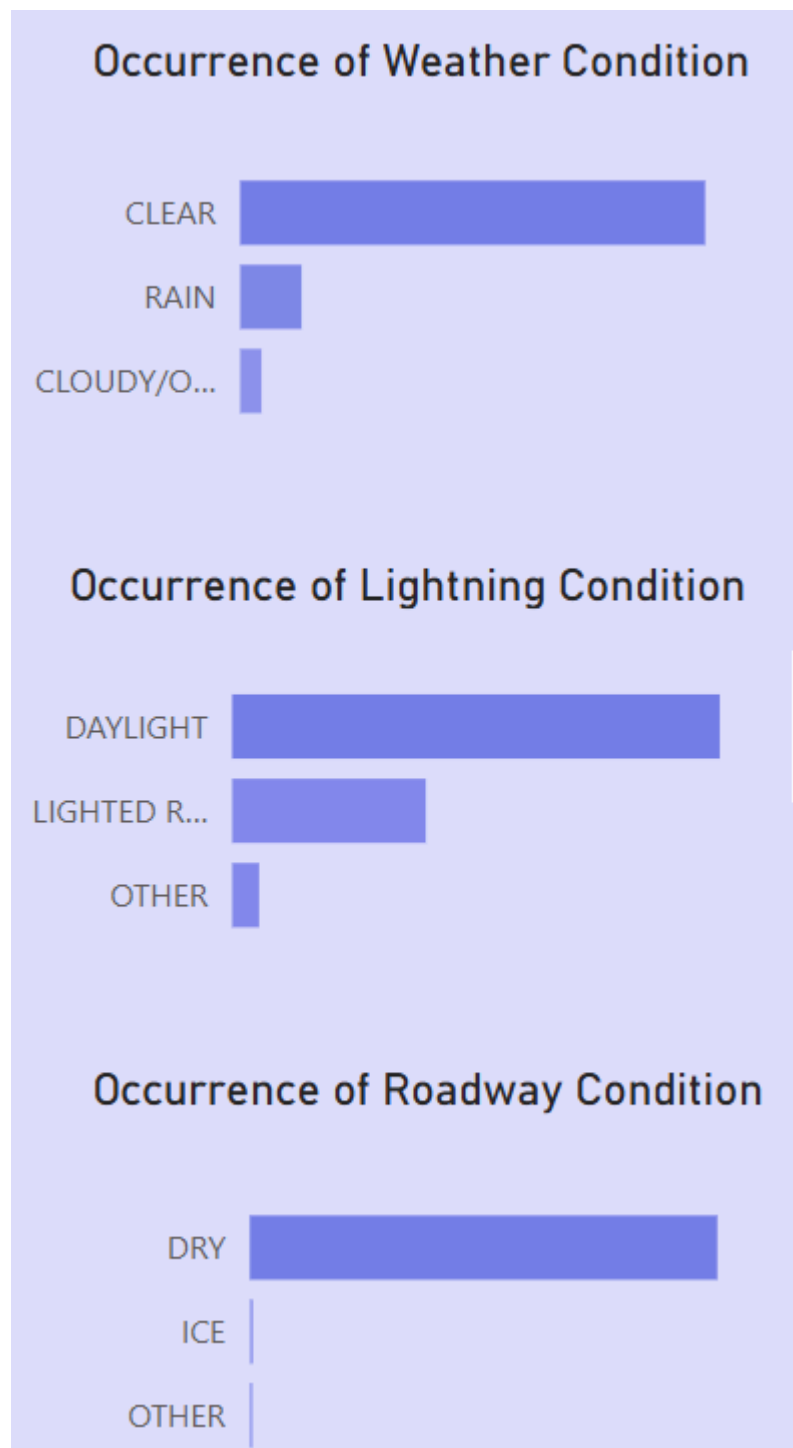
Accidents Per Hour

- Displays accident distribution per hour of the day.
- Highlights high-risk periods where accidents are most frequent (e.g., rush hours).

5.4 Accidents Per Year

- Shows the yearly trend of traffic accidents.
- Helps assess whether accident rates are increasing or decreasing over time.

5.5 Environmental Factors Influence

Occurrence of Weather Condition

CLEAR

RAIN

CLOUDY/O...

Occurrence of Lightning Condition

DAYLIGHT

LIGHTED R...

OTHER

Occurrence of Roadway Condition

DRY

ICE

OTHER

- Examines how weather conditions (rain, fog, snow) impact accident rates.
- Provides insights for improving road safety measures during adverse conditions.

5.6 Injury Severity Breakdown

Crash Overview

Cr... ● No Injury ● Injury/T...

Damage

da... ● OVER $1,500 ▶
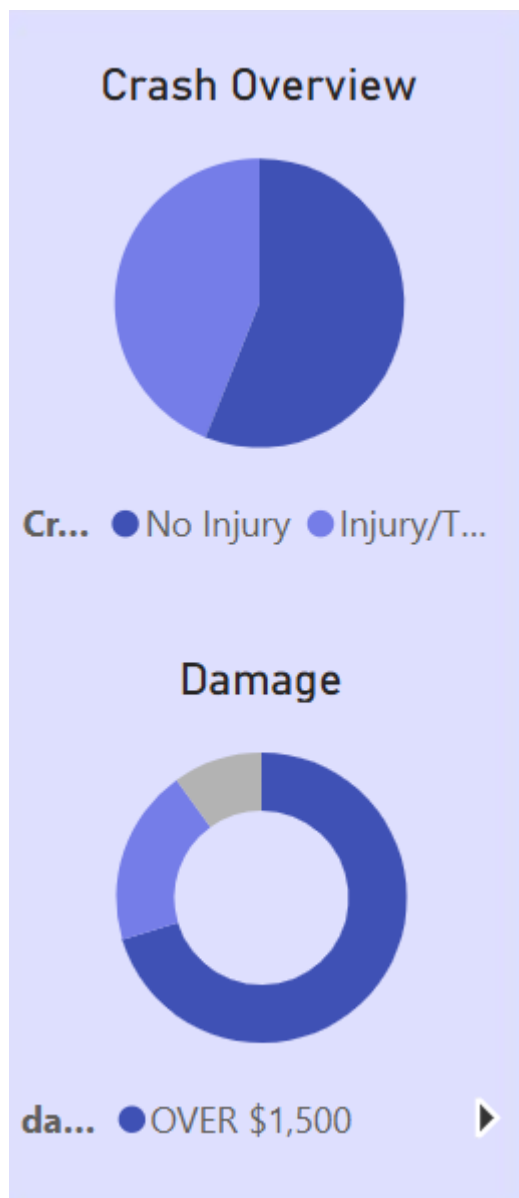
- The **Crash Overview** pie chart categorizes accidents based on injury severity:
    - **No Injury**: A significant portion of accidents result in no reported injuries.
    - **Injury/Treatment Required**: A notable percentage of crashes involve injuries requiring medical attention.
- The **Damage** donut chart provides insight into the financial impact of accidents:
    - **Over $1,500 in damages**: The majority of recorded accidents result in damage exceeding $1,500.
    - **Lower Damage Categories**: Some incidents involve minor damage, as represented by the lighter segments.
- These visualizations highlight the need for preventive measures to reduce both injury severity and financial consequences of accidents.

6. Insights

- **Peak Accident Times**: Most accidents occur during morning and evening rush hours.
- **Environmental Factors**: Rainy conditions correlate with increased accident rates.
- **Injury Trends**: Serious injuries occur more frequently at high-speed zones.
- **Seasonal Variations**: Higher accident rates observed during holiday seasons.

---

7. Recommendations

- **Policy Implementation**: Implement stricter speed regulations during peak accident hours.
- **Public Awareness Campaigns**: Educate drivers on the risks associated with poor weather conditions.
- **Traffic Law Enforcement**: Increase police presence in high-accident zones.
- **Further Research**: Continue data-driven analysis to refine preventive measures.

---

8. Derived Columns (Data Cleaning Reference)

DAX:

```
Season =
SWITCH(
    TRUE(),
    FactTraffic_accidents[crash_month] IN {12, 1, 2}, "Winter",
    FactTraffic_accidents[crash_month] IN {3, 4, 5}, "Spring",
    FactTraffic_accidents[crash_month] IN {6, 7, 8}, "Summer",
    FactTraffic_accidents[crash_month] IN {9, 10, 11}, "Autumn",
    "Unknown"
)
```

M Query:

```
Periods
= Table.AddColumn(#"Renamed Columns", "Periods", each
    if [crash_hour] >= 5 and [crash_hour] <= 11 then "Morning"
    else if [crash_hour] >= 12 and [crash_hour] <= 16 then "Afternoon"
    else if [crash_hour] >= 17 and [crash_hour] <= 20 then "Evening"
    else "Night"
)
```

```
final_road_surface_condition
= if [roadway_surface_cond] = "UNKNOWN" and [weather_condition] = "RAIN" then "WET"
    else if [roadway_surface_cond] = "UNKNOWN" then "DRY"
    else [roadway_surface_cond]
```

```
final_lighting_condition
= if [lighting_condition] = "UNKNOWN" and ([Periods] = "Morning" or [Periods] =
"Afternoon") then "DAYLIGHT"
    else if [lighting_condition] = "UNKNOWN" and ([Periods] = "Evening" or [Periods] =
"Night") then "DARK"
    else [lighting_condition]
```

During the data cleaning process, several derived columns were created using **DAX** and **Power Query (M Query)** to improve data consistency and enable more accurate analysis. The key derived columns include:

DAX-Derived Columns

1. **Season**
    - Categorizes accidents based on the month of occurrence:
        - **Winter**: December, January, February
        - **Spring**: March, April, May
        - **Summer**: June, July, August
        - **Autumn**: September, October, November
    - Implemented using the SWITCH function in DAX.

Power Query (M Query) Derived Columns

2. **Final Lighting Condition**
   - Handles unknown lighting conditions based on crash periods:
     - If `lighting_condition` is "UNKNOWN" and `Periods` is **Morning/Afternoon**, assign **"DAYLIGHT"**.
     - If `lighting_condition` is "UNKNOWN" and `Periods` is **Evening/Night**, assign **"DARK"**.
     - Otherwise, retain the original `lighting_condition`.

3. **Periods (Time of Day Classification)**
   - Classifies crash times into four periods:
     - **Morning**: 5 AM – 11 AM
     - **Afternoon**: 12 PM – 4 PM
     - **Evening**: 5 PM – 8 PM
     - **Night**: 9 PM – 4 AM

4. **Roadway Surface Condition (Handling Unknown Values)**
   - If `roadway_surface_cond` is **"UNKNOWN"** and `weather_condition` is **"RAIN"**, assign **"WET"**.
   - If `roadway_surface_cond` is **"UNKNOWN"**, assign **"DRY"**.
   - Otherwise, retain the original `roadway_surface_cond`.

5. **Standardized missing and inconsistent values in the dataset.**