

# Comprehensive Final Report on Traffic Accidents Analysis



D.S. Abdel Rahman  
Ahmed

D.S. Israa Ibrahim

D.S. Omer Farghal

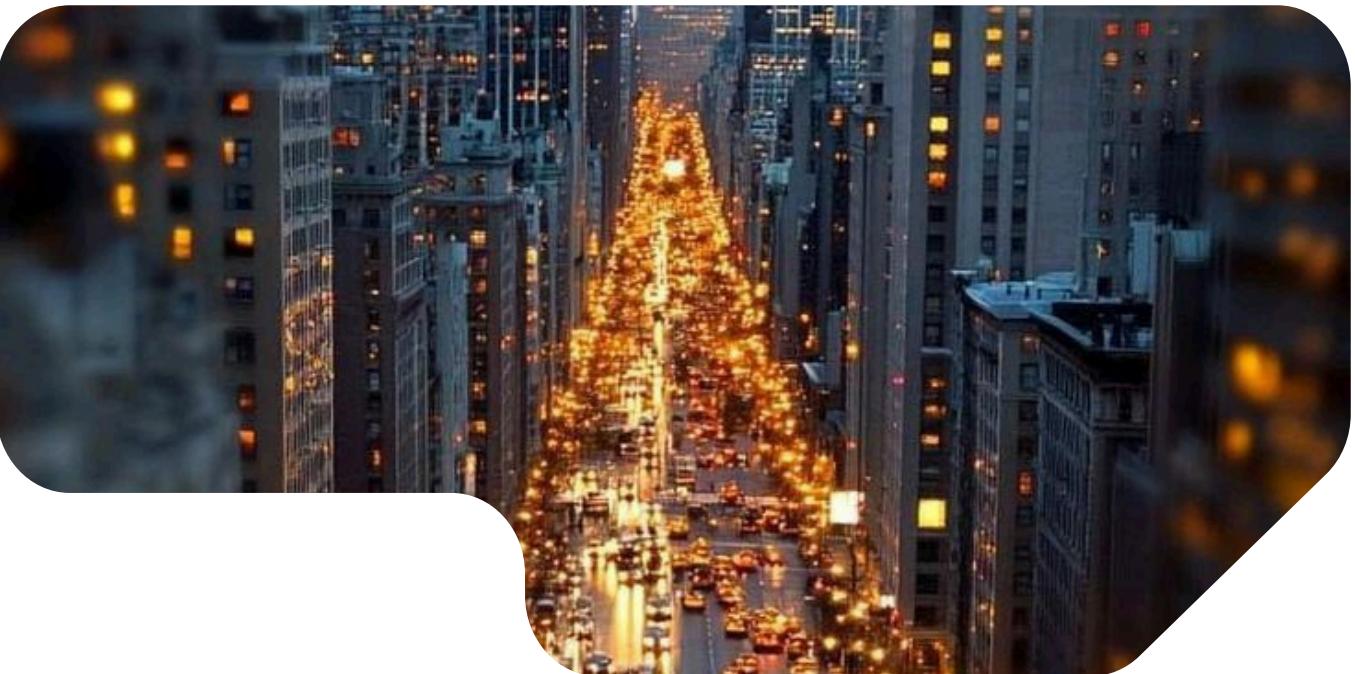
D.S. Abdel-Rahman  
Sameh

D.S. Manar  
Mohammed

D.S. Silvana Samir



# Project Overview



This project analyzes traffic accident data collected across multiple countries (USA, UK, Canada) over various years, months, and regions. The dataset contains 35 variables describing environmental, human, and vehicle-related factors that contribute to accidents. The goal of the analysis is to identify the key patterns, risk factors, and drivers of accident severity, ultimately providing insights that can support prevention strategies and safety improvements.

# Data Quality & Cleaning using Python:

- An inconsistency was found between the region and country fields, where some countries are not mapped to the correct region.
- All columns were kept because each feature contributes meaningful information.
- Data types are already appropriate for all variables.
- The dataset is complete, with all fields fully populated and no missing entries.
- All variables fall within reasonable ranges, and no extreme or unusual values were observed.
- All records are unique, with no duplicated rows.
- Data entries are clean and properly formatted, with standardized text and no typographical issues.



# Data Cleaning via Excel

- The dataset was clean — there were no missing or null values.
- The 'Medical Cost' column contained many decimal places, so it was rounded to 2 decimal places.
- Emergency Response, Visibility Level, and Population Density → Rounded to 2 decimal places.
- Economic Loss → Rounded to 1 decimal place.
- Time → Rounded to 0 decimal places (converted to integer).
- Traffic Volume → Rounded to 1 decimal place.
- Driver Alcohol Level → Rounded to 2 decimal places (values were small percentages).

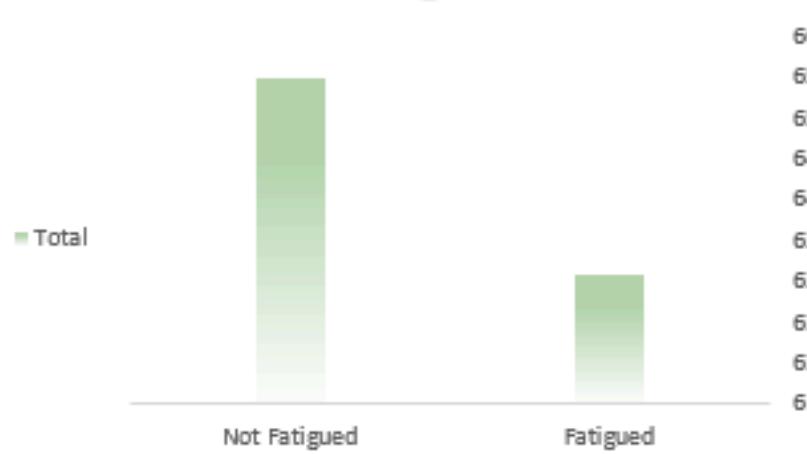
# Data Preprocessing via Excel

- Created a new column named 'Visibility Category' using Add Column → Custom Column.
- Formula: if [Visibility Level] <= 250 then 'Low', else if [Visibility Level] <= 400 then 'Medium' else 'High'.
- Added a new column named 'Total Loss' using Add Column → Custom Column.
- Formula: [Medical Cost] + [Economic Loss].
- Changed the data type to Decimal Number.
- Created a new column 'Severity Numeric' to convert text values to numbers using Add Column
- Conditional Column (Mapping: Minor = 1, Moderate = 2, Severe = 3).
- Added a column named 'Accident Type' using Add Column → Custom Column.
- Formula: if [Accident Cause] = 'Distracted Driving' or [Accident Cause] = 'Drunk Driving' or
- [Accident Cause] = 'Speeding' then 'Human' else if [Accident Cause] = 'Mechanical Failure' then
- 'Mechanical' else if [Accident Cause] = 'Weather' then 'Weather' else 'Other'.
- Created a new column named 'Driver Fatigue (Text)' using Add Column → Conditional Column.
- Formula: if [Driver Fatigue] = 1, then 'Fatigued' else 'Not Fatigued'.
- Pivot Table Analysis
- Created a set of pivot tables to answer all the analytical questions in the dataset.

# EXCEL DASHBOARD

## Accident Data Analytics Dashboard

Count Accident Per Driver Fatigue



Year				
2000	2001	2002	2003	2004
2005	2006	2007	2008	2009
2010	2011	2012	2013	2014
2015	2016	2017	2018	2019
2020	2021	2022	2023	2024

Month				
April	August	December	February	
January	July	June	March	
May	November	October	September	

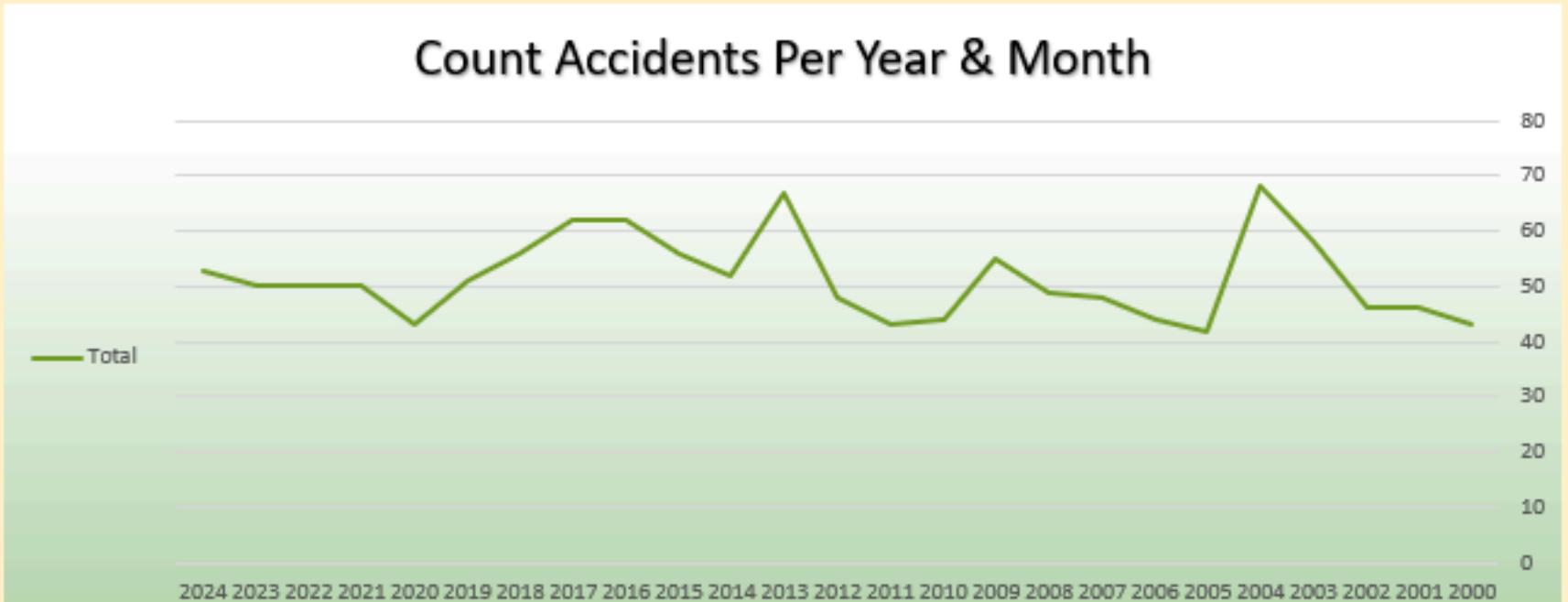
Country				
Austr...	Brazil	Canada	China	Germany
India	Japan	Russia	UK	USA

Total Loss  
101,143,591.4

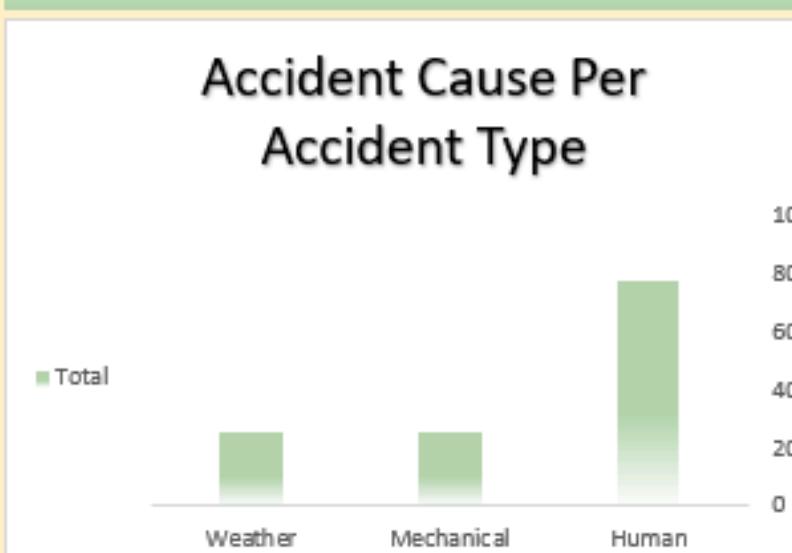
Total Accident  
1286

Total Injuries  
12081

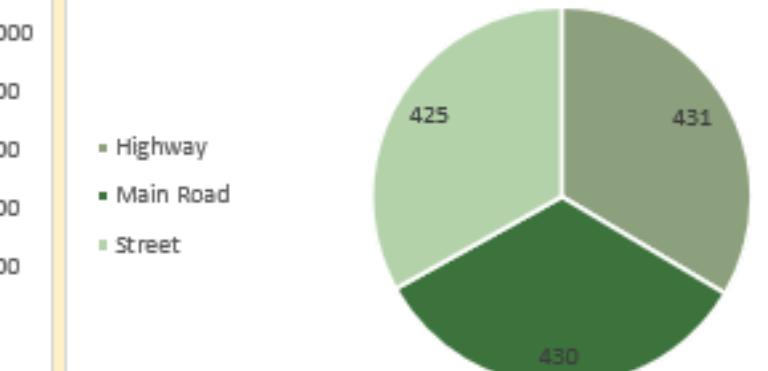
### Count Accidents Per Year & Month



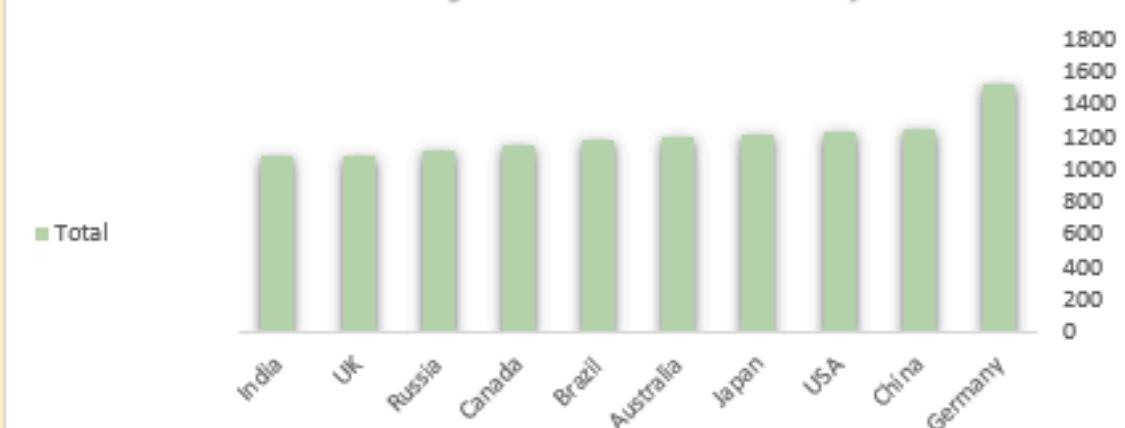
### Accident Cause Per Accident Type



### Count Accidents Per Road Type



### Total Injuries Per Country



# Main Insights via SQL

- Calculated the percentage of accidents resulting in fatalities and injuries.
- Compared countries based on accident frequency and severity levels.
- Analyzed how visibility affects accident severity using average severity scores.
- Identified the most dangerous conditions and trends across the dataset.
- The results help understand which factors increase accident risks and support data-driven decisions for improving road safety.





## Accident Causes & Severity:

- **Causes:** The causes of accidents are almost evenly distributed between Mechanical failures (13,018), Speeding (13,011), and Weather (12,972).
- **Severity of Cause:** For all causes, the accident severity is also spread evenly across Minor, Moderate, and Severe categories.
- **Driver Fatigue:** Fatigue is a critical factor in severe accidents. The number of severe accidents involving fatigued drivers (10,952) is slightly higher than those with non-fatigued drivers (10,778).
- **Fatigue & Vehicle Condition:** For severe accidents where the driver was fatigued, the vehicle's condition made little difference; the numbers were nearly identical for "Good" (3,660), "Moderate" (3,688), and "Poor" (3,604) condition vehicles.
- **Fatigue & Vehicle Condition & Road Condition:** The road condition made little difference; the numbers were identical for "Icy & Wet" (951), less in "Snow Covered" (896).
- **Fatigue & Vehicle Condition & Road Condition & Age Group:** The age group of the driver made little difference; the numbers were nearly identical for "(41-60), (61+), (18-25)" (201, 193, 189).

# POWER BI DASHBOARD P1

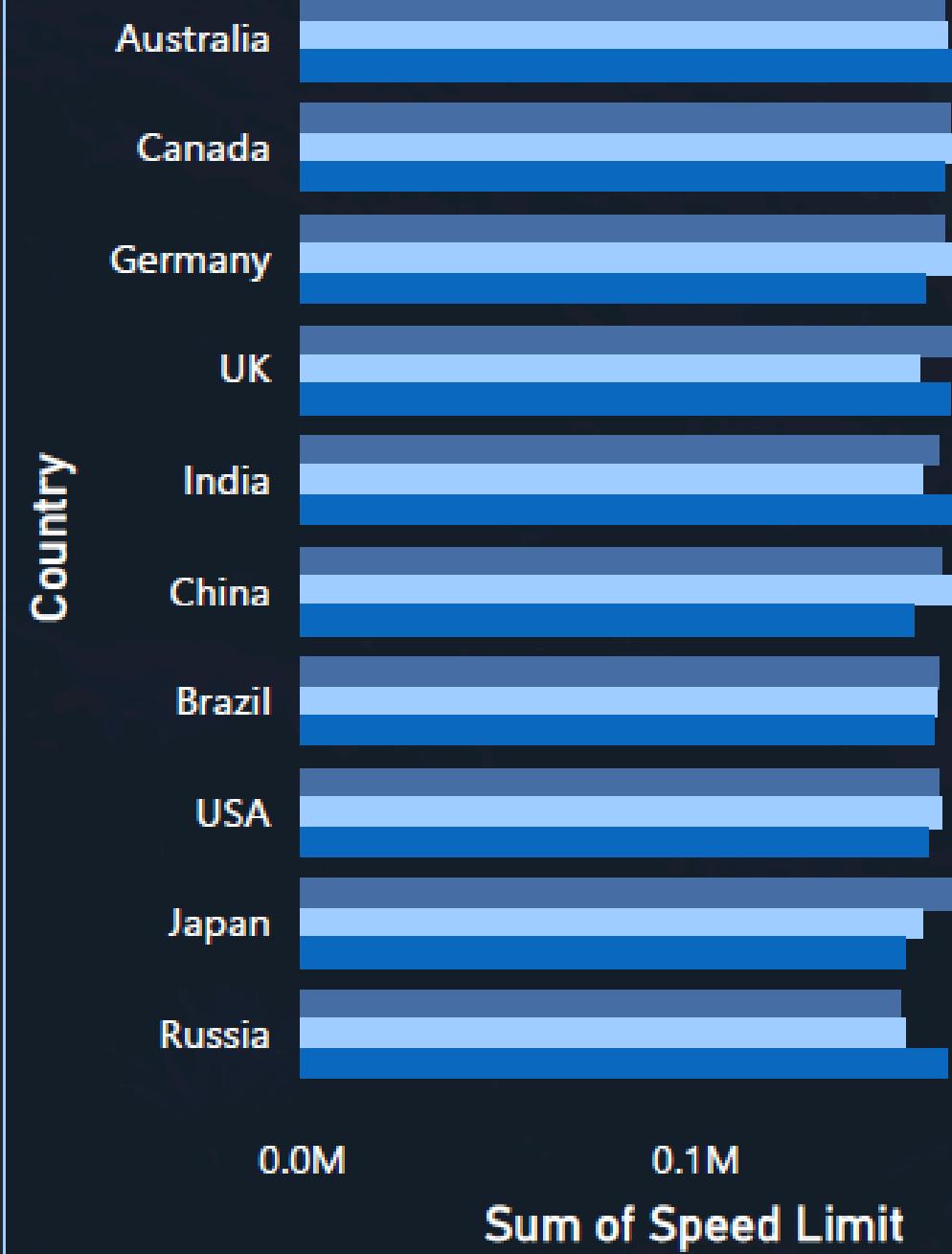
## Accident Factors

Page 2

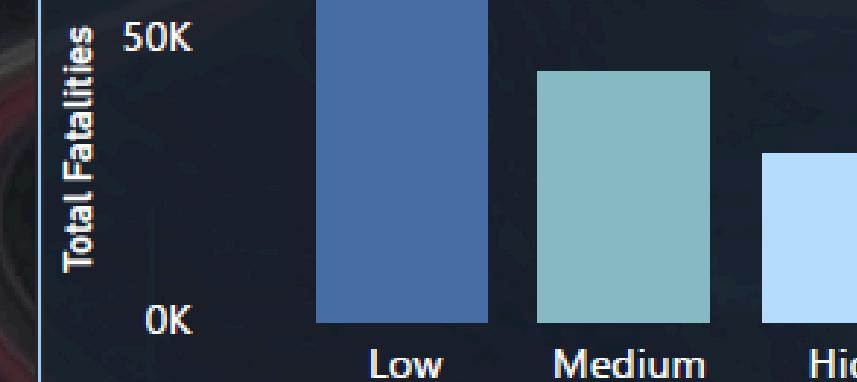
Page 3

### Severity per Speed Limit & Country

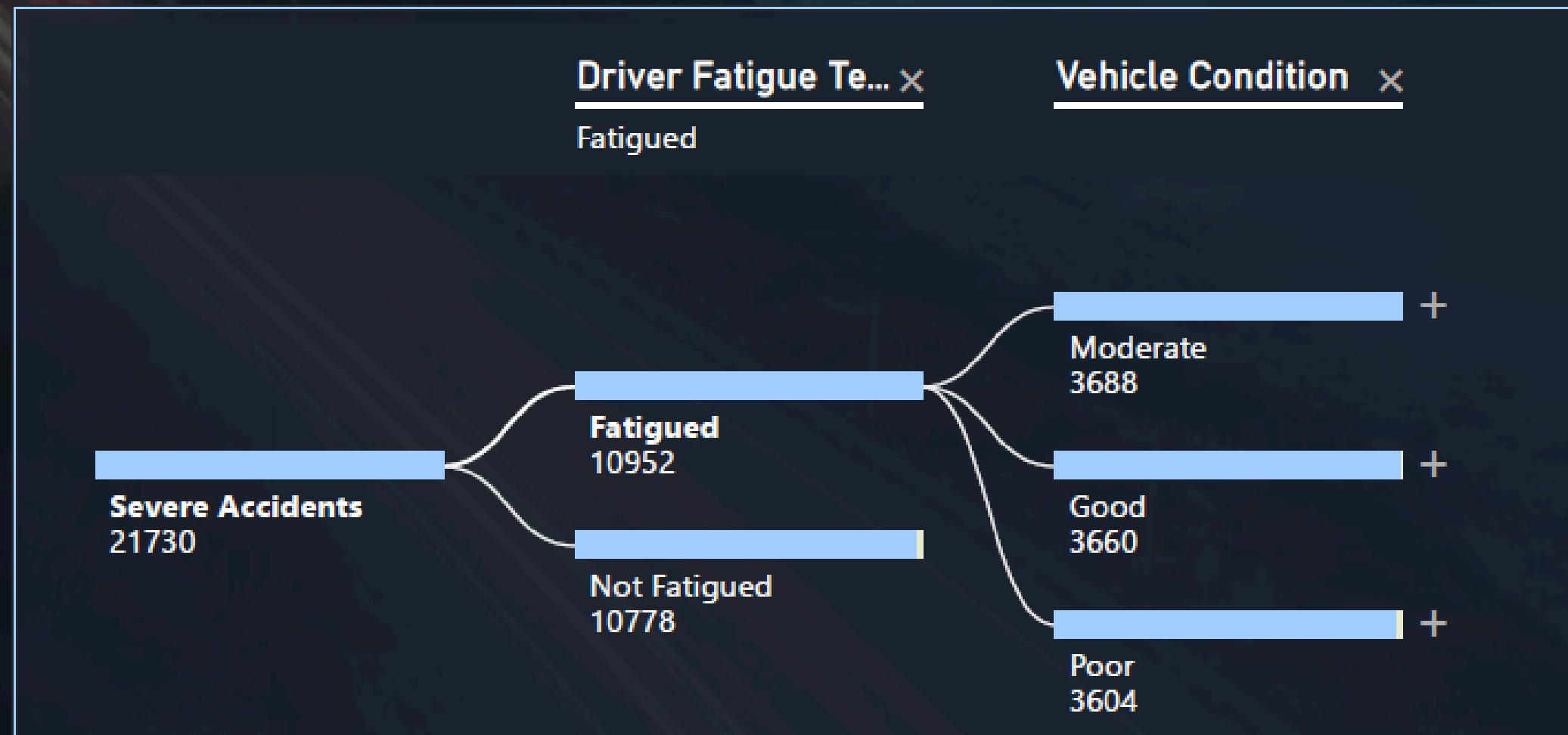
Accident ... ● Minor ● Moderate ● Severe



### Fatalities by Visibility Level



### Fatalities by Urban/Rural

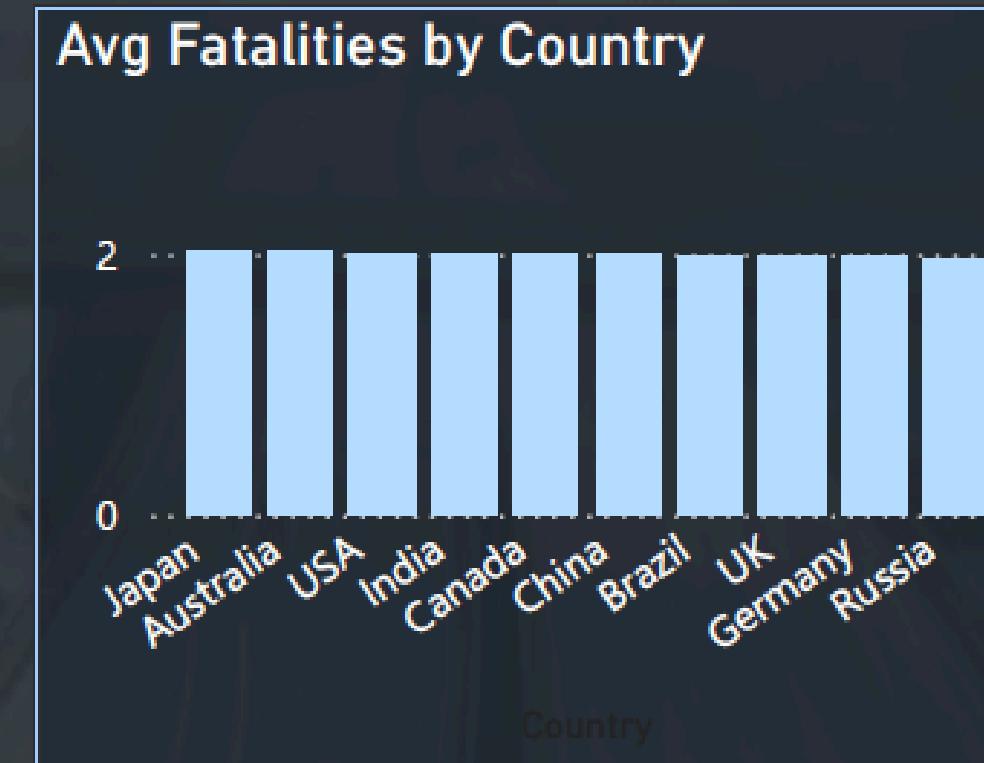
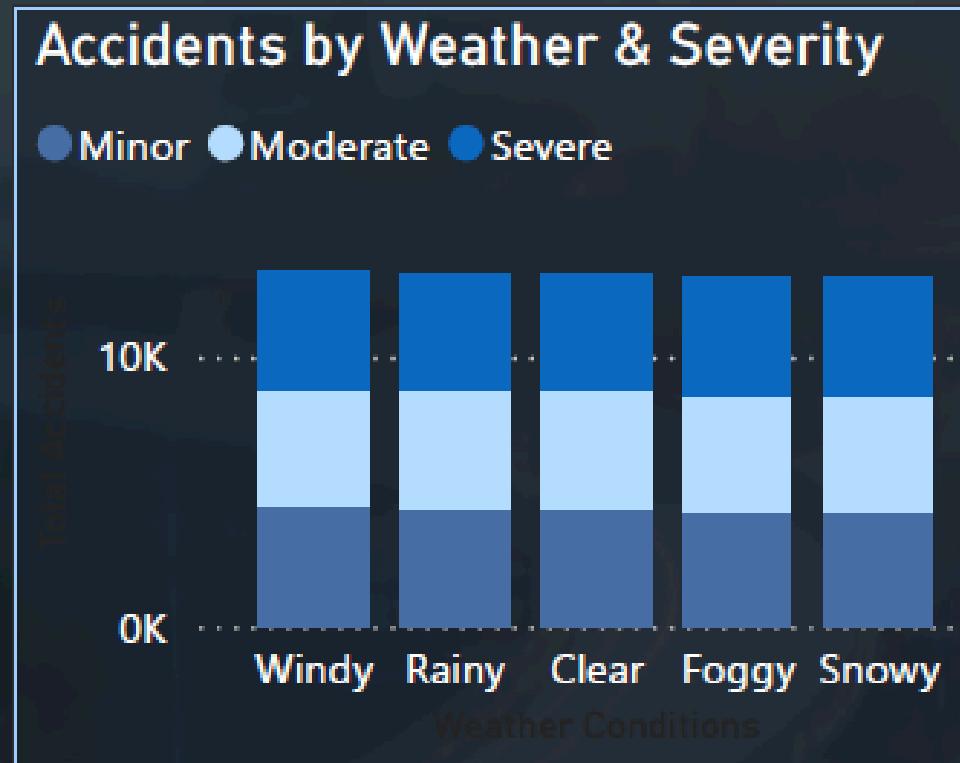


# POWER BI DASHBOARD P2

## Accident Factors

page 1

Page 3



Accident Cause	Minor	Moderate	Severe	Total
Weather	4421	4246	4305	12
Speeding	4335	4330	4346	13
Mechanical Failure	4380	4303	4335	13
Drunk Driving	4375	4376	4287	13
Distracted	4260	4382	4457	13
<b>Total</b>	<b>21771</b>	<b>21637</b>	<b>21730</b>	<b>65</b>

Visibility Category	Weather Conditions	Total Accidents
High	Clear	2861
High	Foggy	2842
High	Rainy	2936
High	Snowy	2871
High	Windy	2924
Low	Clear	5828
<b>Total</b>		<b>65138</b>



# POWER BI DASHBOARD P3

## Temporal & Regional Insights

page 1

Page 2

### Accidents & Growth Rate by Country

● Total Accidents ● Accident Growth Rate



### Severity & Fatalities by Time of Day

● Severe Accidents ● Total Fatalities



April

5402

Total Accidents

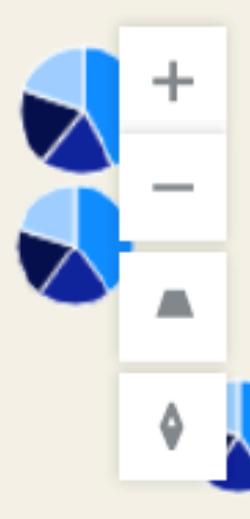
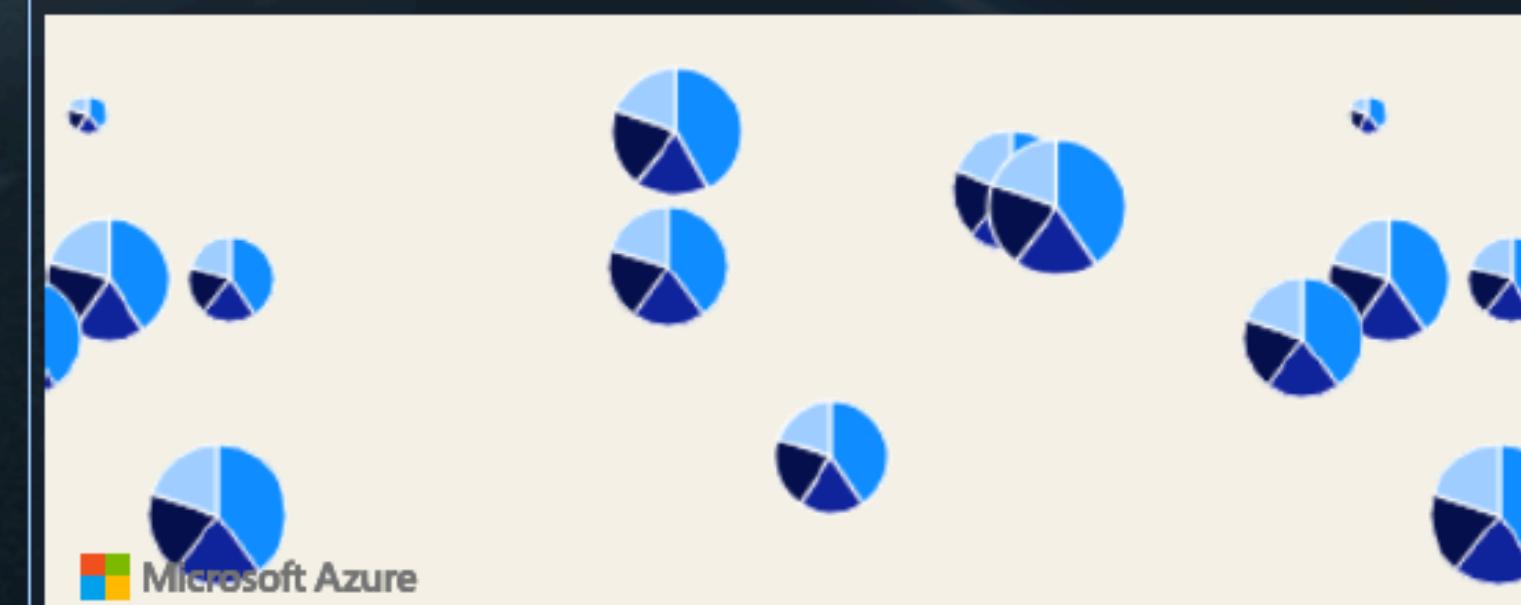
August

5320

Total Accidents

### Accidents by Country & Population Category

● Extremely High ● High ● Low ● Medium



# Environmental & Temporal Factors:

- **Fatalities by Location:** Fatalities are almost evenly split between Rural (65674) and Urban (64414) areas, with rural areas seeing slightly more.
- **Fatalities by Visibility:** A vast majority of fatalities (57846) occur in "Low" visibility conditions, with far fewer in "Medium" or "High" visibility.
- **Time of Day:** "Night" is the most dangerous period, recording the highest number of both severe accidents and total fatalities. The "Morning" appears to be the safest period.
- **Weather Conditions:** The total number of accidents and their corresponding severity levels (Minor, Moderate, Severe) are surprisingly consistent across all weather conditions, including Windy, Rainy, Clear, Foggy, and Snowy.

# Geographical Insights:

- **Average Fatalities:** The average number of fatalities per accident is consistently 2 across all 10 countries analyzed (Japan, Australia, USA, India, Canada, China, Brazil, UK, Germany, Russia).
- **Accident Growth (Australia):** For Australia, while the total number of accidents per year was relatively stable between 2000 and 2013, the accident growth rate was highly volatile, showing a major dip in 2006 followed by a sharp spike in 2007.
- **Global Distribution:** The maps visualize accident data clustered across North America, South America, Europe, Asia, and Australia, broken down by population categories (e.g., Extremely High, High, Low, Medium).

# Accident Distribution Insights:

- Uneven Distribution by Population Category:** Accidents are not evenly spread across population categories. The "Extremely High" category is the largest contributor, accounting for 26,302 (or 40.4%) of all accidents.
- Consistency in Other Categories:** The "High" (12,967), "Low" (12,880), and "Medium" (12,989) categories show a remarkably uniform distribution of accidents.
- Country-Level Consistency:** The total number of accidents is extremely consistent across all 10 countries, with very little variation. Australia has the most (6,634) and Russia the fewest (6,322).

## Fatality distribution:

- The 2:1 Ratio:** The most significant insight is the consistent 2:1 ratio of fatalities to accidents.
- Overall:** The dataset shows 130,088 fatalities from 65,138 total accidents, an average of exactly 2.0 fatalities per incident.
- Per Country:** This 2-to-1 ratio holds true for every single country analyzed. For example, Canada had 6,571 accidents and 13,159 fatalities, while the USA had 6,528 accidents and 13,109 fatalities.

### Combined Insight

Because the "Extremely High" population category accounts for 40.4% of all accidents and the fatality rate is a constant 2.0, you can conclude that this category is also responsible for 40.4% of all fatalities.

# *Thank You*

**Prepared & Developed by:**

# **Abdel-Rahman Sameh**

**Assistant Lecturer | Data Analyst**

