

ROAD ACCIDENT ANALYSIS REPORT (2021–2022)

A Professional Data Analytics Project

Executive Summary

This report presents a comprehensive analysis of road accident data for the years **2021 and 2022**, aimed at identifying high-risk conditions, locations, vehicle categories, and environmental factors contributing to casualties.

The insights from this study support **data-driven decision-making** for government authorities, transport planners, and road safety stakeholders.

The analysis was conducted using **Python, SQL, Power BI, and Excel**, ensuring accurate KPI validation and visually effective dashboards.

Business Problem

Road accidents result in severe social, economic, and public health consequences.

Authorities need **reliable insights** to:

- Reduce accident frequency and severity
- Identify high-risk areas and seasonal patterns
- Improve road infrastructure and lighting
- Allocate emergency and patrol resources efficiently
- Develop targeted awareness and safety campaigns

This project provides the foundations for such data-driven planning.

Dataset Overview

The dataset contains detailed information about each reported accident, including:

- Accident Index
- Date, Day, Time
- Number of Casualties
- Number of Vehicles
- Accident Severity

- Vehicle Type
- Road Type
- Urban/Rural Classification
- Weather & Light Conditions
- Local Authority District

Data was cleaned and standardized before analysis to ensure accuracy.

Methodology

Tools Used

- **Python:** Pandas, NumPy, Matplotlib
- **Power BI:** Interactive dashboards
- **MS Excel:** Initial preprocessing
- **SQL Server:** KPI validation and aggregations
- **Google Colab:** EDA environment

Process Followed

1. **Data Cleaning**
 - Removed duplicates
 - Standardized categories
 - Fixed date formats
 - Handled missing values
 - Added Month and Year columns
2. **Exploratory Data Analysis (Python)**
 - Distribution plots
 - Correlation heatmaps
 - Trend analysis
 - Severity and casualty patterns
3. **SQL-Based KPI Validation**
 - Re-calculated all KPIs visible in Power BI

- Ensured dashboard accuracy

4. Dashboard Creation (Power BI)

- KPIs for accidents & casualties
- Monthly trends
- Severity-based analysis
- Vehicle type & road type breakdown
- Day vs Night comparison
- Urban vs Rural analysis

Key Insights

1. Two-Wheelers Have the Highest Casualty Share

Bike and motorcycle-related accidents form the largest proportion of casualties, indicating a **high-risk vulnerable group**.

2. Night-Time Accidents Show Higher Fatality Rates

Although total accident count is higher during the day, fatality rates are significantly higher at night due to **poor visibility and lighting issues**.

3. Urban Areas Report More Accidents, Rural Areas Report More Fatalities

Urban areas have traffic congestion and higher accident frequency, whereas rural areas lack emergency response speed, leading to more severe outcomes.

4. Single Carriageway Roads Are the Most Dangerous

They contribute the highest proportion of casualties, signaling the need for **infrastructure upgrades**.

5. Seasonal Peaks in November–December

Accidents spike during monsoon and festive periods, likely due to wet surfaces, low visibility, and increased travel.

Recommendations

✓ Improve Two-Wheeler Safety

- Helmet enforcement
- Awareness campaigns
- Dedicated bike lanes

✓ Enhance Night-Time Road Safety

- Street lighting upgrades
- Reflective signs and lane markers
- Speed monitoring systems

✓ Strengthen Rural Emergency Response

- Increase ambulances in rural zones
- Improve first-response infrastructure

✓ Upgrade High-Risk Roads

Especially single-carriageway stretches with high accident rates.

✓ Seasonal Safety Measures

- Monsoon driving awareness
- Festive traffic control strategies

Conclusion

This analysis provides actionable insights to strengthen road safety and support accident prevention strategies.

Using **Python, SQL, and Power BI**, the project delivers a data-driven understanding of accident contributors and highlights priority areas for intervention.

Future enhancements may include:

- Predictive modeling
- Real-time accident monitoring dashboards
- Geo-spatial accident heatmaps

8. Tools & Technologies Used

- Python
 - Power BI
 - MS Excel
 - SQL Server
 - Google Colab
-

Appendix (SQL Queries):

CY Casualties

```
SELECT SUM(Number_of_Casualties) AS CYCasualties  
from road_accident  
where YEAR(incident_date)='2022'
```

OUTPUT

	CYCasualties
1	195737

CY Accidents

```
SELECT COUNT(distinct Accident_Index) AS CY_Accident  
FROM road_accident  
WHERE YEAR(Accident_Date) = '2022';
```

OUTPUT

	CY_Accident
1	144419

CY Fatal Casualties

```
SELECT sum(Number_of_Casualties) AS CY_Fatal_Accident  
FROM road_accident  
WHERE YEAR(Accident_Date) = '2022' AND Accident_Severity='Fatal'
```

OUTPUT

	CY_Fatal_Accident
1	2855

CY Serious Casualties

```
SELECT sum(Number_of_Casualties) AS CY_Serious_Casualties
FROM road_accident
WHERE YEAR(Accident_Date) = '2022' AND Accident_Severity='Serious'
```

OUTPUT

	CY_Serious_Casualties
1	27045

CY Slight Casualties

```
SELECT sum(Number_of_Casualties) AS CY_Slight_Casualties
FROM road_accident
WHERE YEAR(Accident_Date) = '2022' AND Accident_Severity='Slight'
```

OUTPUT

	CY_Slight_Casualties
1	165837

Casualties by Vehicle Type

```
SELECT
CASE
    WHEN Vehicle_Type IN ('Agricultural vehicle') THEN 'Agriculture'
    WHEN Vehicle_Type IN ('Car', 'Taxi/Private hire car') THEN 'Car'
    WHEN Vehicle_Type IN ('Motorcycle 125cc and under', 'Motorcycle 50cc and under',
'Motorcycle over 125cc and up to 500cc', 'Motorcycle over 500cc', 'Pedal cycle') THEN 'Bike'
    WHEN Vehicle_Type IN ('Bus or coach (17 or more pass seats)', 'Minibus (8 - 16
passenger seats)') THEN 'Bus'
    WHEN Vehicle_Type IN ('Van / Goods 3.5 tonnes mgw or under', 'Goods 7.5 tonnes mgw
and over', 'Goods over 3.5t. and under 7.5t') THEN 'Van'
    ELSE 'Others'
END AS Vehicle_group,
```

```

SUM(Number_of_Casualties) AS CY_Casualties
FROM dbo.road_accident
WHERE YEAR(Accident_Date) = 2022
GROUP BY
CASE
    WHEN Vehicle_Type IN ('Agricultural vehicle') THEN 'Agriculture'
    WHEN Vehicle_Type IN ('Car', 'Taxi/Private hire car') THEN 'Car'
    WHEN Vehicle_Type IN ('Motorcycle 125cc and under', 'Motorcycle 50cc and under',
'Motorcycle over 125cc and up to 500cc', 'Motorcycle over 500cc', 'Pedal cycle') THEN 'Bike'
    WHEN Vehicle_Type IN ('Bus or coach (17 or more pass seats)', 'Minibus (8 - 16
passenger seats)') THEN 'Bus'
    WHEN Vehicle_Type IN ('Van / Goods 3.5 tonnes mgw or under', 'Goods 7.5 tonnes mgw
and over', 'Goods over 3.5t. and under 7.5t') THEN 'Van'
    ELSE 'Others'
END;

```

OUTPUT

	Vehicle_group	CY_Casualties
1	Bus	6573
2	Others	1446
3	Car	155804
4	Agriculture	399
5	Bike	15610
6	Van	15905

CY Casualties Monthly Trend

For Year 2022

```

select DateName(MONth,Accident_Date)as Month_Name,
sum(Number_of_Casualties)As 'CY Casualties'
from road_accident
where YEAR(Accident_Date)= 2022
Group by Datename (Month,Accident_Date)

```

OUTPUT

	Month_Name	CY Casualties
1	February	14804
2	June	17230
3	August	16796
4	April	15767
5	May	16775
6	December	13200
7	January	13163
8	September	17500
9	October	18287
10	July	17201
11	November	18439
12	March	16575

For Year 2021

```
select DateName(MONth,Accident_Date)as Month_Name,  
sum(Number_of_Casualties)As 'CY Casualties'  
from road_accident  
where YEAR(Accident_Date)= 2021  
Group by Datename (Month,Accident_Date)
```

OUTPUT

	Month_Name	CY Casualties
1	February	14648
2	June	18728
3	August	18797
4	April	17335
5	May	18852
6	December	18576
7	January	18173
8	September	18456
9	October	20109
10	July	19682
11	November	20975
12	March	17815

Casualties by Road Type

```
select Road_Type,  
sum(Number_of_Casualties)as CY_Casualties from road_accident
```

where Year(Accident_Date)=2022

Group by Road_Type

OUTPUT

	Road_Type	CY_Casualties
1	Single carriageway	144653
2	One way street	3499
3	Roundabout	12683
4	Slip road	2990
5	Dual carriageway	31912

Casualties by Urban/Rural

SELECT

Urban_or_Rural_Area,

CAST(SUM(Number_of_Casualties) AS DECIMAL(10,2)) * 100 /

(SELECT CAST(SUM(Number_of_Casualties) AS DECIMAL(10,2))

FROM road_accident

WHERE YEAR(Accident_Date) = 2022) AS Percentage_Casualties

FROM road_accident

WHERE YEAR(Accident_Date) = 2022

GROUP BY Urban_or_Rural_Area;

OUTPUT

	Urban_or_Rural_Area	Percentage_Casualties
1	Rural	38.0541236455039
2	Urban	61.9458763544960

Casualties by (Day/Night) Light Condition

SELECT

CASE

WHEN Light_Conditions IN ('Daylight') THEN 'Day'

WHEN Light_Conditions IN ('Darkness - lighting unknown', 'Darkness - lights lit', 'Darkness - lights unlit', 'Darkness - no lighting') THEN 'Night'

ELSE 'Unknown'

END AS [Light_condition],

CAST(

SUM(No_of_Casualties) * 100.0 /

(SELECT SUM(Number_of_Casualties)

FROM dbo.road_accident

WHERE YEAR(Accident_Date) = 2022)

```

AS DECIMAL(10,2)) AS CY_Casualties_PCT

FROM dbo.road_accident
WHERE YEAR(Accident_Date) = 2022
GROUP BY
CASE
    WHEN Light_Conditions IN ('Daylight') THEN 'Day'
    WHEN Light_Conditions IN (
        'Darkness - lighting unknown',
        'Darkness - lights lit',
        'Darkness - lights unlit',
        'Darkness - no lighting'
    ) THEN 'Night'
    ELSE 'Unknown'
END;

```

OUTPUT

	Light condition	CY_Casualties_PCT
1	Day	73.84
2	Night	26.16

Top 10 Locations by Casualties

```

Select top 10 Local_Authority_District,sum(Number_of_Casualties)
as Total_casualties
from road_accident
group by Local_Authority_District
order by Total_Casualties DESC

```

OUTPUT

	Local_Authority_District	Total_casualties
1	Birmingham	8611
2	Leeds	5821
3	Bradford	4431
4	Manchester	4366
5	Liverpool	4052
6	Cornwall	3820
7	Sheffield	3737
8	Kirklees	3312
9	County Durham	3295
10	Westminster	3169

SQL queries were used to validate all dashboard KPIs to ensure that the numbers displayed are accurate and consistent with the underlying dataset. This verification confirms that the Power BI dashboard presents correct and reliable insights.

Author:

P LIKHITHA