**Road Traffic Accident Dataset of Addis Ababa City**

Exploratory Data Analysis

**Name: Hamza Saleh**

**Problem Statment**

Every year, road accidents claim countless lives and cause significant economic and emotional losses, making it crucial to understand the factors contributing to these incidents. Between 2017 and 2020, numerous accidents occurred under varying conditions, revealing patterns that could help us prevent future tragedies. This project aims to uncover these patterns by analyzing critical factors such as road and weather conditions, vehicle attributes, and the demographics of those involved. By identifying key risk factors, such as hazardous road surfaces, poor weather, or specific driver behaviors, this analysis will provide actionable insights to improve road safety, reduce accidents, and ultimately save lives.

**Business Impact**

Exploring this data could enable better decision-making regarding road safety measures and infrastructure investments. By analyzing factors such as road conditions, weather, and driver demographics, authorities can identify high-risk areas and implement targeted interventions. This can lead to more effective accident prevention strategies, optimized resource allocation, and improved public safety .

..................................................................................................................................................

### **General Dataset Information**

**Name:** Road Traffic Accident Dataset of Addis Ababa City

**Description:** This dataset contains records of road traffic accidents in Addis Ababa from 2017 to 2020. It includes detailed information on accident causes, locations, times, environmental factors, and demographics. The dataset was collected as part of a Master's research project to analyze traffic accident patterns.

**Dataset Details:** 12,316 Rows & 32 Columns

**Size:** 4,300KB (4.3MB)

**Source**: [Kaggle Dataset - Road Traffic Accidents](https://www.kaggle.com/datasets/saurabhshahane/road-traffic-accidents?select=RTA+Dataset.csv)

**Dataset Creator:** Tarikwa Tesfa Bedane

.................................................................................................................................................

**Columns of interest**

The problem statement outlines the essential factors contributing to road accidents and identifying risk patterns. The key features in question that can be found within the dataset are:

* **Time:** Indicates the specific time of the accident, which helps analyze accident patterns during different hours.
* **Day\_of\_week:** Shows the day the accident occurred, useful for understanding weekday vs. weekend trends.
* **Service\_year\_of\_vehicle:** Represents the age of the vehicle involved, helping to evaluate the impact of vehicle maintenance on accidents.
* **Area\_accident\_occured:** Identifies the location of the accident, allowing for geographic analysis of high-risk areas.
* **Lanes\_or\_Medians:** Details the number and type of lanes or medians, aiding in understanding road design factors.
* **Road\_allignment:** Describes the alignment of the road (e.g., straight, curved), which is critical for assessing road safety.
* **Types\_of\_Junction:** Specifies the type of junction where the accident occurred, providing insights into intersection safety.
* **Road\_surface\_type:** Indicates the type of road surface, helping analyze its impact on vehicle control.
* **Road\_surface\_conditions:** Highlights the condition of the road, such as wet or dry, to determine its impact on accidents.
* **Light\_conditions:** Captures the lighting at the time of the accident, such as daylight or nighttime, to assess visibility factors.
* **Weather\_conditions:** Details weather conditions like rain or fog, which are critical for assessing environmental risks.
* **Age\_band\_of\_driver:** Provides the age group of the driver to analyze demographic trends in accidents.
* **Accident\_severity:** Represents the severity of the accident, the primary target variable to evaluate risk factors.
* **Cause\_of\_accident:** Identifies the root cause of the accident, a key feature for preventive analysis.

.................................................................................................................................................

### **Target Variables**

The key variables to examine in the dataset, based on the problem statement, are:

1. **Primary Target Variable:**

**Accident\_severity:** This is the main focus of the analysis, as it helps evaluate the severity of road accidents and identify high-risk patterns.

1. **Supporting Features (Independent Variables):**

**Road\_surface\_conditions:** To examine how road conditions contribute to accidents.

**Weather\_conditions:** To analyze the impact of environmental factors on accidents.

**Light\_conditions:** To assess how visibility plays a role in accident occurrences.

**Area\_accident\_occured:** To identify geographic accident hotspots.

**Age\_band\_of\_driver:** To analyze the demographic trends contributing to accidents.

**Cause\_of\_accident:** To uncover the primary reasons behind accidents.

**Time** and **Day\_of\_week:** To study temporal patterns of accidents, such as peak hours or weekends.

### **Column Types in the Dataset**

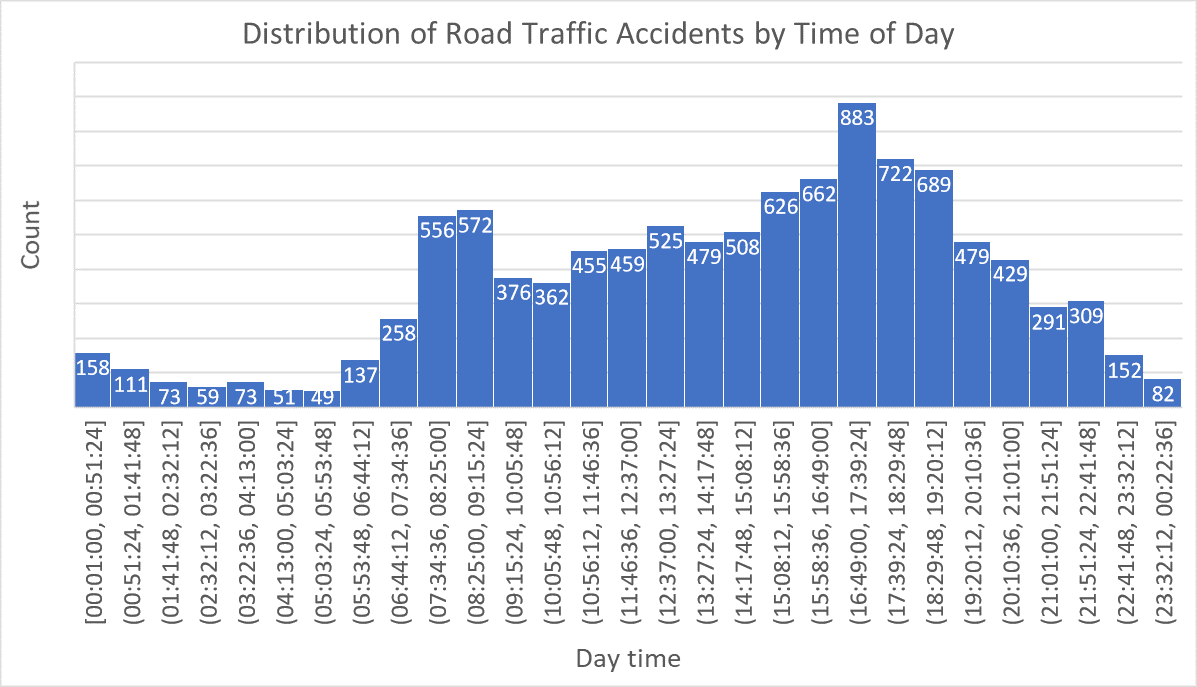
* **Numerical Columns:**
  + Number\_of\_vehicles\_involved, Number\_of\_casualties.
* **Textual Columns:**
  + Area\_accident\_occured, Cause\_of\_accident, Type\_of\_collision, Pedestrian\_movement.
* **Categorical Columns:**
  + Day\_of\_week, Age\_band\_of\_driver, Sex\_of\_driver, Vehicle\_driver\_relation.

Missing Values Summary:

* **Total Missing Entries:** 3,581
* **Overall Percentage of Missing Data:** ~3.24%

....................................................................................................................................................

**Histogram Visualization**

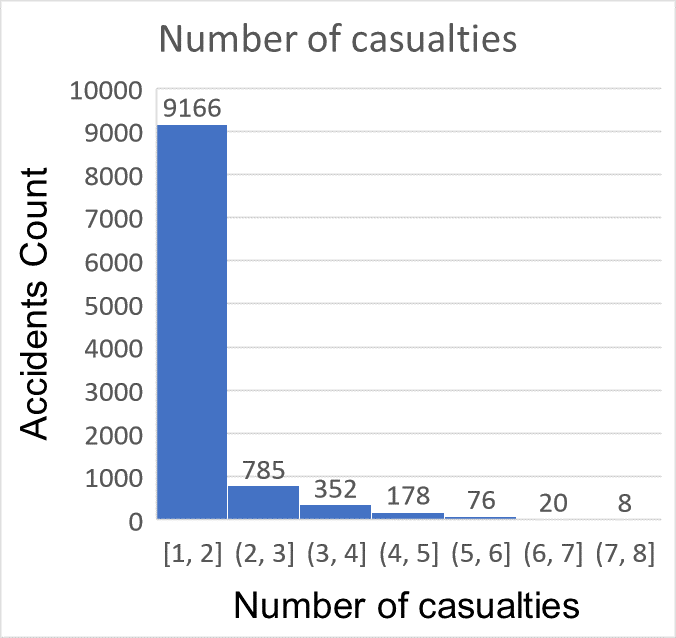
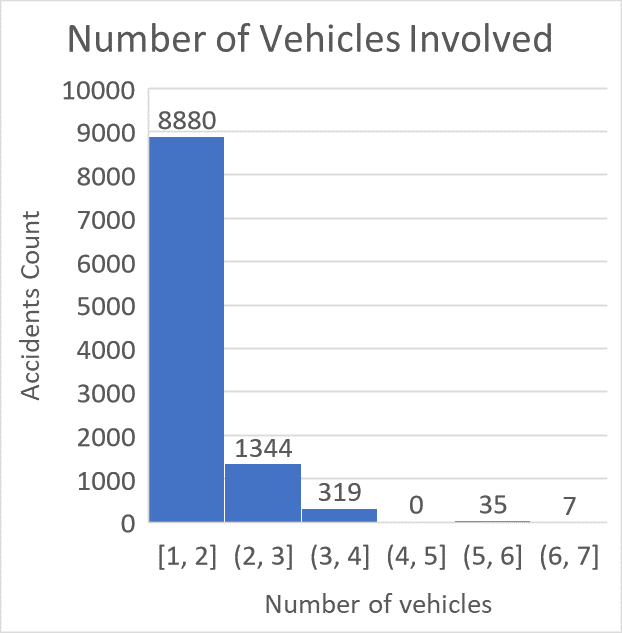


The previous Histogram shows clear peaks during typical rush hour periods:

1. Morning peak: Around 8:00-9:00 (556-572 accidents)
2. Major evening peak: Around 17:00-18:00 (883 accidents - highest point .

The distribution is bimodal, with distinct peaks in the morning and evening rush hours, showing a stronger right skew as accident counts rise more steeply from morning to evening peak (883), followed by a gradual decline through late night hours.

……………………………………………………………………………………………………..



Both distributions are highly right-skewed

Some statistics :

* Sum of Number of vehicles\_involved: 21,613
* Sum of Number of casualties: 16,467
* Max of Number of casualties2: 8
* Max of Number of vehicles involved : 7

......................................................................................................................................

**Subgroup sizes :**

|  |  |
| --- | --- |
| **Row Labels** | **Count of Sex\_of\_driver** |
| Female | 597 |
| Male | 9838 |
| Unknown | 150 |
| **Grand Total** | **10585** |

This indicates that the majority of the drivers involved in the accidents are male, with a relatively small proportion of female drivers

|  |  |
| --- | --- |
| **Row Labels** | **Count of Educational\_level** |
| Above high school | 338 |
| Elementary school | 1925 |
| High school | 1021 |
| Illiterate | 41 |
| Junior high school | 7027 |
| Unknown | 74 |
| Writing & reading | 159 |
| **Grand Total** | **10585** |

the majority of individuals involved in the accidents have completed **Junior high school**,

|  |  |
| --- | --- |
| **Row Labels** | **Count of Accident\_severity** |
| Fatal injury | 131 |
| Serious Injury | 1489 |
| Slight Injury | 8965 |
| **Grand Total** | **10585** |

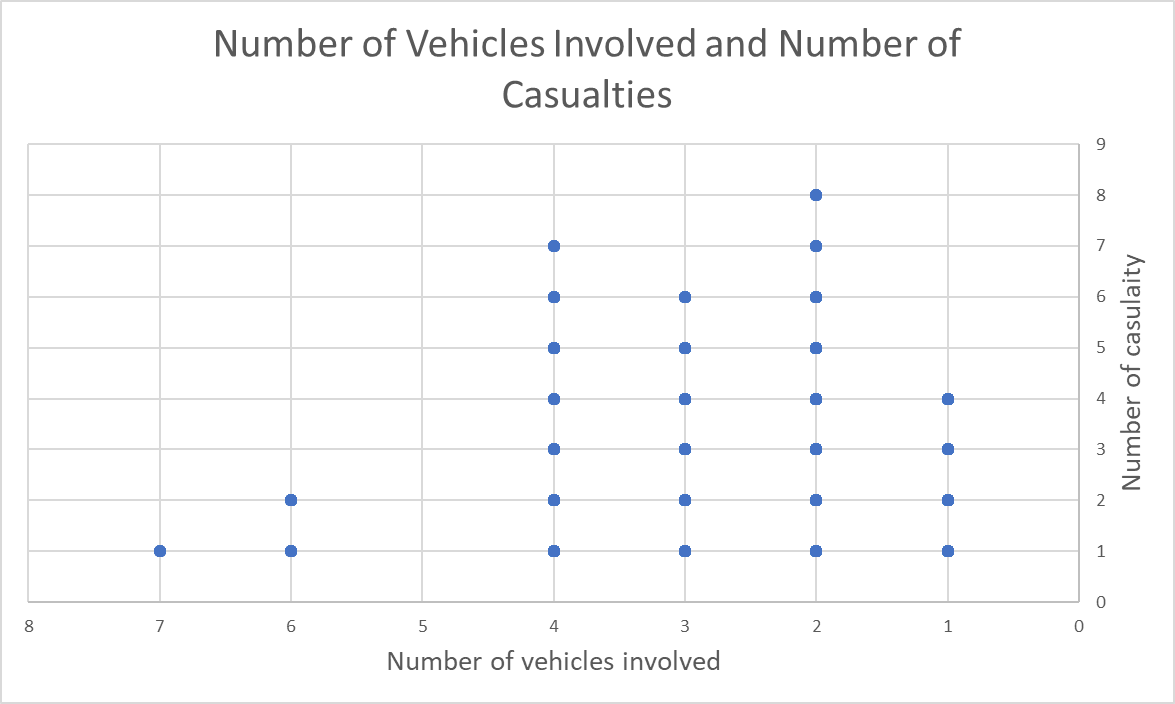
the majority of accidents involve **slight injuries**, followed by **serious injuries**, and a very small number of **fatal injuries**

|  |  |
| --- | --- |
| **Row Labels** | **Count of Age\_band\_of\_driver** |
| 18-30 | 3660 |
| 31-50 | 3537 |
| Over 51 | 1344 |
| Under 18 | 717 |
| Unknown | 1327 |
| **Grand Total** | **10585** |

……………………………………………………………………………………………………………………………………………………………….

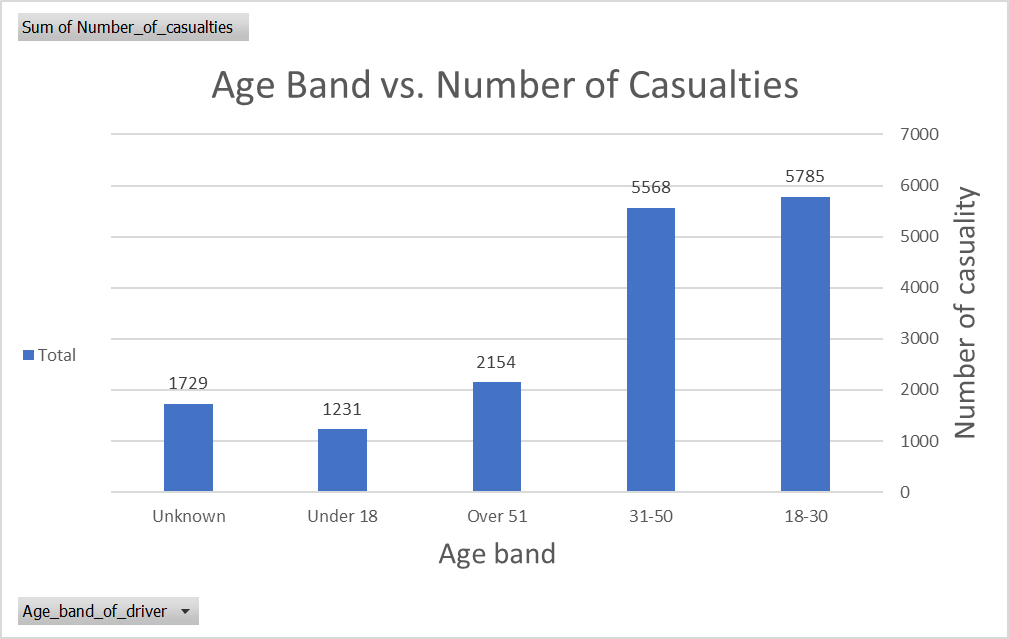
**Analysis supported by charts**

**Relationship Between Number of Vehicles Involved and Number of Casualties**



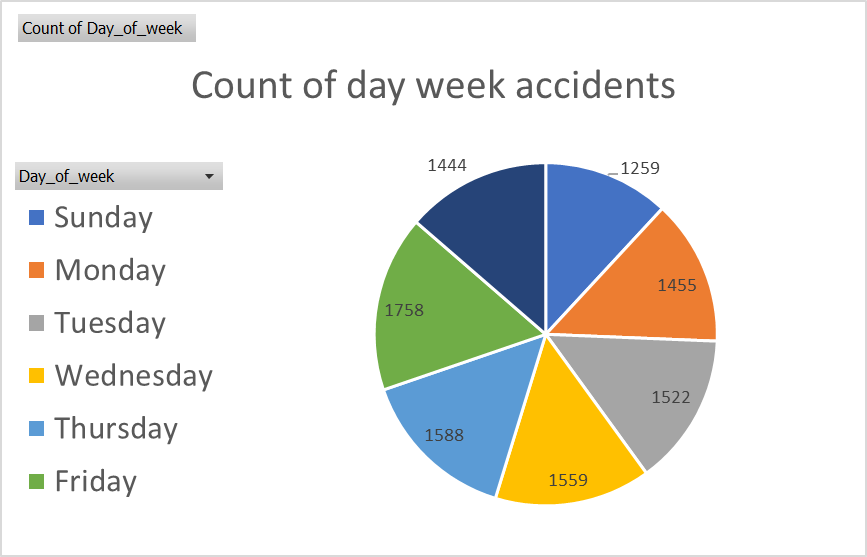
From the previous chart , it can be concluded that the more vehicles involved in an accident, the higher the number of casualties . However there is some variability in the data. For instance, certain accidents involving with seven vehicles resulted in only one casualty, in general there appears to be a positive correlation estimated at approximately 0.6 ,but other factors must be analyzed to fully understand the underlying causes.

**Age Band vs. Number of Casualties**



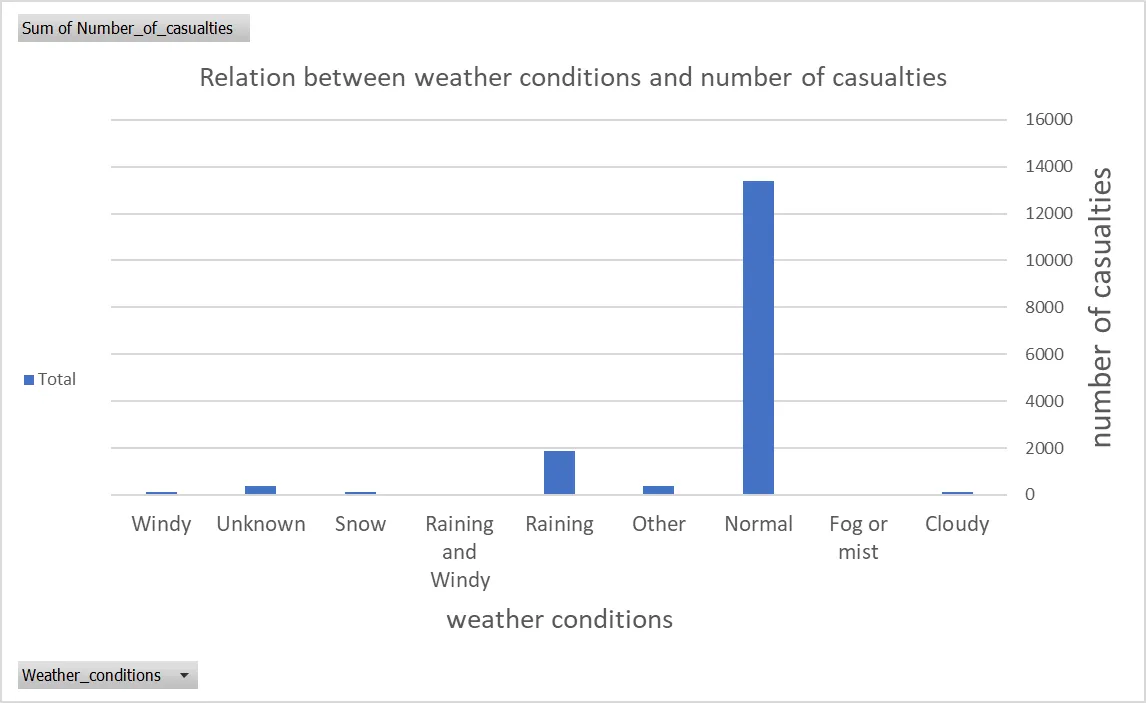
The chart indicates that younger and middle-aged drivers (18-50) are most affected by casualties, with 5785 in the "18-30" age band and 5568 in the "31-50" age band, highlighting the need for targeted safety measures for these groups. In comparison, the "Under 18" age band has the lowest number of casualties at 1231, likely due to fewer young drivers, while the "Over 51" group has 2154 casualties, indicating that older drivers and those with unrecorded ages are less frequently involved in incidents.

**Count of Day week accidents**



The pie chart indicates that Friday, a weekend in Addis Ababa, has the highest number of accidents at 1758, followed by Thursday with 1588 , this could be attributed to a few factors Being a weekend friday often sees increased social and recreational activities, which can lead to higher traffic volumes and possibly riskier driving behaviors , These observations suggest a need for heightened road safety awareness and measures, especially on weekends.

**Weather Conditions vs. Number of Casualties**

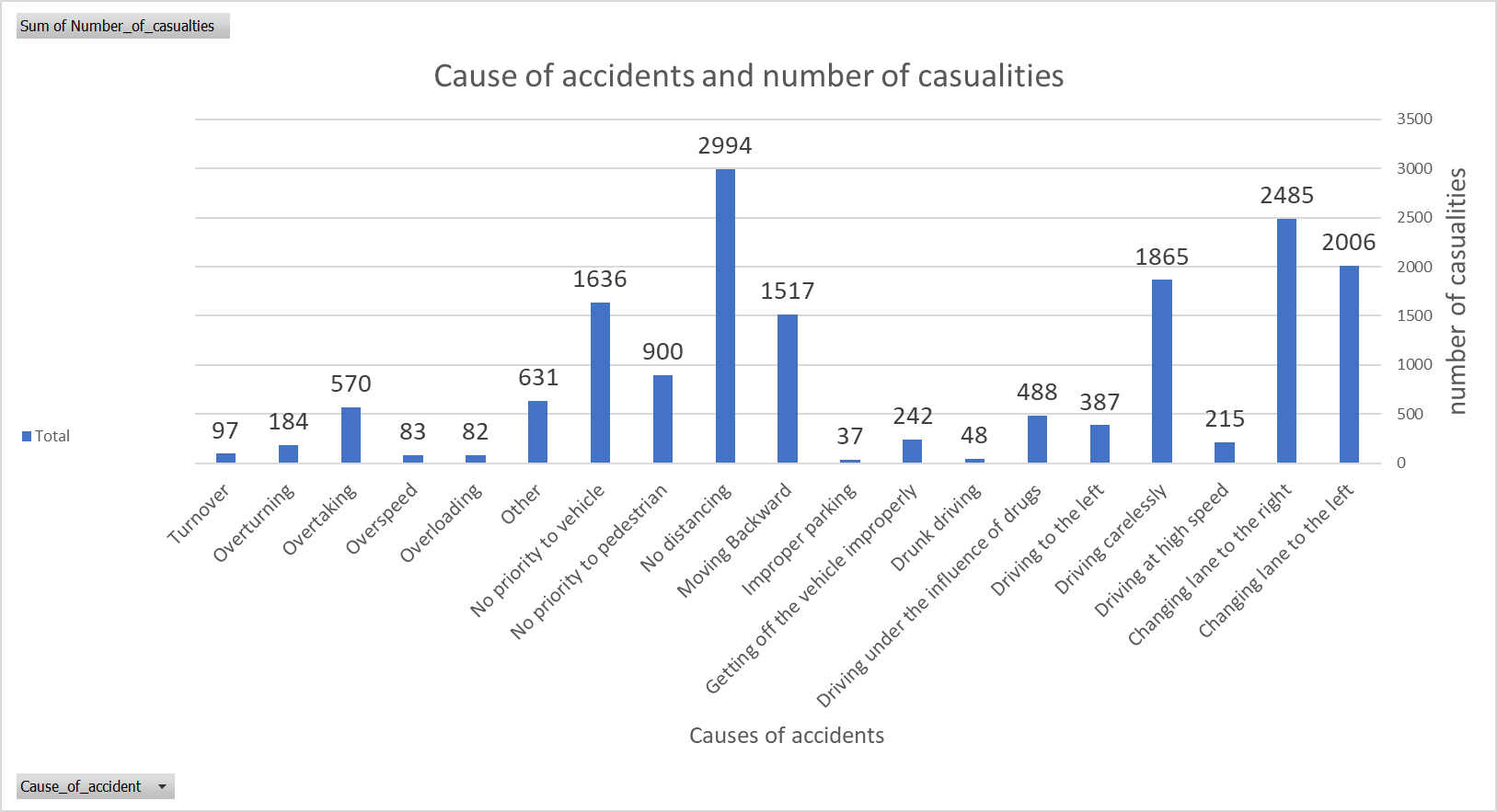


The chart shows the relationship between weather conditions and the number of casualties in road traffic accidents. The data indicates that the weather condition with the highest number of casualties is "Normal", with over 14,000 casualties.

The next highest casualty counts are associated with "Raining and Windy" (around 4,500 casualties) and "Cloudy" (around 1,800 casualties) weather conditions.

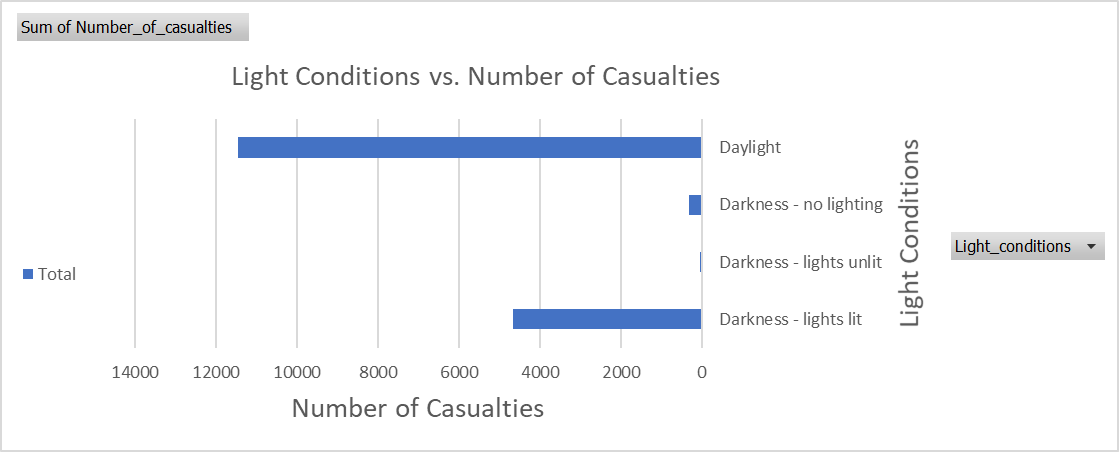
The data suggests that normal, clear weather conditions result in the highest number of casualties, likely due to higher traffic volumes and speeds during ideal driving conditions. Adverse weather like rain and wind also contribute to a significant number of casualties, though not as high as normal conditions.

**Cause of accidents and number of casualities**



The causes with the highest number of casualties are "No distancing" with 2994, "Changing lane to the right" with 2485, and "Changing lane to the left" with 2006. This suggests that issues related to maintaining safe distances and changing lanes are major contributors to accidents.. These insights highlight the need for focused road safety measures and awareness campaigns targeting the most prevalent causes of accidents.

**Light Conditions vs. Number of Casualties**

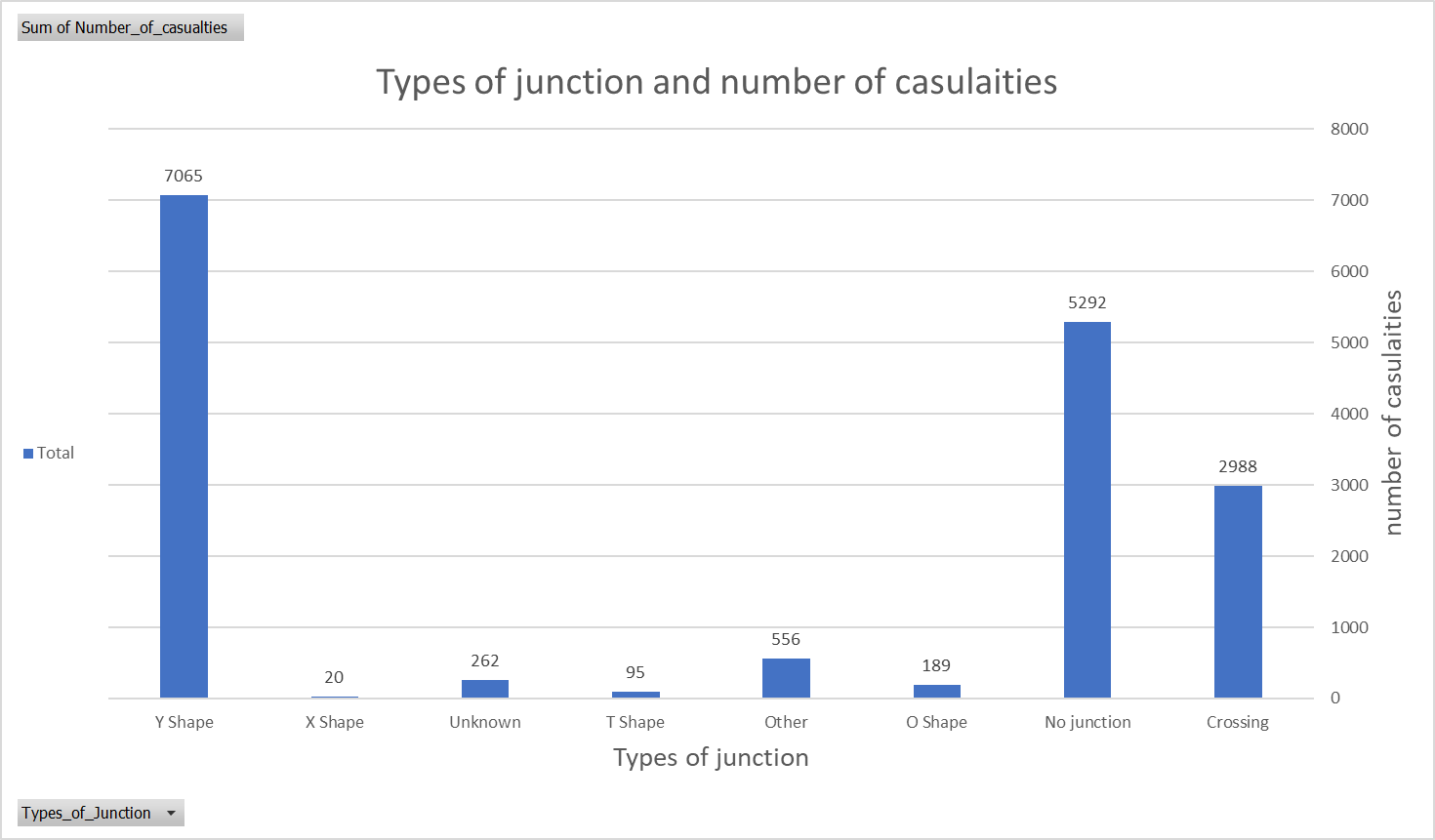


Some insights for the previous chart :

* Daylight conditions account for the highest number of casualties, around 14,000.
* "Darkness - lights lit" has the next highest casualties, around 3,000.
* "Darkness - no lighting" has the lowest casualty count, around 1,500.

This suggests that better lighting conditions, even at night, can help reduce the severity of road traffic accidents in terms of casualties. Driving in complete darkness without any lighting appears to be the most hazardous scenario.

**Types of junction and number of casulaities**



The chart presents a comprehensive analysis of the relationship between types of road junctions and the number of resulting casualties in traffic accidents. The data indicates that "Y Shape" junctions have the highest casualty count at over 7,000, significantly exceeding other junction types .

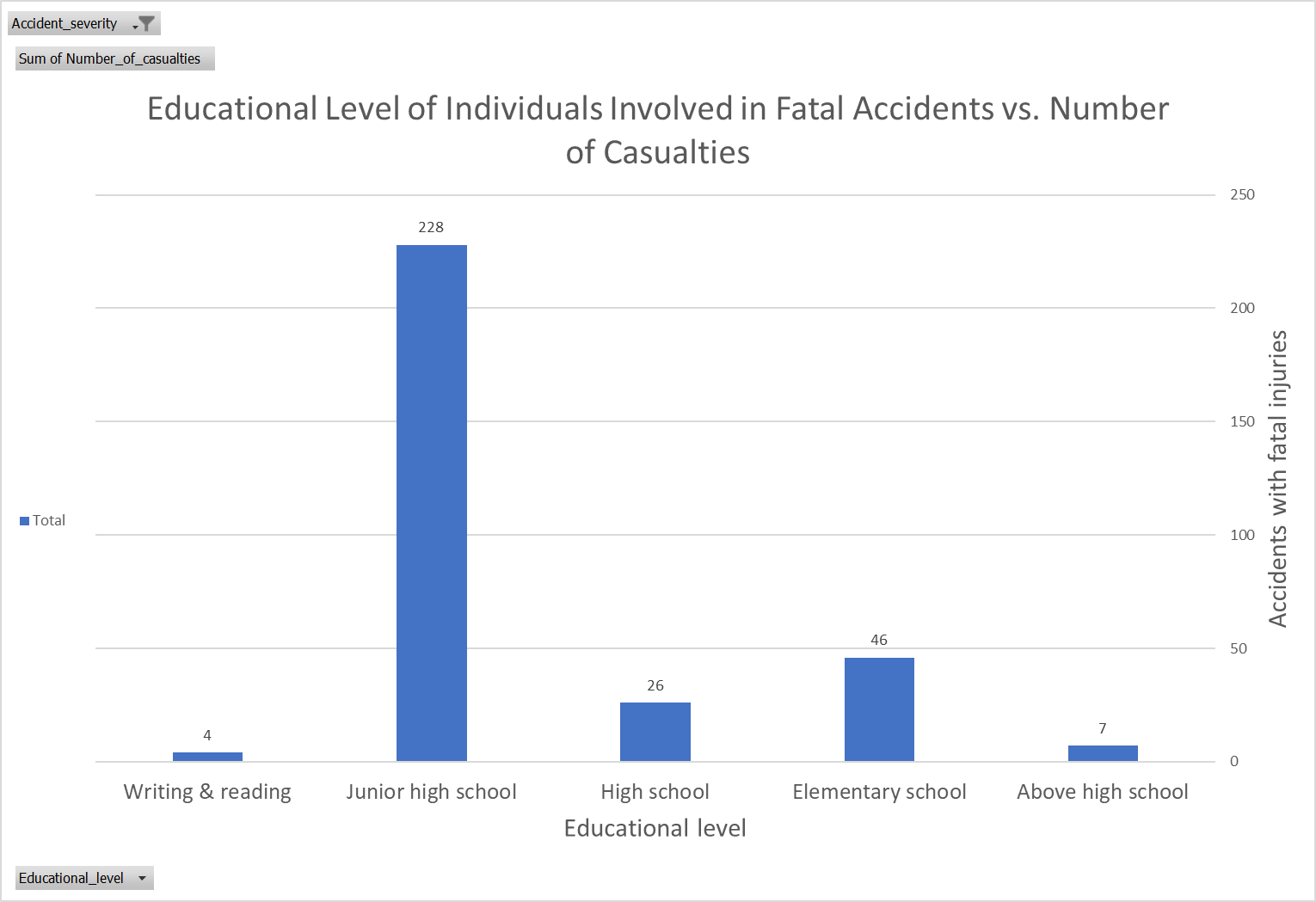
**And the Y-shape junction looks like this :**



From the picture, the "Y Shape" junction has a distinct geometry with three converging roads forming a "Y" configuration. This is the type of junction with the highest number of casualties according to the previous chart, showing over 7,000 casualties associated with "Y Shape" junctions.

We need to take some action to reduce Y-shape accidents from happening , key measures include improving signage and road markings, implementing traffic calming features, considering junction redesign, enhancing visibility, and optimizing traffic management. The goal is to simplify the junction's complexity and provide clear guidance to drivers.

**Educational Level of Individuals Involved in Fatal Accidents vs. Number of Casualties**



The chart shows a clear correlation between the educational level of individuals involved in fatal accidents and the number of casualties. Individuals with a "Junior high school" education level account for the highest number of casualties at 228, significantly exceeding other education levels. This is followed by "High school" at 26 casualties and "Elementary school" at 46 casualties. The chart suggests that education level may be a factor in accident severity, with less educated individuals potentially exhibiting riskier driving behaviors or making poorer decisions that lead to more severe crashes. This information could inform targeted road safety campaigns and educational initiatives to improve driver training and awareness, especially for younger or less educated demographics.

### **Conclusion**

The analysis of traffic accidents in Addis Ababa reveals several critical insights that can guide road safety measures and policy development:

**Demographic Insights:**

* + Male drivers are disproportionately involved in traffic accidents, comprising over 93% of the cases. This highlights the need for targeted awareness campaigns focusing on male drivers.
  + Drivers in the age bands **18–30** and **31–50** are most frequently involved in accidents, collectively accounting for over 68% of the total casualties. These groups should be prioritized in safety measures, such as defensive driving training and awareness programs.

**Education and Accident Severity:**

* + Individuals with a **Junior high school education** represent the largest proportion of casualties in both minor and fatal accidents. This suggests that educational attainment may influence driving behavior, with lower education levels potentially linked to riskier driving practices. Tailored driver education programs could address this gap.

**Accident Severity:**

* + The majority of accidents result in **slight injuries** (84.7%), with **serious injuries** (14%) and **fatal injuries** (1.2%) being less common. While fatal accidents are fewer, their impact warrants further investigation to prevent loss of life.

**Temporal Patterns:**

* + Fridays, coinciding with weekend activities, see the highest number of accidents, followed by Thursdays. This indicates the need for heightened enforcement and safety measures during weekends.
  + Daylight conditions account for the highest casualties, suggesting that traffic volume and speed during clear weather play significant roles in accidents.

**Environmental and Road Conditions:**

* + Normal weather conditions are associated with the highest number of casualties, likely due to increased traffic during favorable conditions. However, adverse weather conditions like rain and wind also contribute significantly, emphasizing the need for improved road infrastructure and driver training for adverse weather scenarios.
  + Poor lighting conditions, particularly **darkness without lighting**, are associated with the most hazardous outcomes. This underscores the importance of better street lighting and reflective road signs.

**Behavioral and Road Usage Patterns:**

* + The primary causes of accidents are **no distancing** and **unsafe lane changes**, together contributing to over 7,400 casualties. Strict enforcement of traffic rules and awareness campaigns on maintaining safe distances and proper lane discipline are essential.
  + "Y Shape" junctions are the most dangerous, with the highest number of casualties among all junction types. Improved signage, traffic signals, and redesigning such junctions can mitigate this risk.

### Recommendations:

* **Targeted Safety Campaigns:** Focus on male drivers, younger age groups, and individuals with lower educational attainment to address risky driving behaviors and improve road safety awareness.
* **Enhanced Law Enforcement:** Increase traffic enforcement on weekends, particularly Fridays, to reduce accidents during peak times.
* **Road Infrastructure Improvements:** Prioritize better lighting in poorly lit areas and redesign hazardous junctions to reduce accident risk.
* **Driver Training Programs:** Implement educational initiatives emphasizing safe driving behaviors, especially regarding lane discipline, maintaining safe distances, and driving in adverse weather.
* **Policy Formulation:** Use these insights to inform citywide transportation policies aimed at reducing accidents, such as stricter licensing requirements or incentives for safe driving.

By addressing the patterns and risk factors identified in this analysis, Addis Ababa can take significant steps toward reducing traffic accidents and improving road safety for all.