**PROJECT ON FRANCE ACCIDENTS**

**EXECUTIVE SUMMARY**

This project is aimed at extracting specific desired information from a dataset containing information about the recorded accidents in France between the year 2005 and 2016. The following questions will be answered in the course of this work:

1. Exploration based on date of accidents

* Is the number of Accidents per year decreasing? (from 2005 to 2016)
* Which months have higher frequency of Accidents?
* Which Day-of-the-Month is most safe to drive?
* Time series of all accidents from 2005 to 2016
* Time series for all accidents in each year

1. Exploration based on roads where accidents occurred:

* Which types of roads are high risk?
* Which type of road gradient is high risk?

1. Exploration based on people involved in the accidents:

* What was the condition of the people after the accident?
* What was the age distribution of the people involved?
* What was the sex distribution of the people involved?

1. Exploration based on use of safety equipment:

* What was the distribution of Safety Equipment used?
* Did use of Safety Eqipment impact condition of people after the accident?

**INTRODUCTION**

There’s been a period of increased accidents in France and it became a thing of worry. Th country knowing it is a time bomb waiting for detonation, had to engage EDA analysis to track the accidents and causes thereof so as to implement control measures.

**METHODOLOGY**

Python programming language was used throughout the work which was done using Jupyter notebook.

Different EDA techniques were deployed in order to achieve the desired goal of the questionnaire, but first data mining and cleaning was deployed to enable accuracy of facts pulled from the given datasets.

Below are word guide to aid understanding of the dataset here within:

*'''#(Num\_Acc : Accident ID*

*jour : Day of the accident*

*mois : Month of the accident*

*an : Year of the accident*

*hrmn : Time of the accident in hour and minutes (hhmm)*

*lum : Lighting : lighting conditions in which the accident occurred*

*1 - Full day*

*2 - Twilight or dawn*

*3 - Night without public lighting*

*4 - Night with public lighting not lit*

*5 - Night with public lighting on*

*dep : Departmeent : INSEE Code (National Institute of Statistics and Economic Studies) of the departmeent followed*

*by a 0 (201 Corse-du-Sud - 202 Haute-Corse)*

*com : Municipality: The commune number is a code given by INSEE. The code has 3 numbers set to the right.*

*Localisation :*

*1 - Out of agglomeration*

*2 - In built-up areas*

*int : Type of Intersection :*

*1 - Out of intersection*

*2 - Intersection in X*

*3 - Intersection in T*

*4 - Intersection in Y*

*5 - Intersection with more than 4 branches*

*6 - Giratory*

*7 - Place*

*8 - Level crossing*

*9 - Other intersection*

*atm : Atmospheric conditions:*

*1 - Normal*

*2 - Light rain*

*3 - Heavy rain*

*4 - Snow - hail*

*5 - Fog - smoke*

*6 - Strong wind - storm*

*7 - Dazzling weather*

*8 - Cloudy weather*

*9 - Other*

*col : Type of collision:*

*1 - Two vehicles - frontal*

*2 - Two vehicles - from the rear*

*3 - Two vehicles - by the side*

*4 - Three vehicles and more - in chain*

*5 - Three or more vehicles - multiple collisions*

*6 - Other collision*

*7 - Without collision*

*adr : Postal address: variable filled in for accidents occurring in built-up areas*

*gps : GPS coding: 1 originator character:*

*M = Métropole*

*A = Antilles (Martinique or Guadeloupe)*

*G = Guyane*

*R = Réunion*

*Y = Mayotte*

*Geographic coordinates in decimal degrees:*

*lat : Latitude*

*long : Longitude*

*Places:*

*Num\_Acc : Accident ID*

*catr : Category of road:*

*1 - Highway*

*2 - National Road*

*3 - Departmental Road*

*4 - Communal Way*

*5 - Off public network*

*6 - Parking lot open to public traffic*

*9 - other*

*voie : Road Number*

*V1: Numeric index of the route number (example: 2 bis, 3 ter etc.)*

*V2: Letter alphanumeric index of the road*

*circ: Traffic regime:*

*1 - One way*

*2 - Bidirectional*

*3 - Separated carriageways*

*4 - With variable assignment channels*

*nbv: Total number of traffic lanes*

*vosp: Indicates the existence of a reserved lane, regardless of whether or not the accident occurs on that lane.*

*1 - Bike path*

*2 - Cycle Bank*

*3 - Reserved channel*

*Prof: Longitudinal profile describes the gradient of the road at the accident site*

*1 - Dish*

*2 - Slope*

*3 - Hilltop*

*4- Hill bottom*

*pr: Home PR number (upstream terminal number)*

*pr1: Distance in meters to the PR (relative to the upstream terminal)*

*plan: Drawing in plan:*

*1 - Straight part*

*2 - Curved on the left*

*3 - Curved right*

*4 - In "S"*

*lartpc: Central solid land width (TPC) if there is*

*larrout: Width of the roadway assigned to vehicle traffic are not included the emergency stop strips,*

*CPRs and parking spaces*

*surf: surface condition*

*1 - normal*

*2 - wet*

*3 - puddles*

*4 - flooded*

*5 - snow*

*6 - mud*

*7 - icy*

*8 - fat - oil*

*9 - other*

*infra: Development - Infrastructure:*

*1 - Underground - tunnel*

*2 - Bridge - autopont*

*3 - Exchanger or connection brace*

*4 - Railway*

*5 - Carrefour arranged*

*6 - Pedestrian area*

*7 - Toll zone*

*situ: Situation of the accident:*

*1 - On the road*

*2 - On emergency stop band*

*3 - On the verge*

*4 - On the sidewalk*

*5 - On bike path*

*env1: school point: near a school*

*USERS:*

*Acc\_number: Accident identifier.*

*Num\_Veh: Identification of the vehicle taken back for each user occupying this vehicle (including pedestrians who are*

*attached to the vehicles that hit them)*

*place: Allows to locate the place occupied in the vehicle by the user at the time of the accident*

*catu: User category:*

*1 - Driver*

*2 - Passenger*

*3 - Pedestrian*

*4 - Pedestrian in rollerblade or scooter*

*grav: Severity of the accident: The injured users are classified into three categories of victims plus the uninjured*

*1 - Unscathed*

*2 - Killed*

*3 - Hospitalized wounded*

*4 - Light injury*

*sex: Sex of the user*

*1 - Male*

*2 - Female*

*Year\_on: Year of birth of the user*

*trip: Reason for traveling at the time of the accident:*

*1 - Home - work*

*2 - Home - school*

*3 - Shopping - Shopping*

*4 - Professional use*

*5 - Promenade - leisure*

*9 - Other*

*secu: on 2 characters:*

*the first concerns the existence of a safety equipment*

*1 - Belt*

*2 - Helmet*

*3 - Children's device*

*4 - Reflective equipment*

*9 - Other*

*the second is the use of Safety Equipment*

*1 - Yes*

*2 - No*

*3 - Not determinable*

*locp: Location of the pedestrian:*

*On pavement:*

*1 - A + 50 m from the pedestrian crossing*

*2 - A - 50 m from the pedestrian crossing*

*On pedestrian crossing:*

*3 - Without light signaling*

*4 - With light signaling*

*Various:*

*5 - On the sidewalk*

*6 - On the verge*

*7 - On refuge or BAU*

*8 - On against aisle*

*actp: Action of the pedestrian:*

*Moving*

*0 - not specified or not applicable*

*1 - Meaning bumping vehicle*

*2 - Opposite direction of the vehicle*

*Various*

*3 - Crossing*

*4 - Masked*

*5 - Playing - running*

*6 - With animal*

*9 - Other*

*etatp: This variable is used to specify whether the injured pedestrian was alone or not*

*1 - Only*

*2 - Accompanied*

*3 - In a group*

*VEHICLES:*

*Num\_Acc*

*Accident ID*

*Num\_Veh*

*Identification of the vehicle taken back for each user occupying this vehicle (including pedestrians who are*

*attached to vehicles that hit them) - alphanumeric code*

*GP*

*Flow direction :*

*1 - PK or PR or increasing postal address number*

*2 - PK or PR or descending postal address number*

*CATV*

*Category of vehicle:*

*01 - Bicycle*

*02 - Moped <50cm3*

*03 - Cart (Quadricycle with bodied motor) (formerly "cart or motor tricycle")*

*04 - Not used since 2006 (registered scooter)*

*05 - Not used since 2006 (motorcycle)*

*06 - Not used since 2006 (side-car)*

*07 - VL only*

*08 - Not used category (VL + caravan)*

*09 - Not used category (VL + trailer)*

*10 - VU only 1,5T <= GVW <= 3,5T with or without trailer (formerly VU only 1,5T <= GVW <= 3,5T)*

*11 - Most used since 2006 (VU (10) + caravan)*

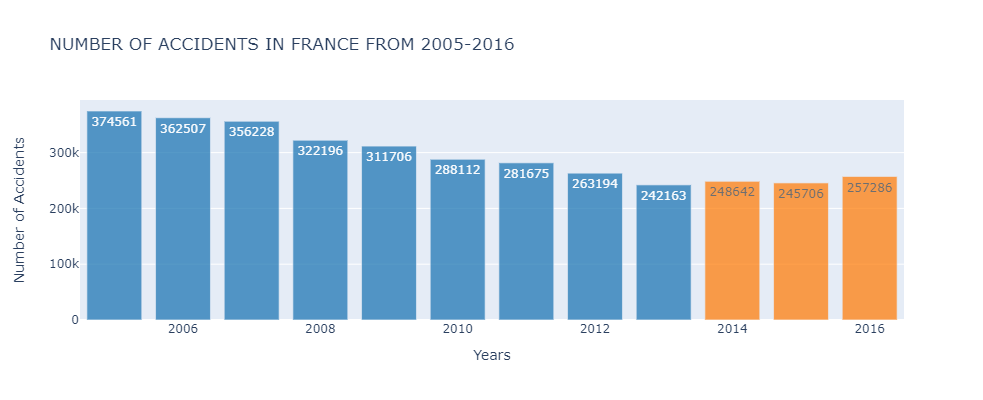
*12 - Most used since 2006 (VU (10) + trailer)*

*13 - PL only 3,5T)'''causes. Low blood sugar may cause similar symptoms*

**RESULTS**

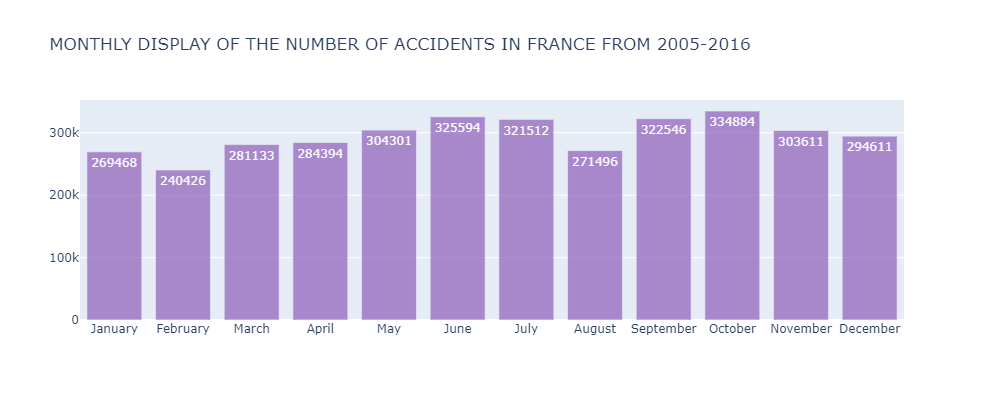
1. **Exploration based on date of accidents:**

* Is the number of Accidents per year decreasing ? (from 2005 to 2016)
* Which months have higher frequency of Accidents ?
* Which Day-of-the-Month is most safe to drive ?
* Time series of all accidents from 2005 to 2016
* Time series for all accidents in each year
* Exploration based on roads where accidents occured

****

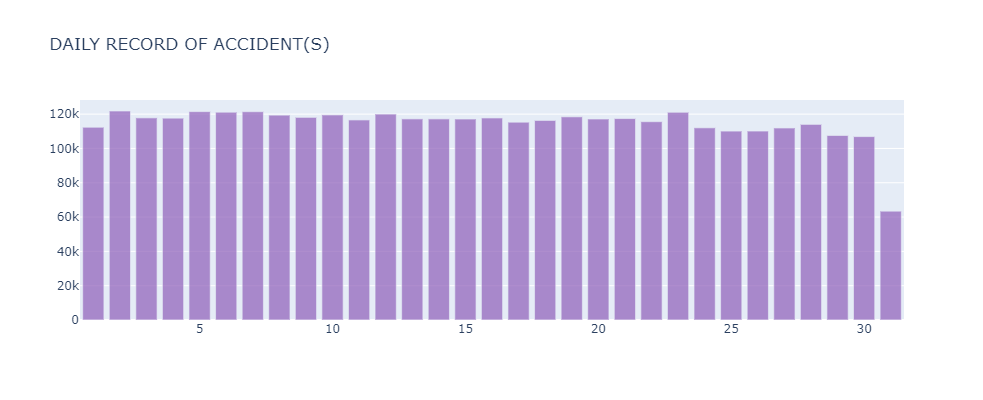
**NOTES:**

* Shows that the number of accidents decreased on a steady rate from 2005 to 2013, and started tending to rise from 2014 to 2016
* Also shows that 2005 has the highest number of accidents according to this dataset

****

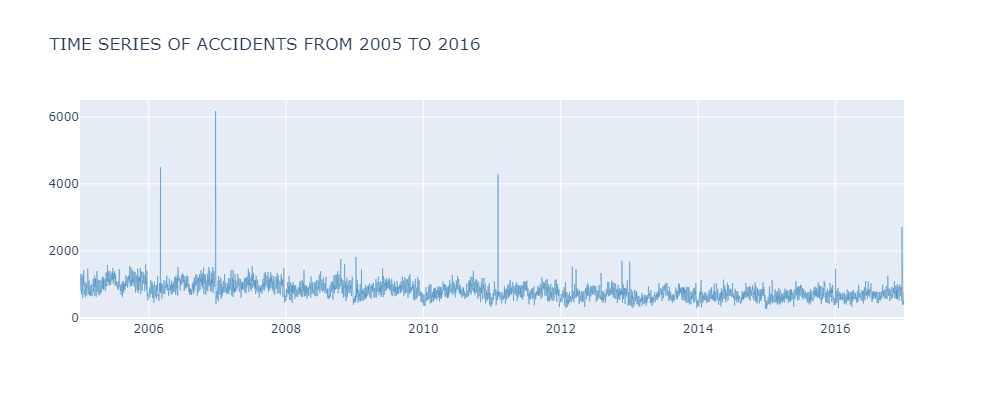
**NOTES:**

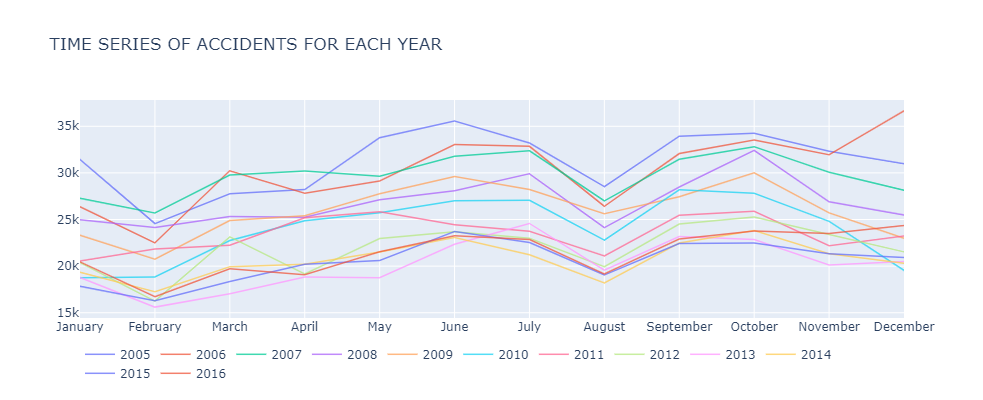
* Shows that October has the highest number of accidents, followed by June and then September



**NOTES:**

* Shows that 2nd, 5th, 6th, 7th, and 23rd have the number of highest accidents in each month





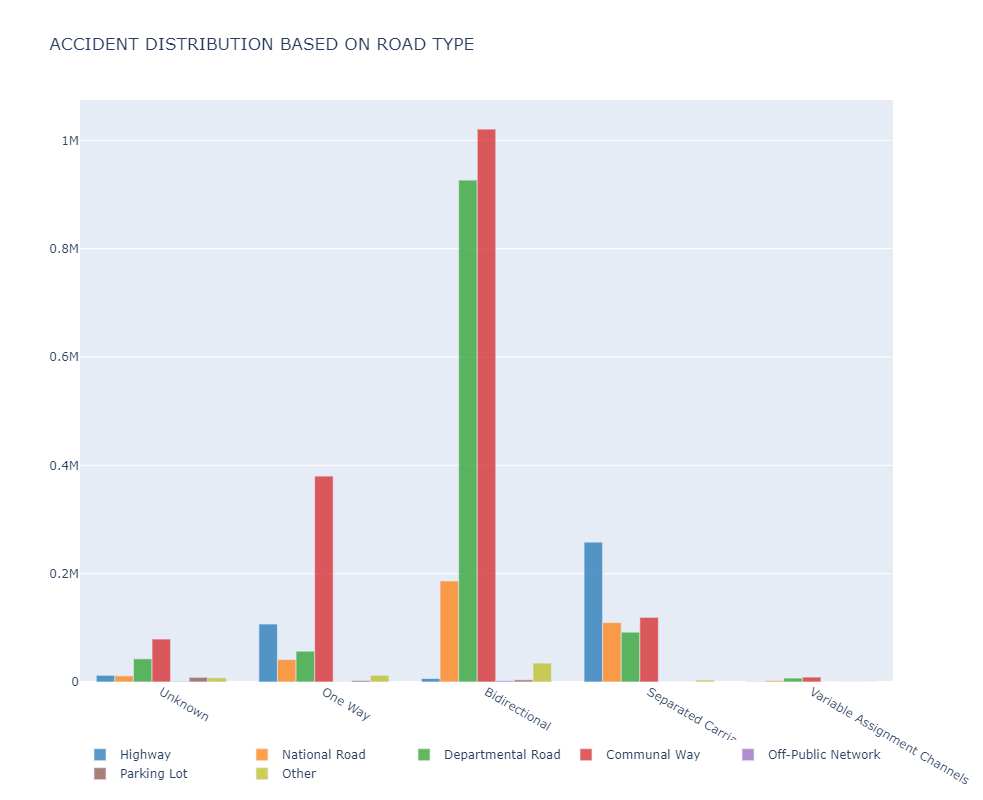
**NOTES:**

* There is an observed sharp rise in the months of June, July, September and October.
* There also are Sharp drops observed in February and August.
* December, 2006 has the highest number of accidents at 36,648.
* February, 2013 has the lowest number of accidents at 15,605.

1. **Exploration based on roads where accidents occurred:**

* Which types of roads are high risk?
* Which type of road gradient is high risk?



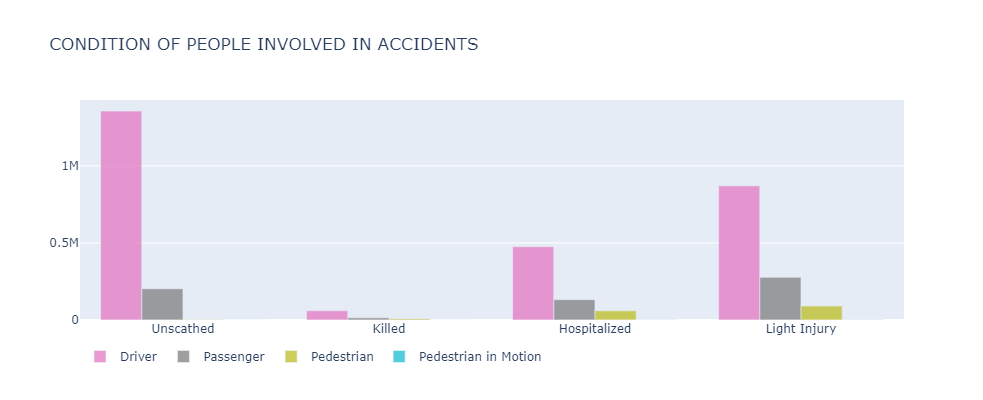


**NOTES:**

* Communal Ways, Departmental Roads, and Bidirectional ways, are roads with higher-risks

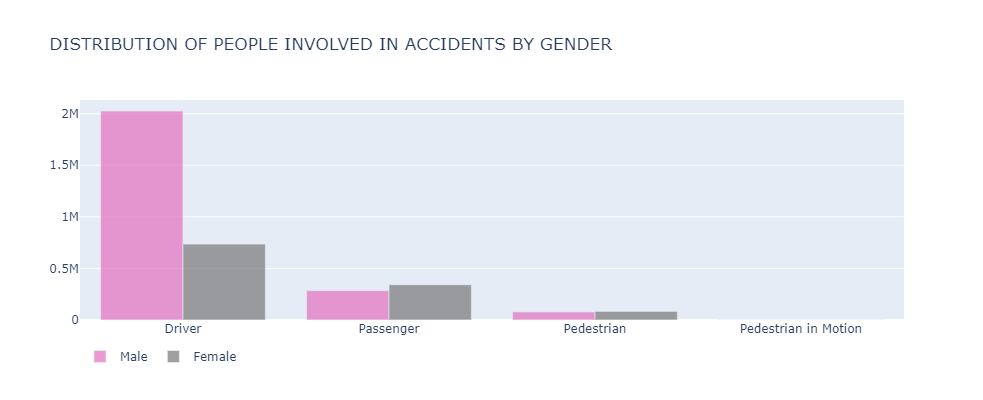
1. **Exploration based on people involved in the accidents:**

* What was the condition of the people after the accident?
* What was the age distribution of the people involved?
* What was the sex distribution of the people involved?

****

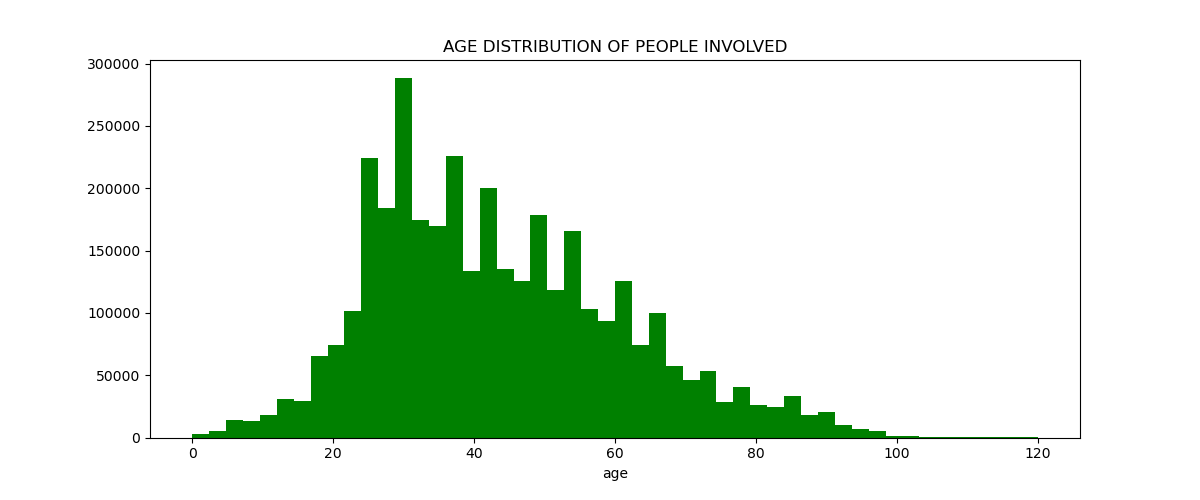
**NOTES:**

* Over 1 million drives were unscathed, about half a million were hospitalized, and about 800 thousand had light injury
* Drivers were the most affected followed by the passengers
* A smaller population of the dataset victims were lost to the cold hands of death.



**NOTES:**

* Male drivers were about 2 Million and female drivers were about 750 thousand in number
* Male passengers were about 300 thousand and female passengers were about 350 thousand in number
* Male and female pedestrians were about 50 thousand respectively

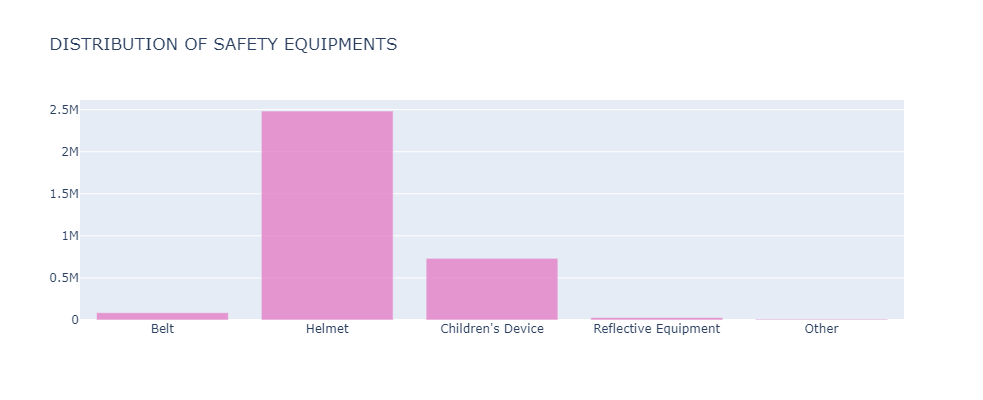
****

**NOTES:**

* The most involved age bracket in accidents in this period are people in their 30s

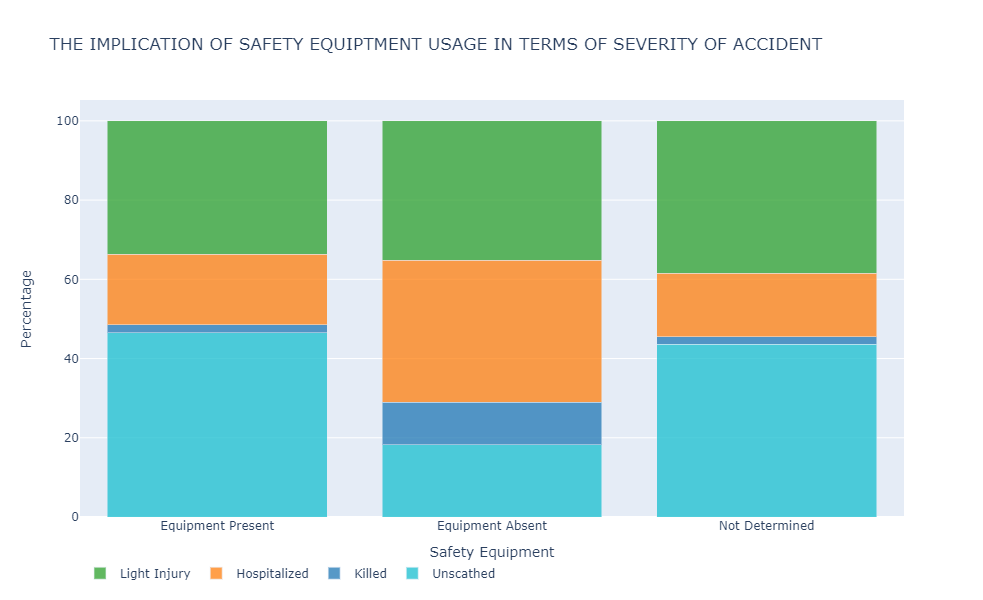
1. **Exploration based on use of safety equipment:**

* What was the distribution of Safety Equipment used?
* Did use of Safety Equipment impact condition of people after the accident?



**NOTES:**

* About 2.5 million persons used the Helmet
* About 750 thousand persons used the Children’s device
* Those who used belts were just about 50 thousand and those who used reflective equipment were barely about 15 thousand in number



**NOTES:**

* In the presence of safety equipment, there was a higher number of unscathed and light injury persons involved with a fewer number of hospitalized, and a very minute number of death recoded, considering the population count involved
* In the absence of safety equipment, there was a slightly higher number of light injured persons, a significantly higher number of hospitalized persons, a significantly smaller number of unscathed persons, and the highest number of death recorded.
* Those who used belts were just about 50 thousand and those who used reflective equipment were barely about 15 thousand in number

**CONCLUSION**

The use of safety equipment cannot be overemphasized as it is proven from this report that there is a lighter danger faced when PPEs are engaged. Road users must therefore be wise and protect themselves first.