

2020

# Predicting Traffic Accident Severity



Applied Data Science  
Capstone

Rautela, Sanjay

# Coursera Applied Data Science Capstone

---

This project is for the Applied Data Science Capstone-car accident severity, this is the final project for Data Science Professional certificate by IBM.

## Data

### Exploring Vehicle Accidents in France

- This notebook represents the exploration and visualization of the dataset Accidents in France Dataset.
- This data set consists of all the recorded accidents in France since 2005 to 2016.

### Reading Data

There are 5 tables and it will be a good idea to read, explore and analyses all the tables separately. To avoid this situation, we need to merge all the tables into one data frame based on the attribute representing accident ID i.e. Num\_Acc attribute. Once the data from 5 tables are merged we can check the dimension of our dataframe.

#### *Shape of Dataframe*

- Rows: 3553976
- Columns: 52

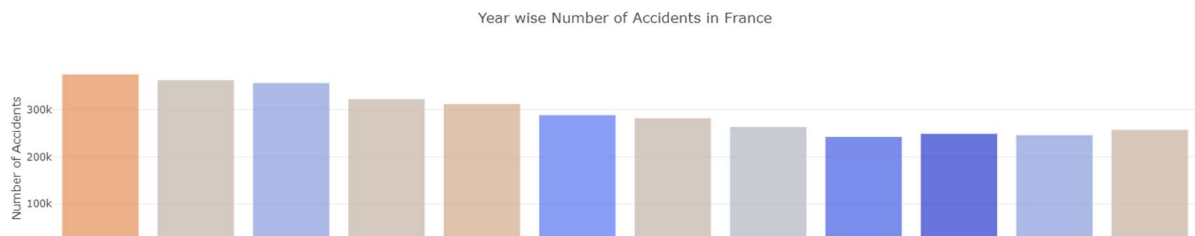
### Cleaning the Data

- We identify the percentage of NaN values in the dataset and eliminate the columns where most of the values are NaN
- Drop the column which are relevant for data exploration.

# Exploration

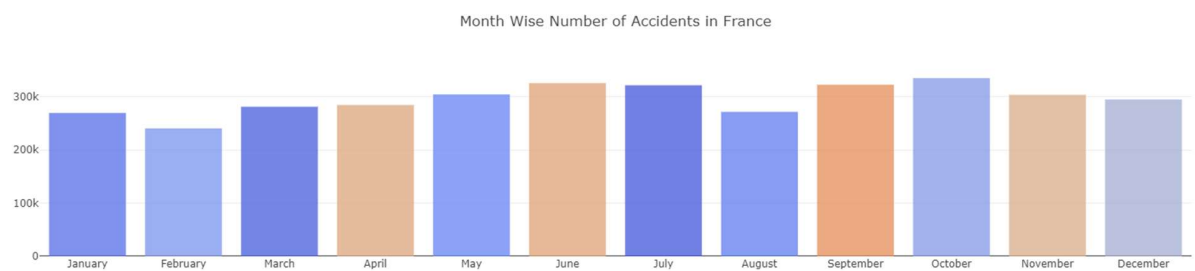
## Exploration based on date of accidents

### *Trend in Accidents per year from 2005 to 2016*



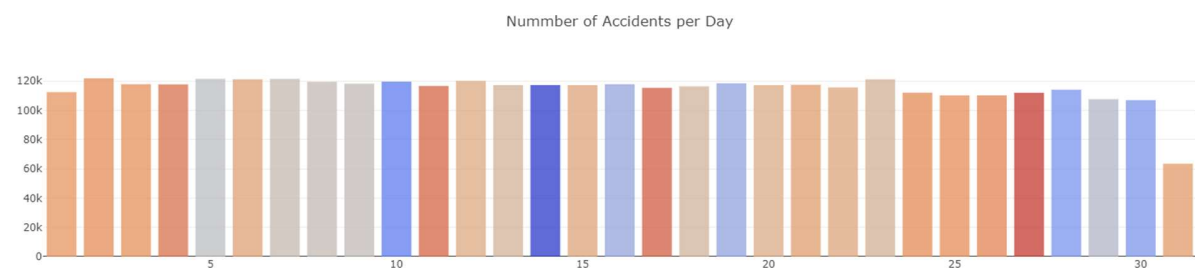
- Number of accidents in France have steadily decreased from 2005 to 2013, but an increasing trend is observed from 2013 to 2016.
- In this dataset, 2005 had the largest number of accidents equal to 374561.

### *Frequency of Accidents on monthly basis*



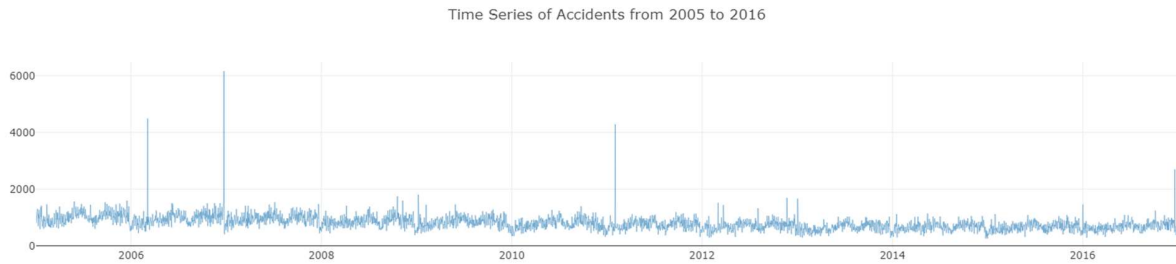
- June, July, September, October have the highest number of accidents, while February has the lowest.
- On an average about 296,164 accidents occur every month in France.
- October has the highest number of accidents (with about 334,884 incidents) than any other month
- Weather in France during September and October is cold and wet whereas, June and July form the peak tourist season

### *Safest day-of-the-month to drive*

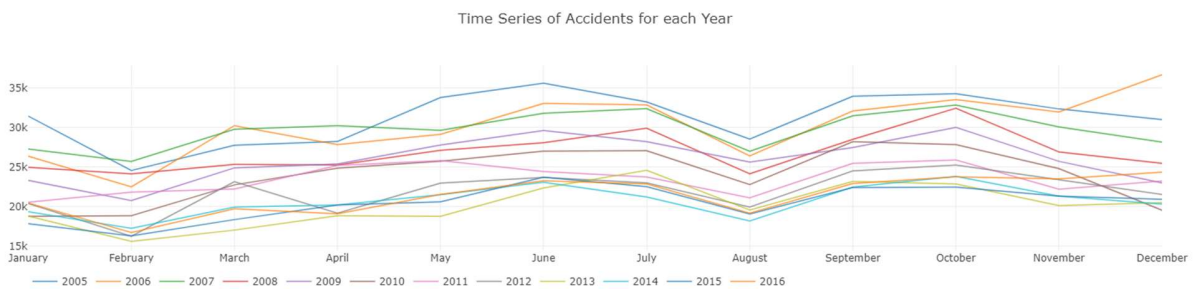


- Number of accidents per month-day is mostly uniform. 31st is lowest because only 7 months have 31 days.
- On an average about 114,644 accidents occur every day in France.

### *Time series of all accidents from 2005 to 2016*

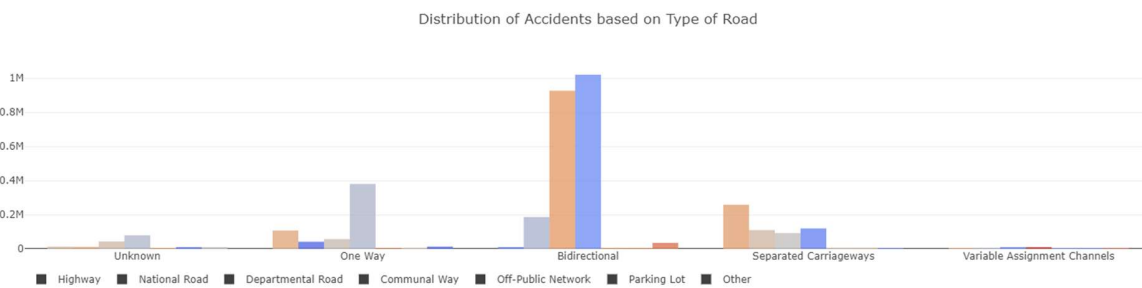


### *Time series for all accidents in each year*



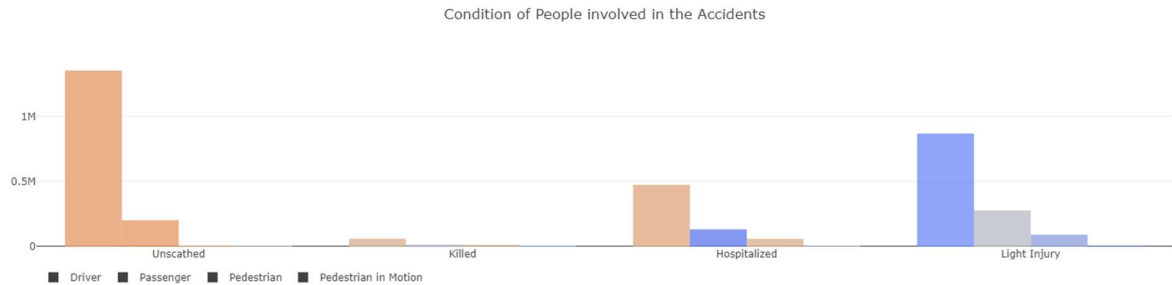
- A sharp rise is observed in the months of June, July, September and October.
- Sharp drops are observed in February and August.
- December 2006 has the highest number of accidents at 36,648.
- February 2013 has the lowest number of accidents at 15,605.

*Which types of roads are high risk?*

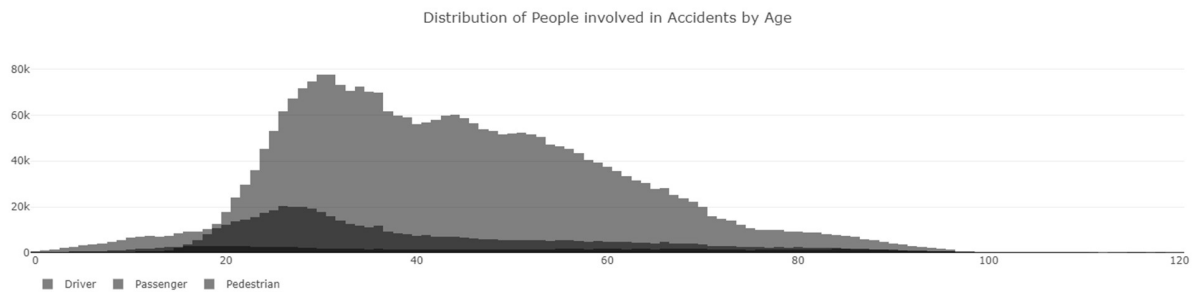


## Exploration based on people involved in the accidents

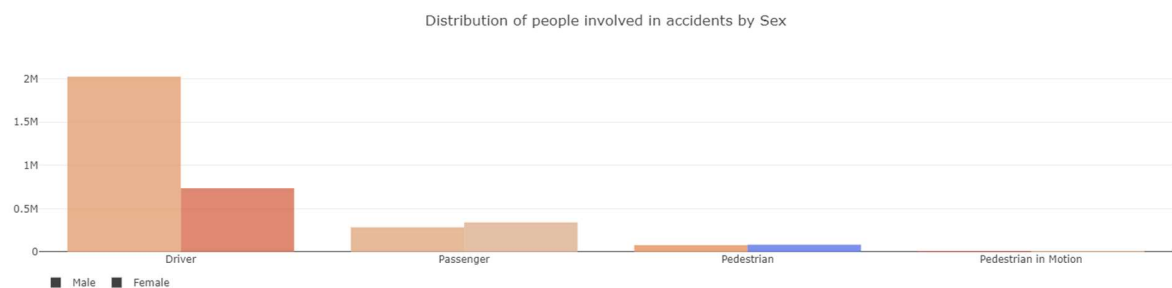
*What was the condition of the people after the accident?*



*What was the age distribution of the people involved?*

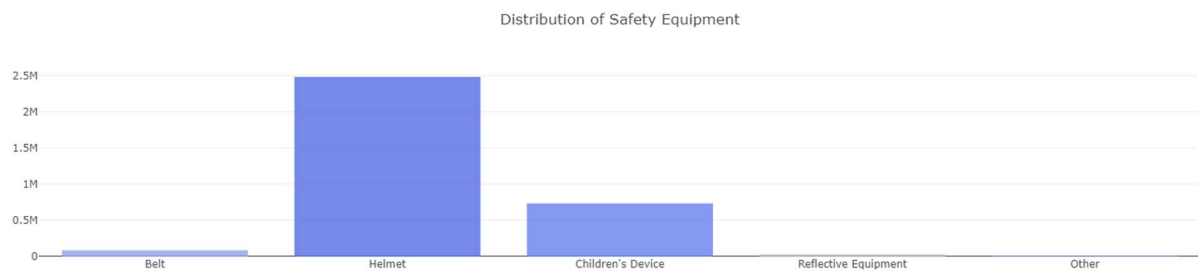


*What was the sex distribution of the people involved?*

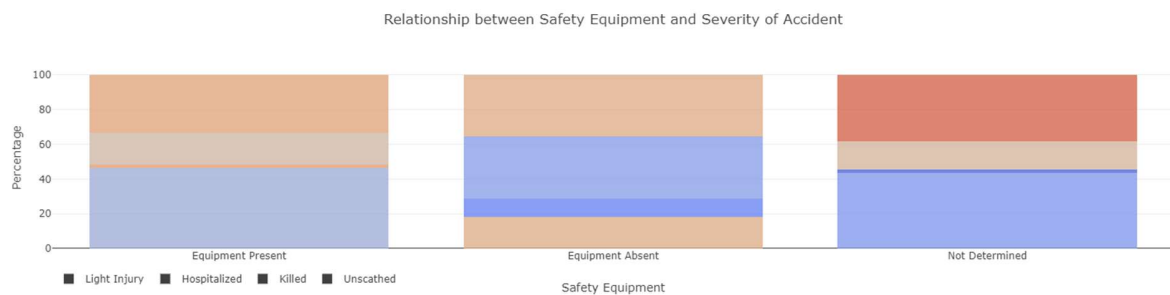


## Exploration based on use of safety equipment

*What was the distribution of Safety Equipment used?*



*Did use of Safety Equipment impact condition of people after the accident?*



- 1.9% of the people were killed and 17.8% were hospitalized if Safety Equipment was used.
- 10.7% of the people were killed and 35.8% were hospitalized if Safety Equipment was not used.