Analysis on Traffic Crashes in Chicago

Raghul Sekar, Raj Deelip Sounder Raj, Mohammad Rafi Shaik

10/30/2021

Datasets:

Crashes Dataset:

Crash data shows information about each traffic crash on city streets within the City of Chicago limits and under the jurisdiction of Chicago Police Department (CPD)

People Dataset:

This data contains information about people involved in a crash and if any injuries were sustained.

Vehicle Dataset:

This dataset contains information about vehicles (or units as they are identified in crash reports) involved in a traffic crash.

Cleaning the data in Crashes Dataset

There are exactly 1131 null values in multiple columns. Upon checking further, they are all in same rows. So deleted the rows.

Checked null value percentage in other columns and deleted the columns with more than 80% null values.

Cleaning the data in Vehicles Dataset

Deleted columns with null values greater than 80%

Cleaning the data in People Dataset

```
##            PERSON_ID          PERSON_TYPE       CRASH_RECORD_ID
##                 0.00                 0.00                  0.00
##                RD_NO           VEHICLE_ID           CRASH_DATE
##                 0.77                 1.98                  0.00
##              SEAT_NO                 CITY                 STATE
##                79.57                26.49                 25.60
```

```
##               ZIPCODE                   SEX                   AGE
##                 32.88                  1.51                 28.75
## DRIVERS_LICENSE_STATE DRIVERS_LICENSE_CLASS      SAFETY_EQUIPMENT
##                 41.02                 49.21                  0.29
##       AIRBAG_DEPLOYED             EJECTION INJURY_CLASSIFICATION
##                  1.90                  1.22                  0.05
##              HOSPITAL            EMS_AGENCY            EMS_RUN_NO
##                 81.50                 88.02                 98.04
##         DRIVER_ACTION         DRIVER_VISION    PHYSICAL_CONDITION
##                 20.62                 20.65                 20.56
##       PEDPEDAL_ACTION   PEDPEDAL_VISIBILITY     PEDPEDAL_LOCATION
##                 98.13                 98.14                 98.13
##            BAC_RESULT      BAC_RESULT.VALUE        CELL_PHONE_USE
##                 20.53                 99.87                 99.91
```

Although PEDPEDAL_ACTION, PEDPEDAL_VISIBILITY, PEDPEDAL_LOCATION have more than 90% null values, we are not deleting the values as those columns are applicable only to Pedestrians and bicyclists who are nearly 2%

```
##     PEDPEDAL_ACTION PEDPEDAL_VISIBILITY   PEDPEDAL_LOCATION
##                   2                   2                   2
```

Removed other columns with more than 80% null values.

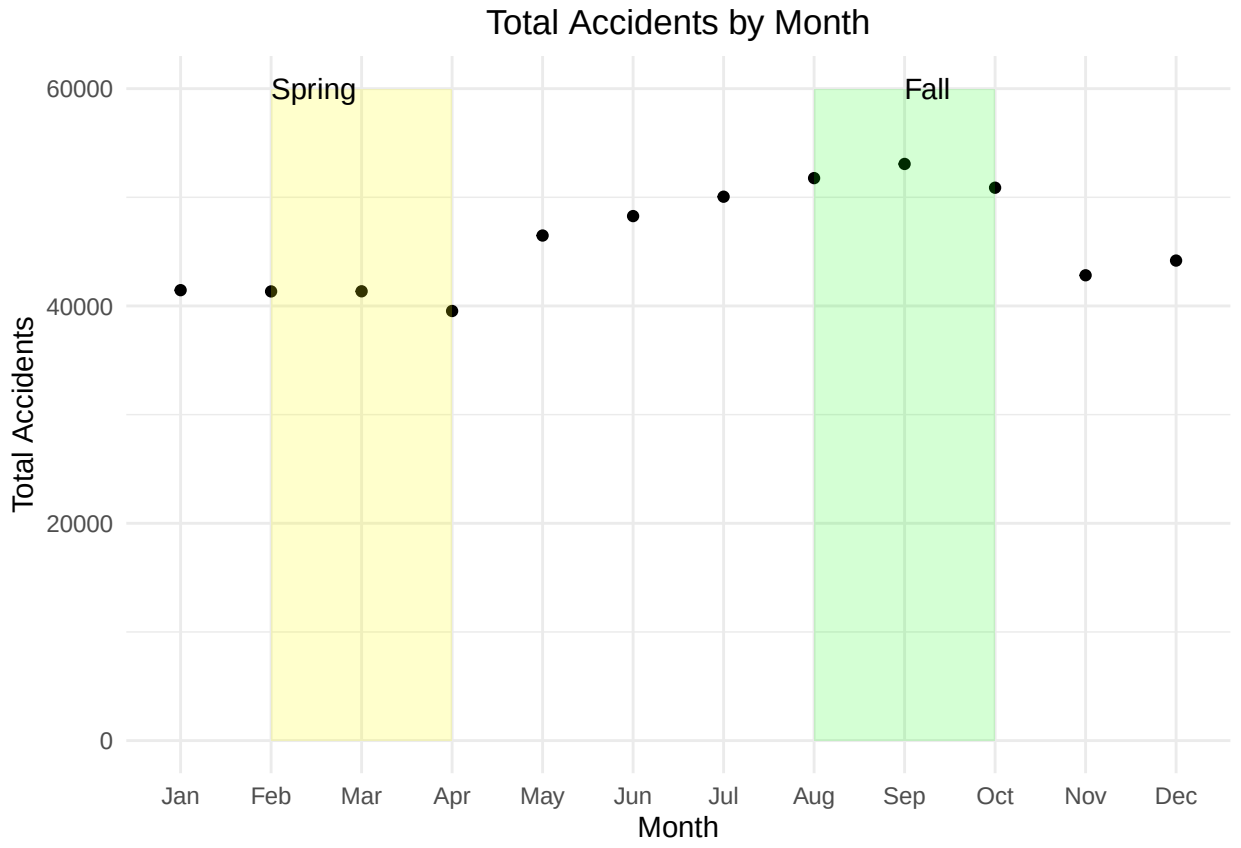
Modified date columns to Datetime format in all three datasets.

Business Questions

Spread of accidents over the time period

Monthwise distribution of crashes

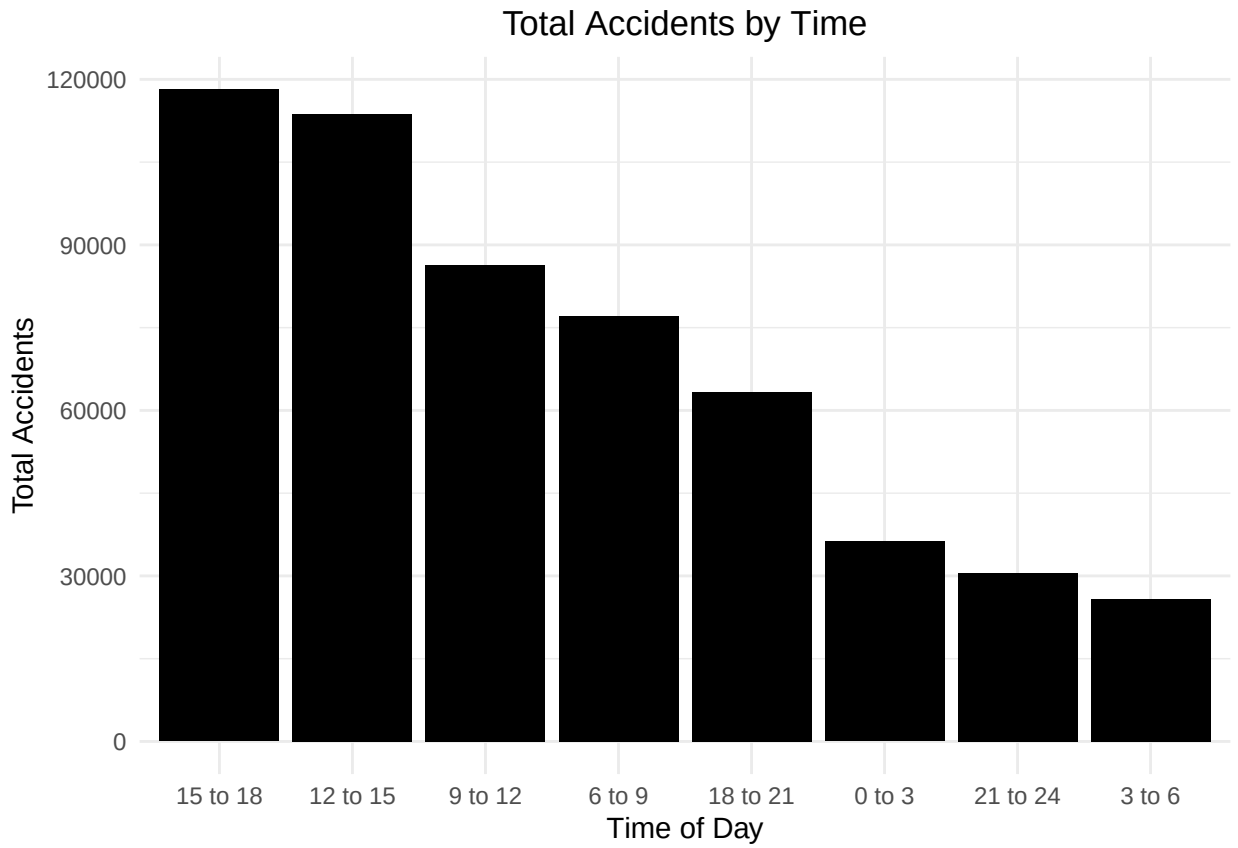
Grouped the dataset on month and aggregated crash counts. Created scatter plot with timeline on X-axis and crashes on Y-axis.



Accidents are more probable to happen during the second half of the year especially during the fall and starts to dip during the winter.

Time of Day vs number of crashes

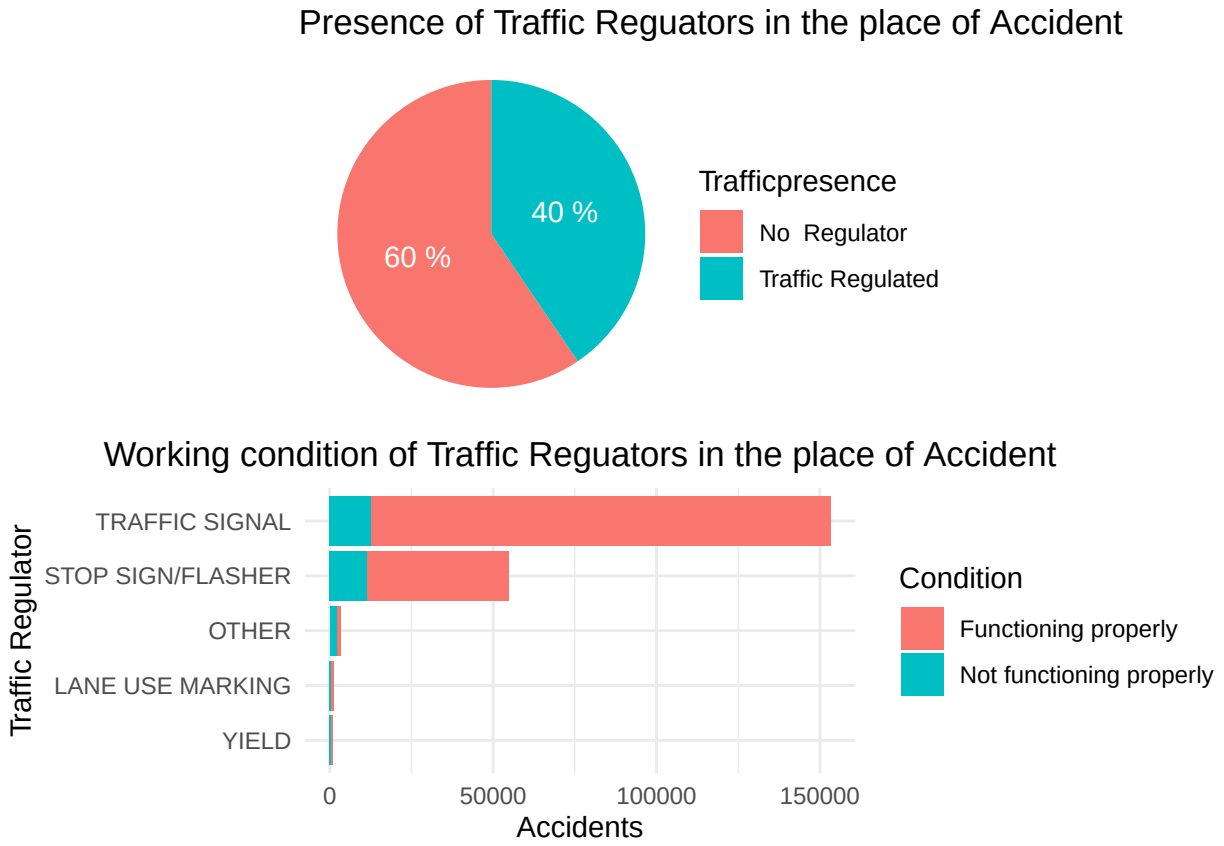
Classified hours into different ranges and grouped the data to aggregate crashes. Shown bar chart in ordered groups with timeline on X-axis and crashes on Y-axis.



Most number of accidents are reported during 12 PM to 6 PM. This is logical as majority of student,employees and others commute during this time. Its advisable to deploy more highway patrol officers to oversee the traffic during this time period.

Analysis on the effect of presence of Traffic Regulators

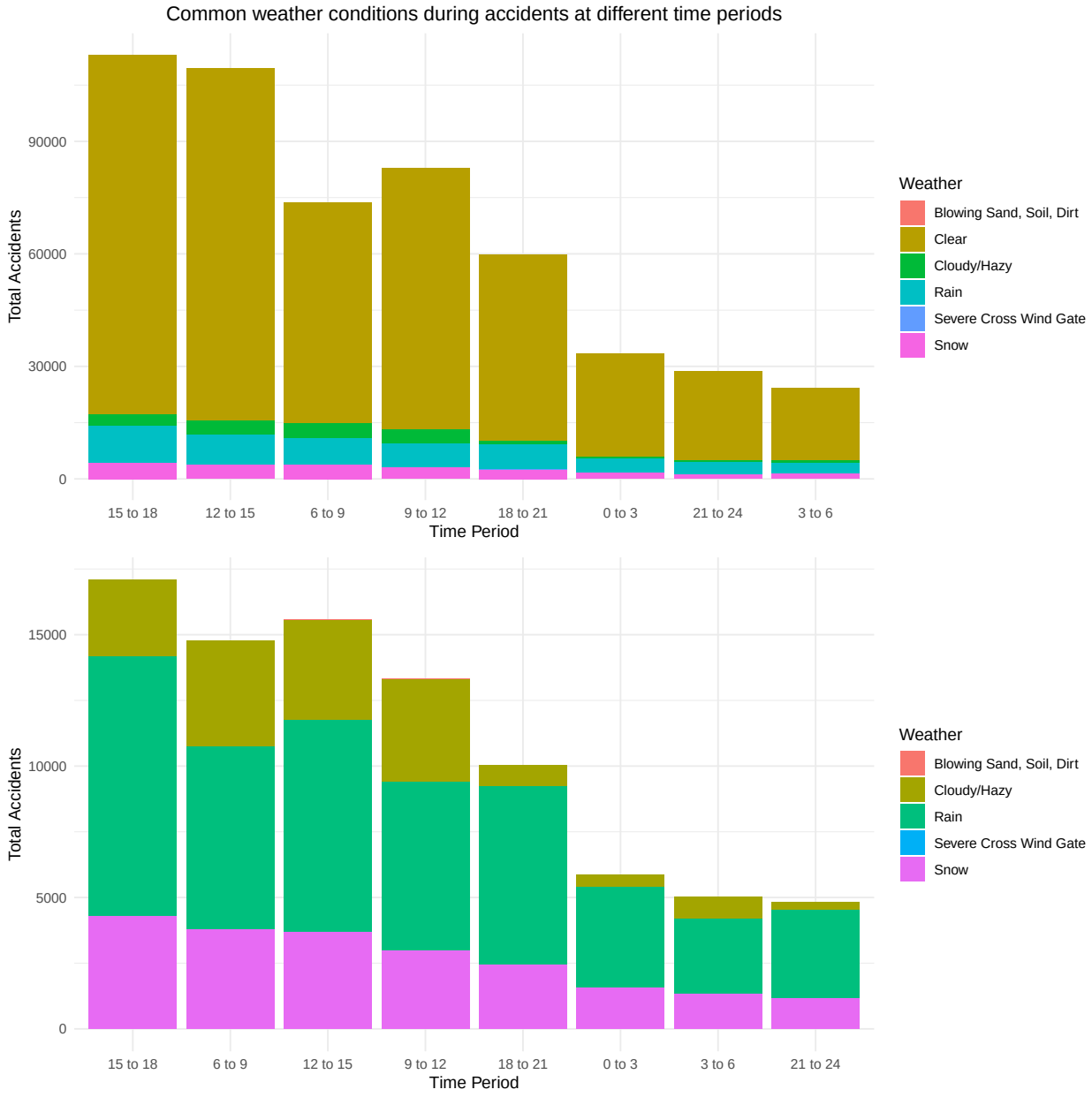
Grouped the data by the presence of Traffic Regulators to check the percentage of accidents under each scenario. Then filtered records where traffic regulators were present and checkied the working condition of the regulators where the accidents occurred.



About two-thirds of the total accidents happened in places without Traffic Regulators. In spite of fine working, accidents have been reported from places with traffic regulators (especially where traffic signals are present).

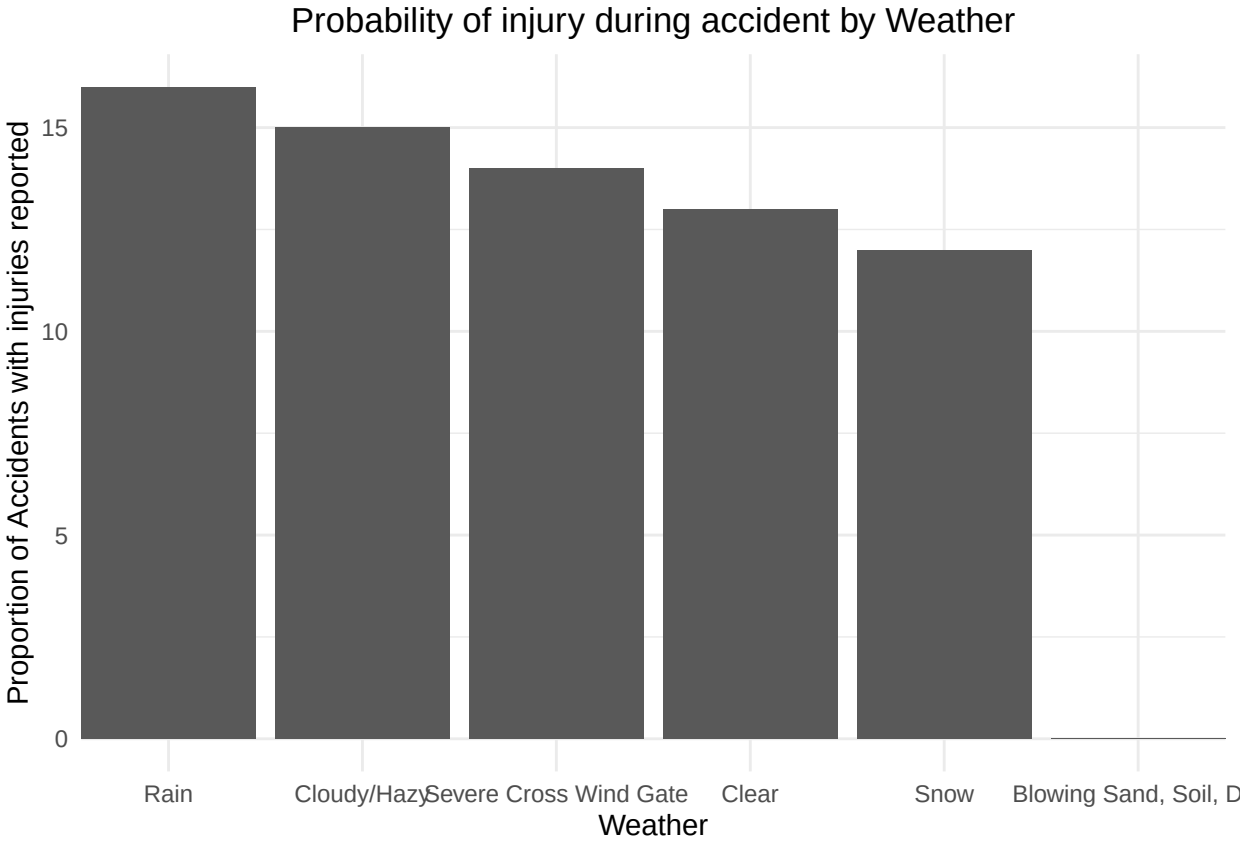
Analysis of Environmental facotors on crashes

Created a new column to group similar weather conditions together and then grouped by Weather and Time to find the common weather conditions during different times of day. Since the weather in Chicago is Clear for most of the days, we filtered the records where weather was not clear and plotted again.



Most of the accidents reported have occurred on a clear day. This is expected as majority of the days in a year in Chicago are clear. On excluding Clear days, we see that accidents are more likely to happen on Rainy or snowy days particularly in late hours of the day.

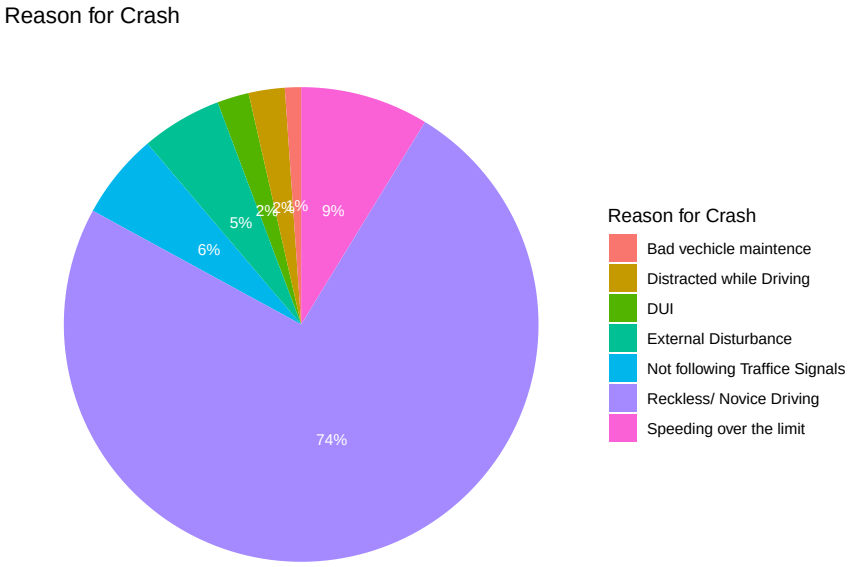
Filtered the accidents which caused injuries and aggregated them based on weather conditions to find which weather condition leads to severe accidents.

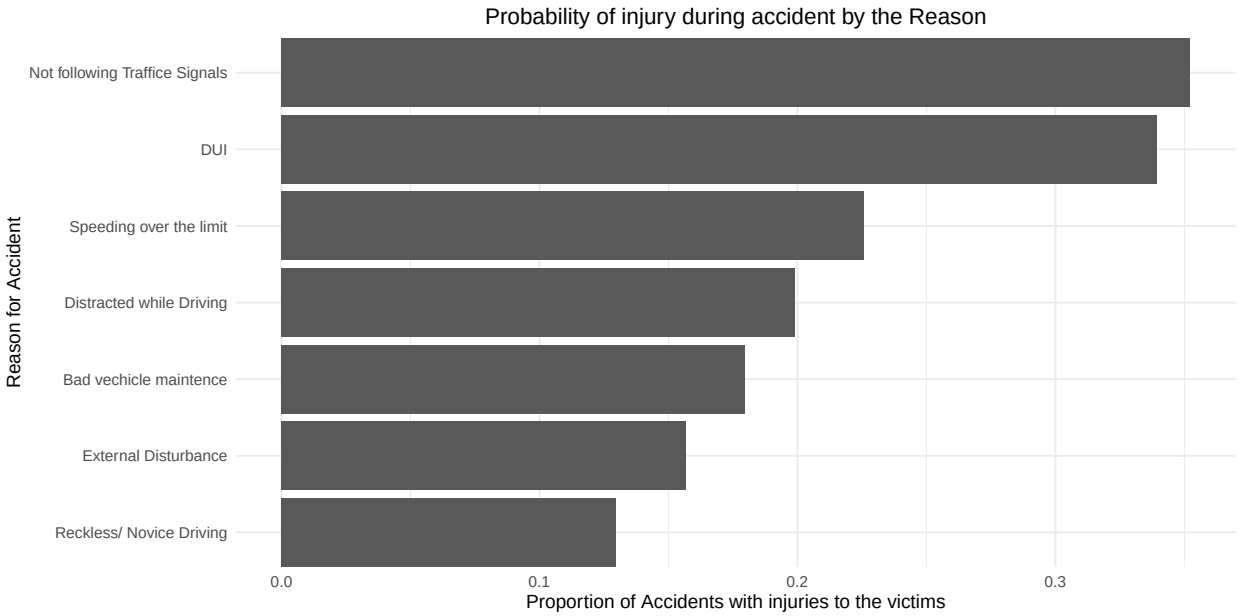


Although most accidents occurred on Clear days they are minor ones with less or no injury. Accidents happening on Rainy, Cloudy or Windy are more likely to be severe in nature.

Analysis of the different primary causes on accidents.

Grouped similar causes together and plotted a pie chart to show the common causes. Then filtered the injurious accidents and plotted a horizontal bar chart to show the most common causes which lead to severe accidents.

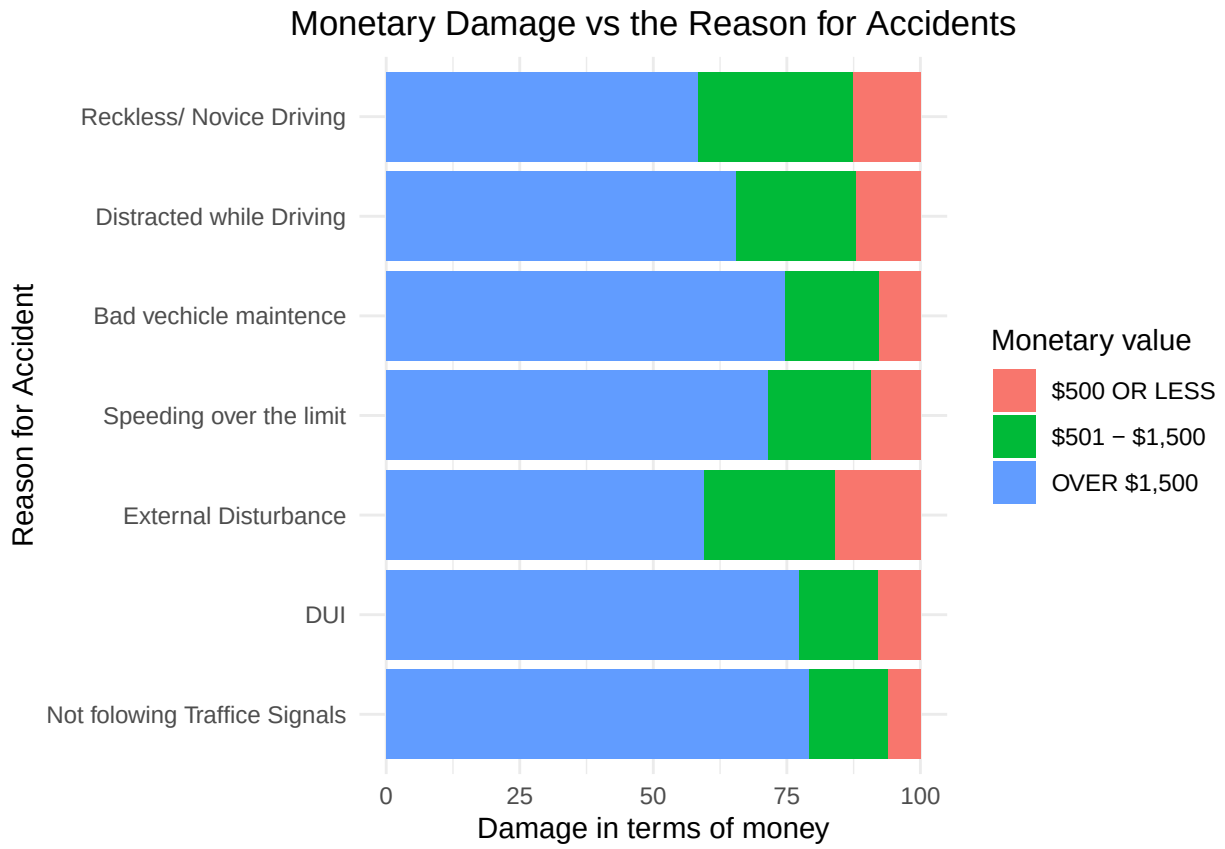




In about 3/4th of the accidents, reckless/Novice driving skills have been reported as the primary reason. Speeding and not following traffic rules are other major primary reasons. Driving under influence and not following traffic traffic signals are reportedly more dangerous.

Analysis of the damage caused with different cause of accidents

Filtered the rows which are not applicable and grouped similar causes together and cut the data by the damage interms of monetary values.



Irrespective of the reason for accident, majority of the accidents reported have caused damage worth more than 1500 dollars. Damage cost due to DUI and traffic rule violations are more likely to be high

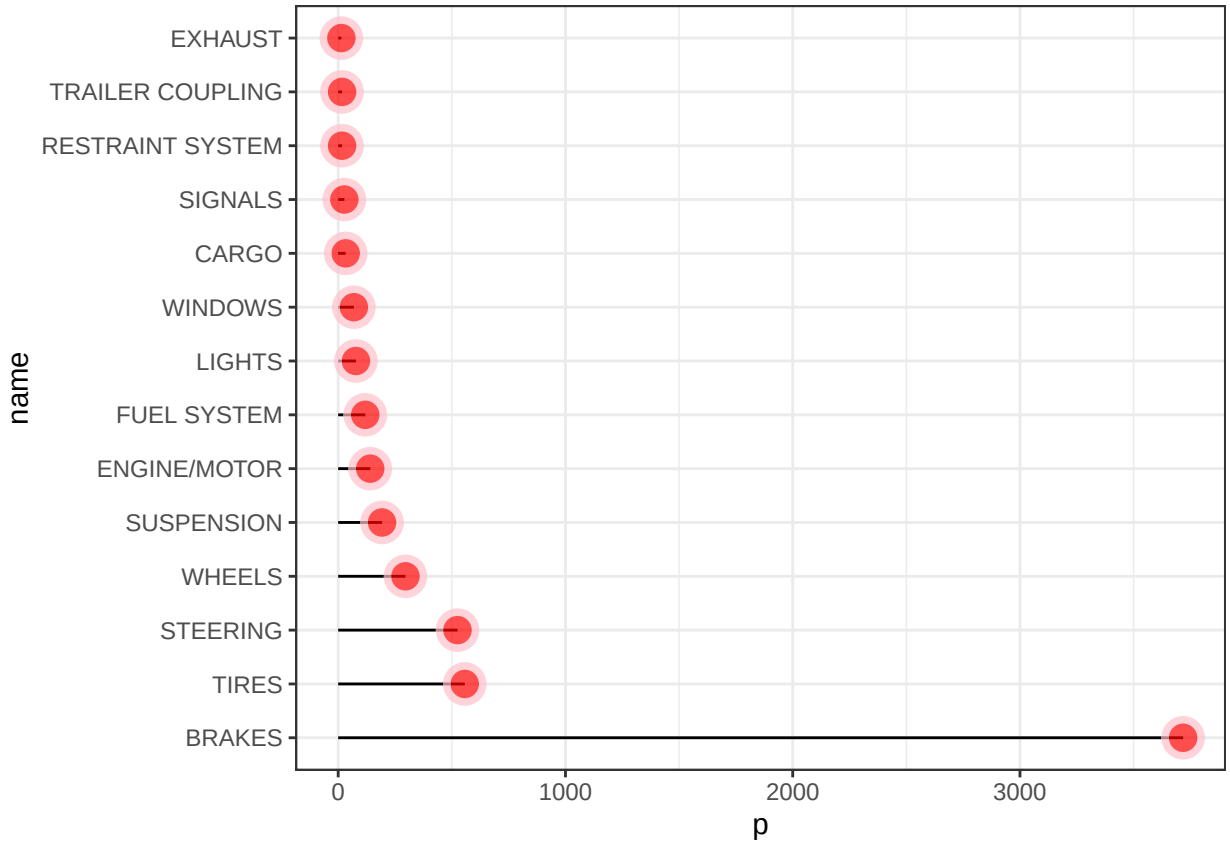
Analysis on the usage of safety equipments

Classified injuries and safety_equipment into broader groups and grouped similar data together to find the effect of safety equipments usage on severity.

Safety_Equipment_Used	Injury_Type	Percentage
Not Used	Non Severe	94.12
Not Used	Severe	5.88
Used	Non Severe	99.32
Used	Severe	0.68

Analysis of vehicle defects on crashes

Grouped records based on Vehicle_Defects plotted a bar chart to find the most common defects in vehicles causing accidets.



Among records where defects are found in vehicles, defects in brakes is the most common. Defects in Tires and steering are other causes.

Analysis of accidents involving pedestrians/bicyclists

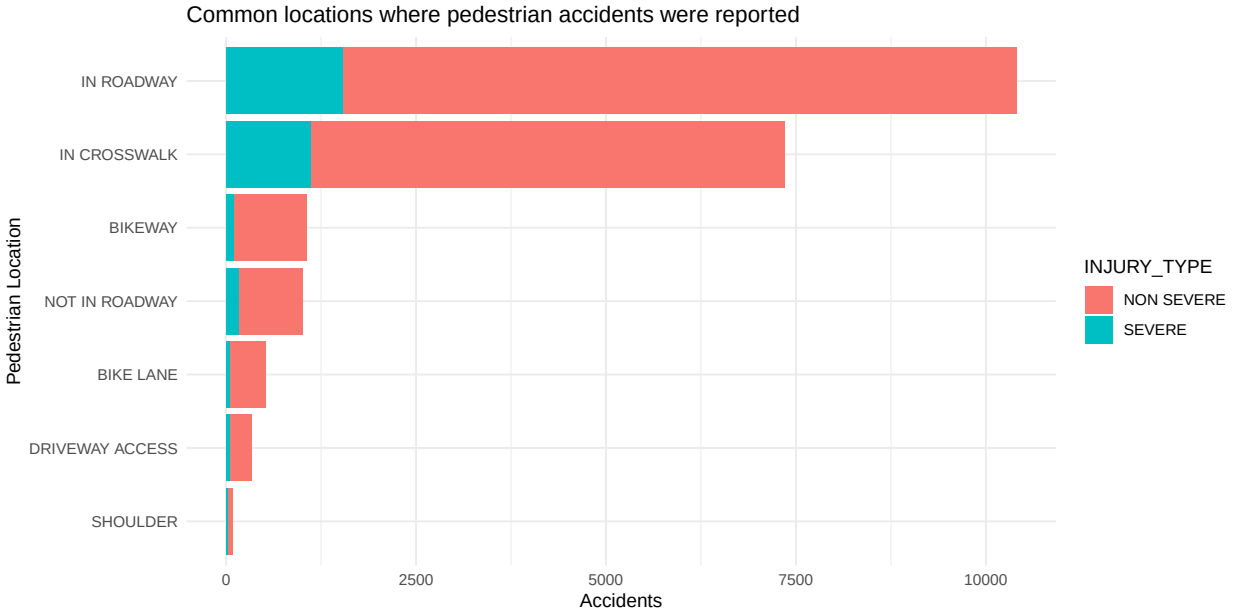
Filtered out pedestrians and Bicyclists data to find the percentage of accidents where pedestrians are involved.

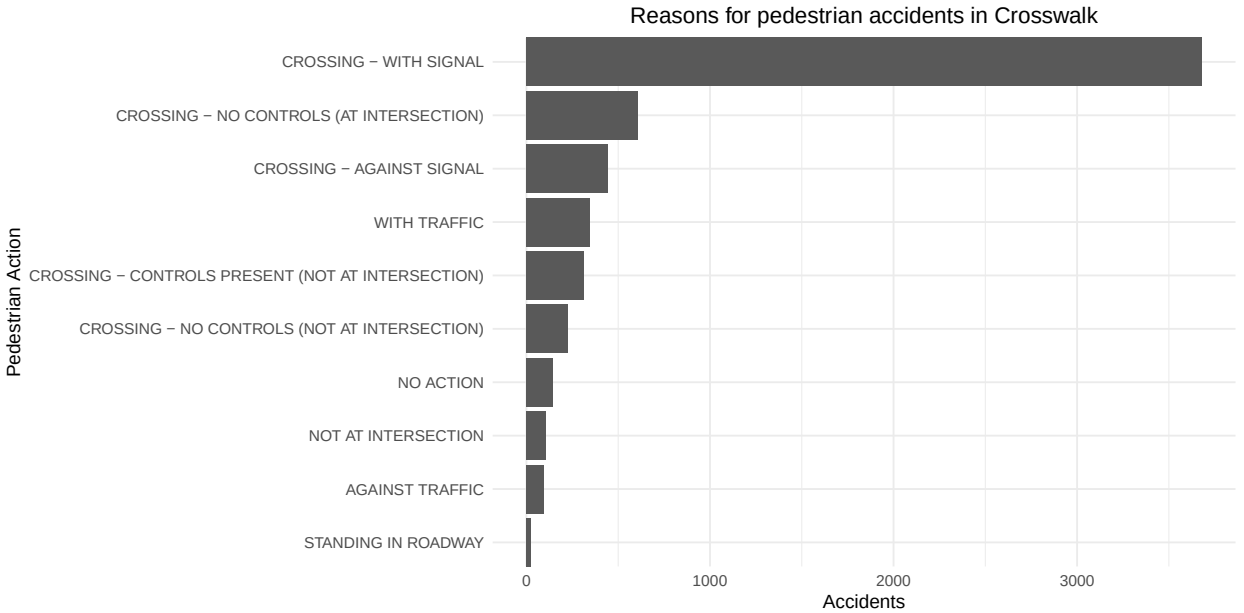
Percent of accidents involving pedestrians
3.9

Pedestrians are involved in almost 4% of the total crashes.

Common location where pedestrians realted accidents happened

Categorised injuries as Severe and Non Severe and then grouped the data based on location where accidents are reported. Plotted a stacked bar chart to find the common locations with pedestrian accidents and find where severe crashes occurred. After that we took the common location and analysed the reason for accidents at that particular location





Most pedestrian related accidents happened in Roadway and Crosswalk. Its advisable for pedestrians to stay away from Roadways and walk on platforms. In Cross walks, most accidents happened when people were crossing the roads with signal. The probable reason could be that the vehicles did not follow the traffic signal or failed to yield way to pedestrians. Other than that more accidents transpired at intersections without traffic signals and crossing against signals.

Recommendations

Drive carefully especially during the fall season. More highway patrol officers should be on duty during the later half of the day.

Avoid driving under the influence and respect the traffic rules. Do regular maintence check of your vehicles, espescially check if the brakes are functioning properly.

Pedestrians should be careful while crossing the road even if they are crossing in crosswalks

References

1) https://www.r-graph-gallery.com/

2) https://sebastiansauer.github.io/figure_sizing_knitr/