Car Collision severity report

Ankita Sinha

Introduction

Background:

According to cdc.gov (https://www.cdc.gov/injury/features/global-road-safety/index.html), "Road traffic crashes are a leading cause of death in the United States for people aged 1–54 and the leading cause of non-natural death for healthy U.S. citizens residing or traveling abroad... Each year, 1.35 million people are killed on roadways around the world. Every day, almost 3,700 people are killed globally in road traffic crashes involving cars, buses, motorcycles, bicycles, trucks, or pedestrians. More than half of those killed are pedestrians, motorcyclists, and cyclists. Road traffic injuries are estimated to be the eighth leading cause of death globally for all age groups and the leading cause of death for children and young people 5–29 years of age. More people now die in road traffic crashes than from HIV/AIDS."

With the increasing number of vehicles that hit the road every year, the number of accidents and fatalities is only expected to increase. The reduction of accidents is hence of immense importance to people, healthcare organizations, private corporations and government bodies across the world to prevent the tragic loss of life, limb and property. The Seattle Police Department maintains a record of all the car collision incidents that occur within its jurisdictional limits, which is updated weekly on its website. An analysis of the car collision dataset may reveal interesting observations and patterns regarding the occurrence of car collisions, such as:

- Days of week which are more prone to accidents.
- Time of the day in which accidents are more likely to occur.
- Accident prone localities and streets.
- Whether some areas have a high occurrence of accidents during nighttime.
- Whether weather conditions affect the occurrence/severity of accidents.
- Whether road conditions affect the occurrence/severity of accidents.

Audience:

The analysis could be of particular interest to:

- Traffic departments and civic agencies who might be looking for ways to reduce the accidents and resulting injuries/fatalities by identifying zones and situations which have a higher incidence of collisions and help them take appropriate measures for the same.
- Residents, drivers and pedestrians who could be warned about the likelihood of an accident on their travel route based on the location, weather, road conditions, etc.

Data description

The Data-Collisions.csv dataset has 194673 rows and 38 columns. Each accident has an unique primary and secondary key. Each accident location is uniquely identified through ESRI unique identifiers and geometry fields. Description of the general location of the accident site, type of the address(whether it is an alley/block/intersection), codes identifying the lane/crosswalk are also available in addition to the time and date of the accident. The severity of the accident is denoted by the severity code. The type of collision, the

number of people and vehicles involved, total number of injuries, further segregated by serious injuries and fatalities are also available. The role of the driver in the accident, whether it was due to inattention, intoxication, weather/road conditions, speeding, whether right of way was given to pedestrian and whether it involved colliding with a parked car is also available in the dataset. A few more columns are also available in the dataset which do not seem to be very useful and would be dropped for the purpose of the analysis.

The dataset can be downloaded at this link. The metadata can be downloaded here

Methodology

A Jupyter notebook is used to perform the data analysis. The libraries used are numpy, pandas, matplotlib and seaborn. The dataset is checked for size and completeness. The target variable is Severity code and hence the columns relevant to the analysis are looked at and where useful, plotted using graphs to check how much they influence the severity of a collision. Data cleaning is performed so that categorical variables with descriptive values are transformed into numerical categories.

```
import pandas as pd
import numpy as np
import matplotlib as plt
import seaborn as sns
print("Hello Capstone Project Course!")
```

```
Hello Capstone Project Course!
```

Fig 1: imported python libraries

```
df.dtypes
```

```
SEVERITYCODE        int64
X                 float64
Y                 float64
OBJECTID            int64
INCKEY              int64
COLDETKEY           int64
REPORTNO           object
STATUS             object
ADDRTYPE           object
INTKEY            float64
LOCATION           object
EXCEPTRSNCODE      object
EXCEPTRSNDESC      object
SEVERITYCODE.1      int64
SEVERITYDESC       object
COLLISIONTYPE      object
PERSONCOUNT         int64
PEDCOUNT            int64
PEDCYLCOUNT         int64
VEHCOUNT            int64
INCDATE            object
INCDTTM            object
JUNCTIONTYPE       object
SDOT_COLCODE        int64
SDOT_COLDESC       object
INATTENTIONIND     object
UNDERINFL          object
WEATHER            object
ROADCOND           object
LIGHTCOND          object
PEDROWNOTGRNT      object
SDOTCOLNUM        float64
SPEEDING           object
ST_COLCODE         object
ST_COLDESC         object
SEGLANEKEY          int64
CROSSWALKKEY        int64
HITPARKEDCAR       object
dtype: object
```

Fig 2: data types of columns in dataset.

```
df['SEVERITYDESC'].value_counts().plot(kind='bar')
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x25a84934630>
```

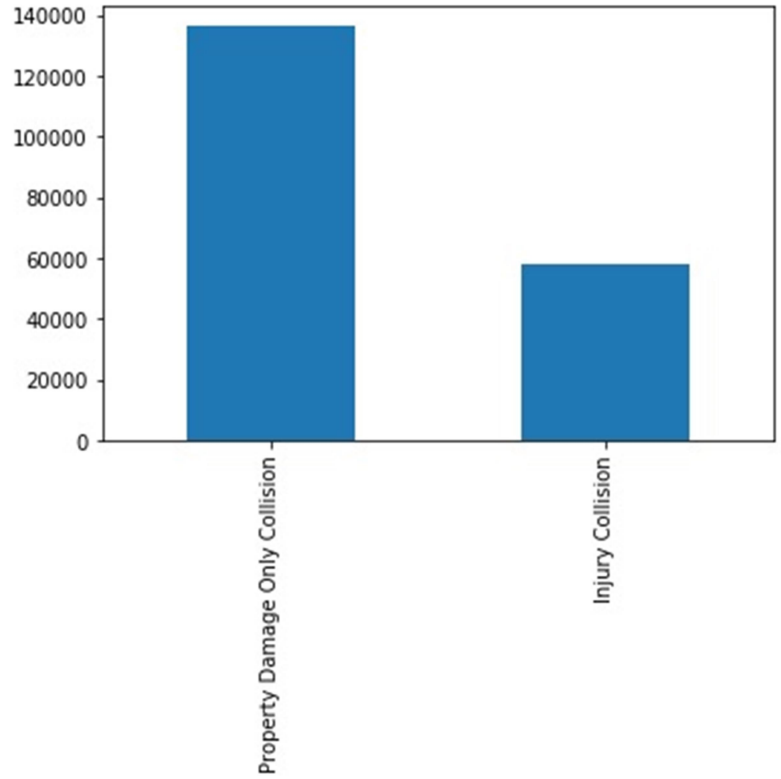

Fig 3: Collisions grouped by severity codes.

```
df['dayofweek'].value_counts().plot(kind='bar')
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x25a8c392438>
```

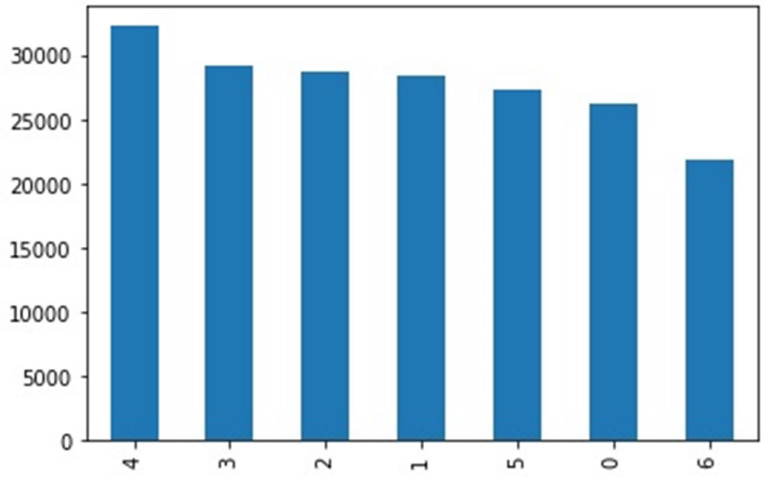

Fig 4: Collision grouped by day of week

```
df.groupby(['SEVERITYDESC'])['LIGHTCOND'].value_counts()
```

SEVERITYDESC	LIGHTCOND	
Injury Collision	Daylight	38544
	Dark - Street Lights On	14475
	Dusk	1944
	Dawn	824
	Unknown	605
	Dark - No Street Lights	334
	Dark - Street Lights Off	316
	Other	52
	Dark - Unknown Lighting	4
Property Damage Only Collision	Daylight	77593
	Dark - Street Lights On	34032
	Unknown	12868
	Dusk	3958
	Dawn	1678
	Dark - No Street Lights	1203

Fig 5: Collision grouped by light conditions

```
df.groupby(['SEVERITYDESC'])['WEATHER'].value_counts()
```

SEVERITYDESC	WEATHER	
Injury Collision	Clear	35840
	Raining	11176
	Overcast	8745
	Unknown	816
	Fog/Smog/Smoke	187
	Snowing	171
	Other	116
	Sleet/Hail/Freezing Rain	28
	Blowing Sand/Dirt	15
	Severe Crosswind	7
	Partly Cloudy	3
Property Damage Only Collision	Clear	75295
	Raining	21969
	Overcast	18969
	Unknown	14275
	Snowing	736
	Other	716
	Fog/Smog/Smoke	382
	Sleet/Hail/Freezing Rain	85
	Blowing Sand/Dirt	41
	Severe Crosswind	18
	Partly Cloudy	2

Name: WEATHER, dtype: int64

Fig 6: Collision grouped by weather conditions

```
df.groupby(['SEVERITYDESC'])['VEHCOUNT'].value_counts()
```

SEVERITYDESC	VEHCOUNT	
Injury Collision	2	35949
	1	14105
	3	5470
	0	1227
	4	1078
	5	261
	6	60
	7	22
	9	6
	8	5
	11	3
	10	2
Property Damage Only Collision	2	111701
	1	11643
	3	7540
	0	3858
	4	1348
	5	268

Fig 7: Collisions grouped by vehicle count

```
df.groupby(['SEVERITYDESC'])['ADDRTYPE'].value_counts()
```

```
SEVERITYDESC                  ADDRTYPE
Injury Collision              Block           30096
                              Intersection    27819
                              Alley              82
Property Damage Only Collision Block          96830
                              Intersection    37251
                              Alley             669
Name: ADDRTYPE, dtype: int64
```

```
df.groupby(['SEVERITYDESC'])['LOCATION'].value_counts()
```

```
SEVERITYDESC                  LOCATION
Injury Collision              AURORA AVE N BETWEEN N 117TH PL AND N 125TH ST                  120
                              6TH AVE AND JAMES ST                                            107
                              N NORTHGATE WAY BETWEEN MERIDIAN AVE N AND CORLISS AVE N          94
                              RAINIER AVE S BETWEEN S BAYVIEW ST AND S MCCLELLAN ST             94
                              AURORA AVE N BETWEEN N 130TH ST AND N 135TH ST                    88
                              BATTERY ST TUNNEL SB BETWEEN AURORA AVE N AND ALASKAN WY VI SB    86
                              BATTERY ST TUNNEL NB BETWEEN ALASKAN WY VI NB AND AURORA AVE N    78
```

Fig 8: Collisions grouped by address type and location

```
df.groupby(['SEVERITYDESC'])['JUNCTIONTYPE'].value_counts()
```

SEVERITYDESC	JUNCTIONTYPE	
Injury Collision	At Intersection (intersection related)	27174
	Mid-Block (not related to intersection)	19404
	Mid-Block (but intersection related)	7297
	Driveway Junction	3234
	At Intersection (but not related to intersection)	623
	Ramp Junction	54
	Unknown	2
Property Damage Only Collision	Mid-Block (not related to intersection)	70396
	At Intersection (intersection related)	35636
	Mid-Block (but intersection related)	15493
	Driveway Junction	7437
	At Intersection (but not related to intersection)	1475
	Ramp Junction	112
	Unknown	7

Fig 9: Collisions grouped by junction type

```
sns.heatmap(df_new.corr(),linewidths=.5,cmap="YlGnBu")
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x25a8a9819b0>
```

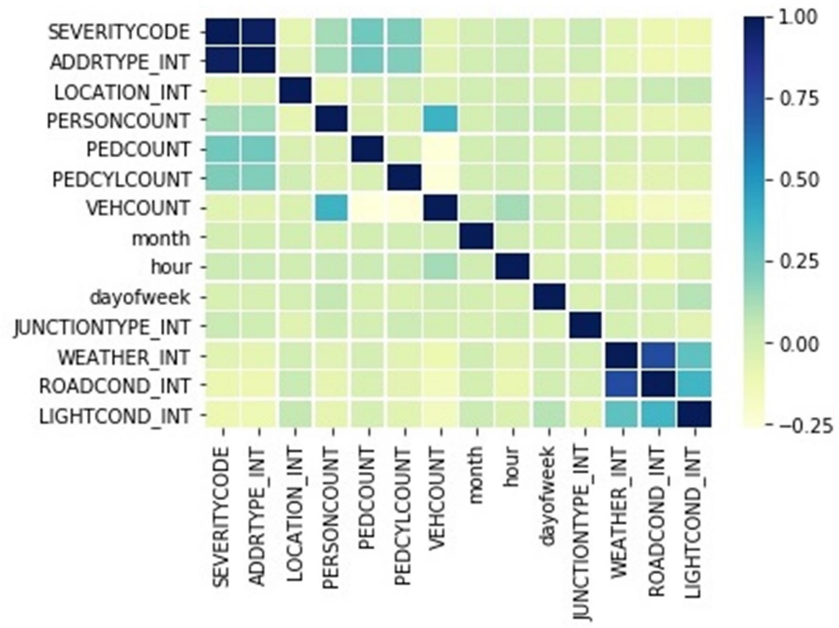

Fig 10: Collisions dataset correlation heatmap

Results

From the above heatmap, one can observe the correlation between severity code and address type as most of the injury collisions happen in address types blocks and intersection while nearly three-fourths of the property damage collisions happen in blocks. Also there is a correlation between person count and vehicle count which is not surprising as more the number of vehicles, the more the number of likely passengers and people

involved in accident. Also weather, road conditions and light conditions are strongly correlated which is unsurprising as weather impacts road conditions as well as light conditions.

Discussion

From the analyzed data one can observe that:

- The likelihood of property damage collisions is twice as more likely than injury collisions.
- Collisions are more likely to happen under high visibility conditions like daylight or dark with streetlights.
- Collisions are more likely when two vehicles are involved.
- Most accidents occur on a Friday. Hence traffic officials and people need to be more vigilant on that day of the week.
- Both injury and property damage only collisions are frequently occurring at Intersection (intersection related), Mid-Block (not related to intersection) and Mid-Block (but intersection related). Highly visible warning signs and traffic signals need to be put up at such locations.
- There are some locations which are more prone to injury collisions but property damage related collisions seem to be evenly distributed between locations. Proper and highly visible warning signs and traffic signals needs to be put up at accident prone locations.
- Accidents happen more commonly during clear weather, followed by rainy and overcast weather. The road conditions also vary accordingly. Hence, the driver needs to be more vigilant during such weather and road conditions.

Conclusion

Drivers, people and traffic officials need be more vigilant about accidents at certain locations, address types, weather, junction-types and on particular days of the week. Accordingly, resources of the traffic department can be judiciously allocated so that the number of traffic accidents can be reduced. Additionally studies may be done to analyze why certain locations are prone to accidents, for instance whether it is due to faulty road design, alignment, etc and how the accidents at such locations can be reduced.