Tens of thousands of people die on American roads every year. Bringing that number down to zero by 2050 is possible. We would have to change how we think about road safety, stop accepting car crashes as accidents, and make smart investments in technology.

1. Introduction

There is a huge impact on society due to traffic accidents where there are great costs of fatalities and injuries. In recent years, there is an increase in the research's attention to determine the significantly affect the severity of the driver's injuries which are caused due to road accidents accurate, and comprehensive accident records are the basis of accident analysis. the effective use of accident records depends on some factors, like the accuracy of the data record relation, and data analysis. There are many approaches applied to this scenario to study the problem. A recent study illustrated that the residential and shopping sites are more hazardous than village areas might have been predicted, the frequencies of the casualties were higher near the zones of residence possibly because of the higher exposure. A study revealed that the casualty rates among the residential areas are classified as relatively deprived and significantly higher than those from relatively affluent areas.

accidents have become very common these days. Nearly 1.25 million people die in road crashes each year, on average, 3,287 deaths a day. Moreover, 20—50 million people

are injured or disabled annually. Road traffic crashes rank as the 9th leading cause of death and accounts for 2.2% of all deaths globally. In this contest to reduce the severity of the accidents machine learning and neural techniques have been used for analysis. Car accidents are one of the common types of Collisions occurring everywhere globally every day. By analyzing the different factors which cause the collision. In this section, we are discussing the data Capstone project topic Car accident severity.

```
In [39]: df.corr()
```

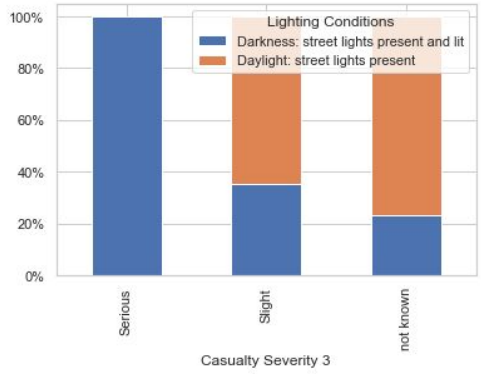
Out[39]:

	Grid Ref: Easting	Grid Ref: Northing	Number of Vehicles	Accident Date	Time (24hr)	Age of Casualty 1
Grid Ref: Easting	1.000000	0.007166	0.094315	0.151429	-0.045607	-0.132149
Grid Ref: Northing	0.007166	1.000000	-0.029322	-0.073697	0.051562	-0.092974
Number of Vehicles	0.094315	-0.029322	1.000000	-0.042695	0.044298	-0.004039
Accident Date	0.151429	-0.073697	-0.042695	1.000000	0.040475	-0.067686
Time (24hr)	-0.045607	0.051562	0.044298	0.040475	1.000000	-0.181658
Age of Casualty 1	-0.132149	-0.092974	-0.004039	-0.067686	-0.181658	1.000000

Casualty Severity 3

```
In [44]: import matplotlib.ticker as mtick
         import matplotlib.pyplot as plt
         df.groupby(['Casualty Severity 3', 'Lighting Conditions']).size().groupby(level=0).apply(
             lambda x: 100 * x / x.sum()
         ).unstack().plot(kind='bar', stacked=True)
         plt.gca().yaxis.set_major_formatter(mtick.PercentFormatter())
         plt.show
```

```
Out[44]: <function matplotlib.pyplot.show(*args, **kw)>
```




```
In [ ]:
```

Modeling

The machine learning methods ave thrived in the applications of language and text modelling -in -recent years which can potentially counter the challenges In processing and classifying the tweets. In most of the studies, language modelling can be taken as a kind of information extraction from the text messages, which is the process of converting¶ the unstructured text information into a structured database and solving it as a supervised or unsupervised learning task. The limited word features can be utilized for specific research.

Data modelling and deployement

For the current project KNN and Decision tree classifier has been used car accident servity. c) KNN: KNN is a classification algorithm which is based on feature similarity. It analyses the data and measure the distance and similarities between data and cluster them based on K values. Distance is calculated in man ways, for this research (Labib 2019), we used Euclidean distance measurement. The class of new in closest one. The posteri9r p true th 'b" is true. Decision tree classification In general decision tree classifier has goOdJuccuracy. ki)free induction is a typical inductive approach to learn knowledge on glassitjcaiior.pecision tregs slassify Ostaqees by sorting them down the tree from the root to some leaf node, which provides the cl"sification of the instance. Decision Trees usually mimic human thinking ability while making a decision, so it is easy to understand. The logic behind the decision tree can be easily understood because it shows a tree-like structure.

In [ ]:

In [43]:
```
import matplotlib.ticker as mtick
import matplotlib.pyplot as plt
df.groupby(['Casualty Severity 3', 'Weather Conditions']).size().groupby(level=0).apply(
    lambda x: 100 * x / x.sum()
).unstack().plot(kind='bar', stacked=True)
plt.gca().yaxis.set_major_formatter(mtick.PercentFormatter())
plt.show
```

Out[43]: <function matplotlib.pyplot.show(*args, **kw)>