

**Capstone Project - Car accident severity**

**Date: 9/10/2020**

**By: Taha Ansari**

|  |  |
| --- | --- |
| **Content** | **Page number** |
| **Introduction** | **3** |
| **Data** | **4** |
| **Methodology** | **5** |
| **Results** | **8** |
| **Discussion** | **9** |
| **Conclusion** | **10** |

**Introduction**

**Background information**

Road accidents are a very common problem in todays day and age. The severity of accidents varies from minor(which can result in a scratch on your car) to fatal(which can lead to a loss of life). I think we can all agree that we would all love to prevent accidents from occurring and so I would like to share with you my aim for this project. My aim is to build a model that can help us understand the severity of an accident and why it has happened.

**Who is the target audience?**

The target audience that my model would be helpful would be the local authority. This is because they can find the cause of accidents and find ways to prevent them from occurring.

**What are the aims of this project?**

The aim of this project is simple.

1. Find out whether accidents are increasing or decreasing
2. Find out whether accidents are happening on a certain day/ certain time
3. Find out whether accidents are happening at a certain location
4. Find out how severe accidents are and how we can prevent them from happening in the future.

**Data**

**What data is required?**

The data that is required is a database showing the traffic accidents from one time period to another. We need the data to be relatively new and the time period to be large. The reason for a large time period would be because we need to observe trends of the data over time. The bigger the time period we have, the easier it may be to observe any apparent trends.

**Where is the data from?**

We will use the data from a database that contains information about UK car accidents from 2005-2015.

**What will it contain?**

The database contains:

* Number of accidents
* Number of accidents on each day of the week
* Number of casualties
* Type of casualties
* and much more

**Methodology**

Our first step was to import the library and clean the data. If the data was cleaned, it would be easier to use which was very important.

The task was then to compress the data and visualize it so that it would be much easier to read. Watching the videos that were set in IBM Data Science, we remember that Darkhorse analytics was mentioned. Their motto was to visualize data that would be easier to the target audience to read, and that was exactly what I was aiming to do.



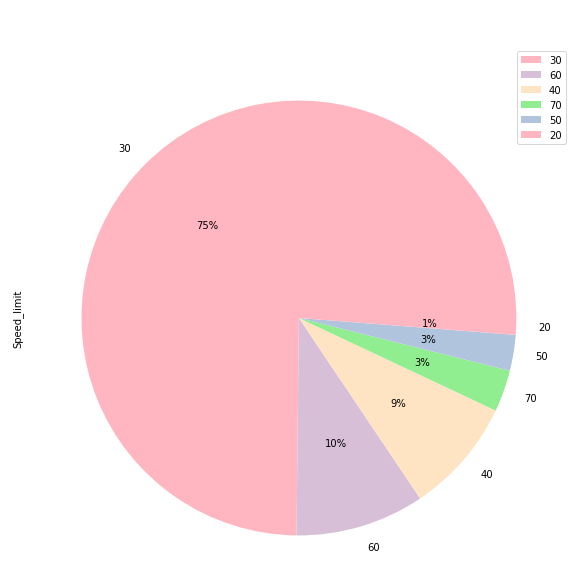
**\***the code above imported all the relevant libraries and cleaned data

Now that the code was imported and cleaned we now had to start the data visualization process.

Our first step was to find out the total number of people included in the data. As seen in the notebook that was 1,045,076.

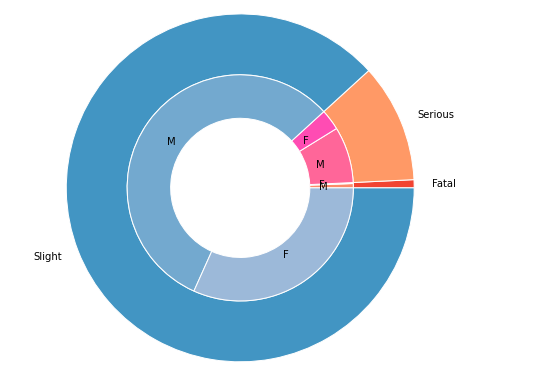
We then started visualizing data to help us answer some of our important questions. As seen in our notebook it shows how many accidents happened each year, what days they happened on, etc.

When observing the data there are elements of it that may come shock us. The first element that shocks us is the percentage of accidents in a certain speed zone.



This surprises us as most accidents are happening when the speed is comparatively slower( 75% of accidents were happening when the speed limit was 30). Most people would assume that the majority of accidents would be happening at faster speed zones, but this is not the case.

Data was continued to be visualized using simple to view methods such as bar and pie charts.   
  
Towards the end I wanted to find more correlations between the data, thus I decided to make the charts more informative.



\*The chart above shows the severity of accidents and which gender had caused which type

**Results**

The results have shown us a lot of things. There are a lot of surprises in these results. For example: Most accidents happened on Friday, whereas the least amount of accidents happened on Sunday. This may be that People are going out to places late night on Friday and are too tired to drive back, so they aren't as alert or that they are rushing home as soon as possible from work. The results are important as we understood how severe the accidents were and where most of these accidents are happening

**Discussion**

I've observed that the large amount of accidents happened in single carriageways (over 700,000). I would recommend local authority to review footage of these roads and inspect them further incase there as a more underlying reason for this. I think that there should be a bigger campaign for driving safely on Fridays as there is definitely a reason for why most accidents happen on that day. I also think people should be taught to be more aware, this is because 75% of accidents are happening in a speed zone of 30. This means that people are not as aware on slower roads, thus a campaign about this would be very beneficial or more training in driving schools regarding this matter.

**Conclusion**

Thanks to the database we were able to extract lots of information that will be valuable to the local authorities. Using this data they may decide to take action via campaign or more training, or even look at roads and see how they can be improved. Although crashes are not fatal, the information can be used by local authority to find ways to reduce accidents and even help predict the trend of whether they will increase or decrease in the years to come.