INTRODUCTION

Road accidents are an issue of concern in every part of the world. We all know that some countries have a high percentage of deaths and injuries caused by road accidents. There are several factors on which it is dependent, and every year the government checks the data based on these cases, including every factor that is related to an accident. In this project, STATS19, which is a data set released by the government, has been analysed. STATS19 is strongly associated with road crashes that happened throughout the UK. In this report, data on accidents, vehicles, and casualties will be examined and visualised to determine the reason for these accidents and later come up with a conclusion using proposed ideas for how they can be reduced. All the ideas that will be discussed in this report to reduce accidents will have public interest so that the public can be aware and do more safety measures on the basis of what is going to be explained in this project, and accident severity can be reduced.

EVALUATION OF DATA SETS

Evaluation on data sets and analysis is based on the following factors in the accidents and casualties data for 2016:

- Number of vehicle
- Place of accident
- Time of accident
- Speed of vehicle
- Speed limit of particular place
- Accident severity
- Vehicle Age

ETHICAL, REGULARITY AND PRIVACY ISSUES

- ETHICAL ISSUES: When dealing with big data, it is important to keep ethical issues in mind, which include systemizing, defending, and recommending concepts of right and wrong in relation to data.

- REGULARITY ISSUES: Regulatory issues are a concern when dealing with data that has been released by an authorised body to provide knowledge about some facts that can be helpful to others. It is important that human and organisational rights are not violated when working with this kind of data. In our project, the data provided is vast and was released by the UK government, for which the regulatory laws should be in place while we work on it without violating any kind of knowledge.

- PRIVACY ISSUES: Data privacy is an issue that must be handled properly and not shared with anyone in big data projects, where you are dealing with a massive amount of data related to large organisations and government bodies such as the UK government, as it includes massive amounts of data about a country, including individual's data. There needs to be consent, notice, and legal use of data while working on it.

DATA VISUALISATIONS AND ANALYSIS

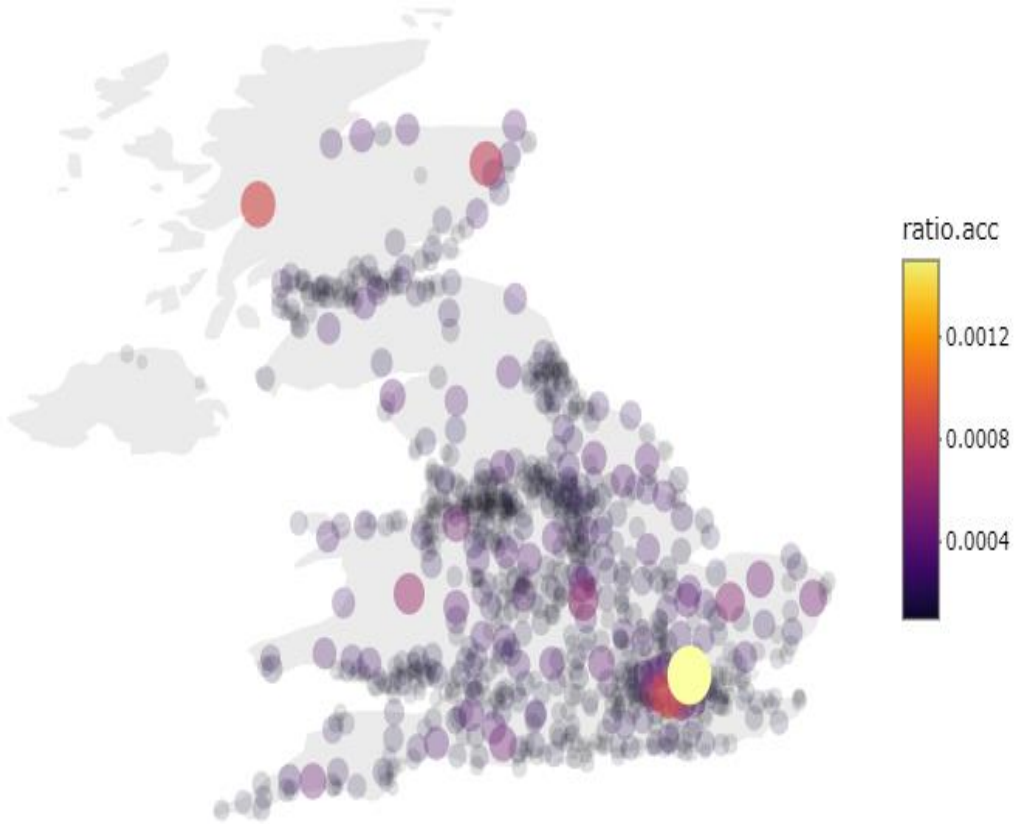
In this part, we are required to come up with three ideas or questions that may help us and the public get an idea of what are the major reasons for accidents and casualties on the road, using the road safety data on which this project is based.

- IDEA 1: Analysis of pedestrians and cyclists, how vulnerable they have been in accidents.
- IDEA 2: Highest accidental region (Top 3) and Correlation of highest contributing factors with speed limit. Also, Compare them over the years with respect to speed limit? Has road safety improved?

IDEA 1: Analysis of pedestrians and cyclists, how vulnerable they have been in accidents.

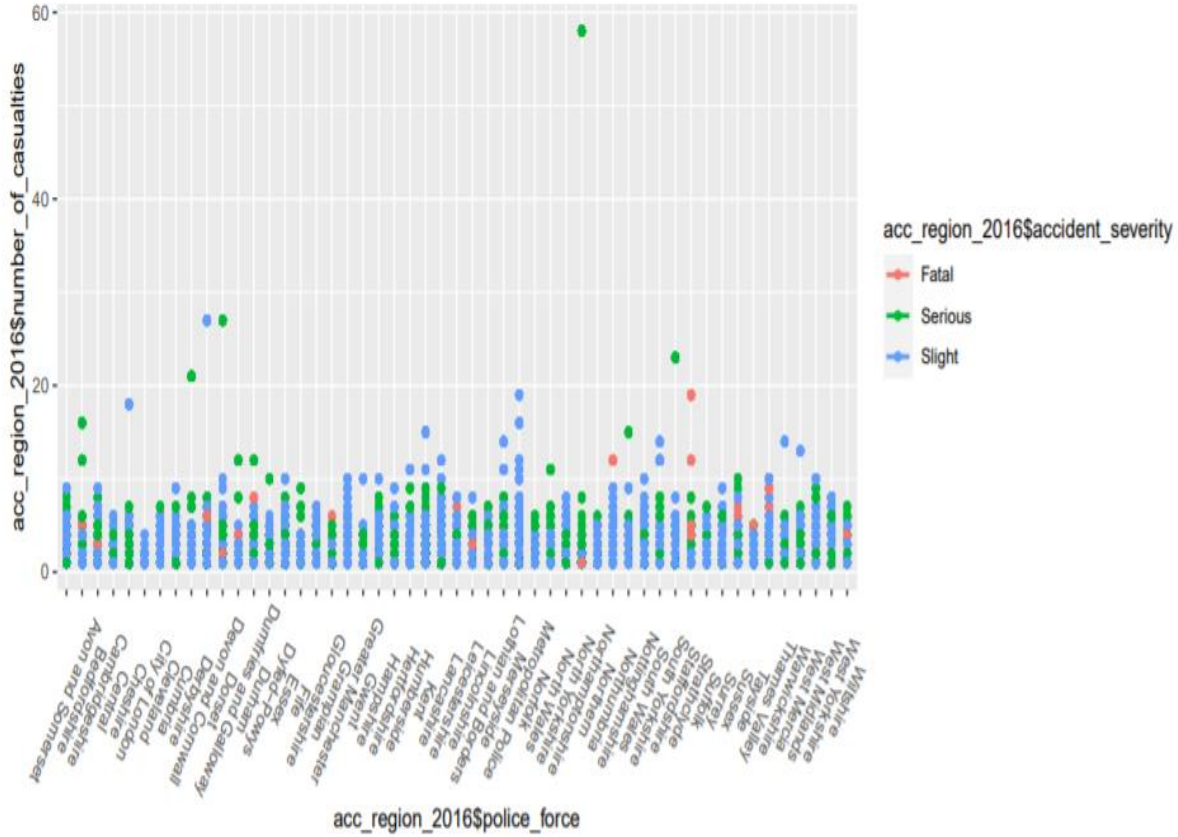
- As we are aware, the most vulnerable road users to accidents are cyclists and pedestrians, with higher chances of severity during the accident because they have no protection while using the road. A speed limit of 40 and less than 40 is taken because those are the only roads which is used by pedestrians and cyclists.

- From the plot above, it can be seen where and how badly pedestrians and cyclists are prone to accidents. Most of the incidents happened in the midlands and south of England, where more strict measures need to be put in place for the safety of these users.



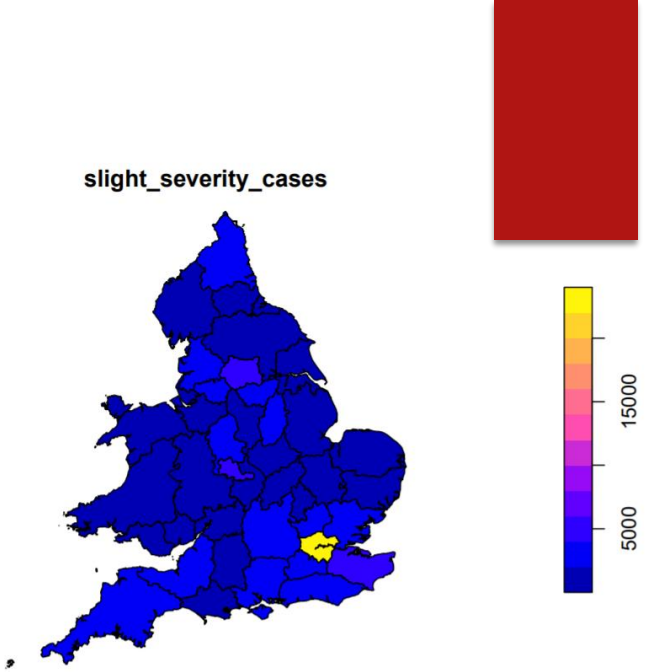
IDEA 2: Highest accidental region (Top 3) and Correlation of highest contributing factors with speed limit. Also, Compare them over the years with respect to speed limit? Has road safety improved?

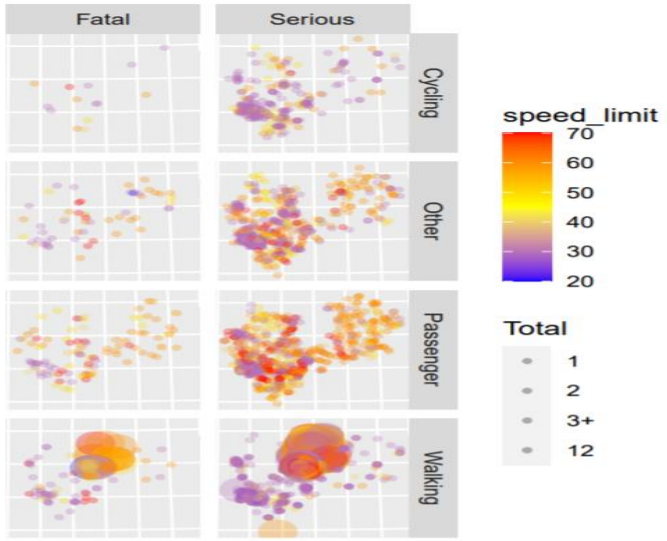
- In this idea we have plotted a gg plot to see where are most highest taking place and on what seep limit and what was the severity of accidents, is it fatal, serious or slight.
- According to the above plot, the top three regions with the highest accidental count are North Yorkshire, Devon and Cornwall, and Derbyshire.



OTHER ANALYSIS

- Some basic plots give ideas about fatal and serious accidents that happened in 2016 involving passengers, cyclists, pedestrians, and others, with a heat map and at different speeds.
- Most vulnerable and people involved in accidents are Walking and passengers on a speed more than 40 mph.
- Another plot for country wide map showing which place has more severity of the cases in 2016 is below:





CONCLUSION

After thoroughly considering and evaluating all the data sets from Stats19 for 2016 and 2020, it can be concluded that road safety has increased over the years. According to the statistics, pedestrians and cyclists appeared to be the most susceptible to accidents among the others, which has gradually improved as the accident rate has appeared to be minimal. With more safety measures on the road by the public and government, the rate of accidents, or at least their severity, can be reduced.

FUTURE SCOPE

From all the analysis done so far, there are many things that can be done using this data or different data from Stats 19. There are several factors that need to be focused on to extract more helpful information that can be more beneficial for the public and improve road safety.

Several Future Scope：

- The age factor needs to be considered when comparing the severity of accidents.
- Accidents are more fatal when young drivers are on road- could be a reason they have lack of experience in driving
- Data for roundabouts and junctions should be analysed because they are the most critical points on the road.
- For public interest, more strict laws and measures need to be imposed to road more safe.
- analysing the top 10 prone areas, examining their numbers, and imposing more careful road rules in those areas.