

COCHIN UNIVERSITY OF SCIENCE AND TECHNOLOGY

Department of Computer Science

September 18, 2022



ACCIDENT DATA ANALYSIS AND GAUSSIAN GRAPH PLOTTING

Submitted by :

Akshadha.A 80521004

Namitha.M.S 80521013

Submitted to :

Dr. Madhu.S.Nair

Department of Computer Science

ACKNOWLEDGEMENT

We would like to express our heartfelt gratitude to Mr Madhu S Nair, our project guide, for his invaluable guidance and constant support, along with his capable instruction and persistent encouragement.

We are grateful to our Head of Department, Dr. Philip Samuel for his steady support and for the provision of every resource required for the completion of this project.

We would like to take this opportunity to thank all other department faculties for having designed an excellent learning atmosphere. Our thanks are due to all others for providing us with the help required to carry out the groundwork of this project.

ABSTRACT

This project is an analysis of datasets obtained regarding previously occurred accidents in a specific region from which inferences are drawn. The main aspects of the project converge on the inferences drawn using the different machine learning approaches on these datasets. The solution to issues faced in maintaining safety on roads also lies in the selection of which specific problems are to be tackled in which regions based on the data collected.

Problem Statement: “To analyse the given accident data from a particular region and to obtain inferences regarding the influences and contributions of several attributes to the severity of an accident and hence provide specific measures in order to reduce the rate of these accidents”

Also in the end Gaussian graph is plotted for the accident data set which makes it easy comparison of total cases, injuries and deaths. iii

Contents

1	Introduction	1
1.1	Scope	1
1.2	Objectives	1
1.3	Tools	1
2	Software and Tools	2
2.1	Software Requirements	2
2.2	Functional Requirements	2
2.2.1	Data collection	2
2.2.2	Handling Data	3
2.2.3	Scalability	3
2.2.4	Libraries used	3
2.3	System design	4
2.4	Project Outcome	4
2.5	Result	5
3	Conclusion	10
4	References	11

1 Introduction

1.1 Scope

The purpose of Accident data analysis is to find possible causes of accident-related to driver, vehicle, and roadway. Accident analyses are made to develop information such as drivers, pedestrians, vehicle, and roadway.

Scene management and scene assessment (secure the scene, make sure it is safe for investigators to do their job), Witness management (provide support, limit interaction with other witnesses, interview), Investigate the incident, collect data, Analyze the data, identify the root causes are the major steps in accident analysis.

1.2 Objectives

Based on the survey, 1.3 million people die every year due to roadway accidents.

The two objectives of the smart accident management system are

- 1) accident analysis.
- 2) accident prevention.

1.3 Tools

This program has been totally developed on google colab platform by creating a jupyter notebook by using pandas library and many other graph plotting tools.

2 Software and Tools

In this section, tools used and several related concepts will be explained for better understanding the process of project.

2.1 Software Requirements

Jupyter Notebook : The Jupyter Notebook is an open-source web application that allows to create and share documents that contain live code , equations , visualizations and narrative text.

Uses include: data cleaning and transformation, numerical simulation, statistical modeling, data visualization, machine learning, and much more.

2.2 Functional Requirements

2.2.1 Data collection

1. Severity of accident with respect to light conditions and alcohol consumption : We use modules and a decision tree to find out how light conditions and alcohol consumption as independent factors affect the severity of accident. It is often seen that light conditions in area and with cases in which alcohol consumption is involved, the accidents can be more severe. This analysis can help us reduce the amount of accidents that are highly severe.

2. Severity of accident with respect to day of the week, light conditions and speed zone: A decision tree helps us to tell us how the day of the week, light conditions and speed zone as independent factors

affect the severity of accident. The speed zones of areas often attribute to how severe accidents can be. When factors like light conditions and day of the week are considered, we see that drivers can be reckless in certain areas and/or on certain days. The information extracted from this analysis would help in taking proactive measures.

2.2.2 Handling Data

The data requirements are very simple and direct. The data is to be supplied to the model in the form of a dataset which will be worked upon using the Pandas library. A database may be used to retrieve the data in which case we may use an SQL based framework, however since this is a purely machine learning-based project we may use data available directly for the model to work upon.

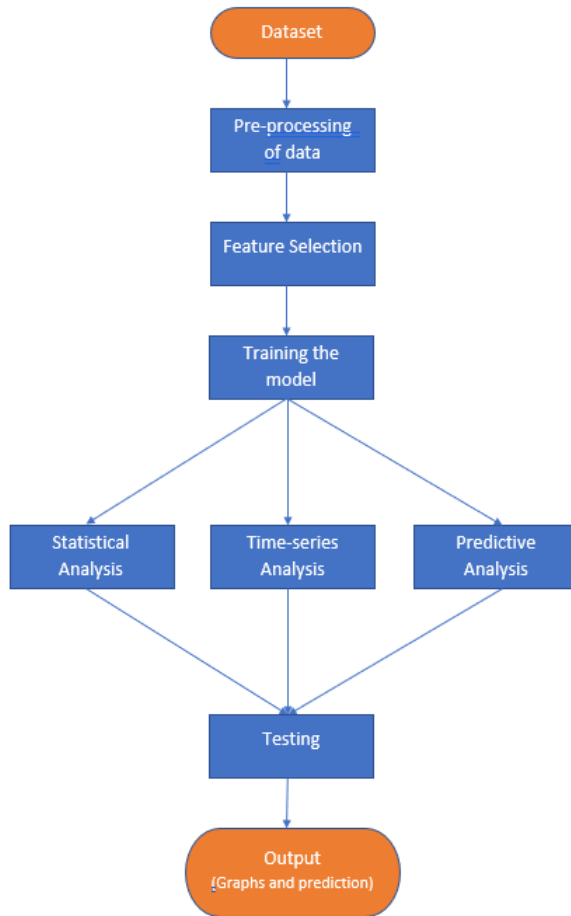
2.2.3 Scalability

The scalability of the model depends on the amount of data provided to be studied. The model is only as accurate in prediction as the quality of the data set provided. With more data a better working model can demonstrate more accurate predictions.

2.2.4 Libraries used

- 1.pandas
- 2.numpy
- 3.matplotlib

2.3 System design



2.4 Project Outcome

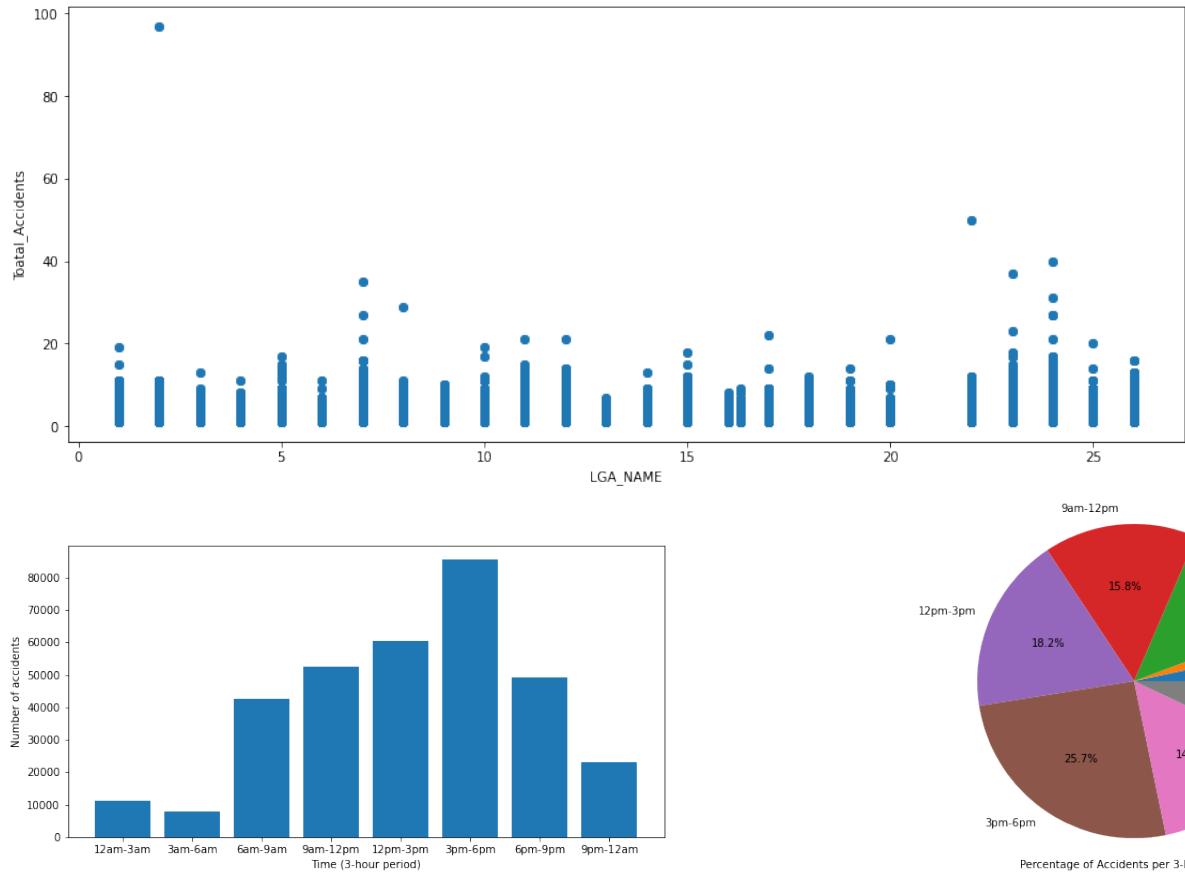
The project contain tow modules.

In first module we set an accident data set which contain many attributes like accident day,time,type of vehicle,accident reasons like light condition ,alcohol consumption etc.Then we do calculate and plot countplot type of graph which make it easier to understand the data in a better way.Countplots between light condition and severity,severity and alcohol time,light condition and alcohol time,severity and speed zone,police attend and speed zone ,female and severity are plotted in

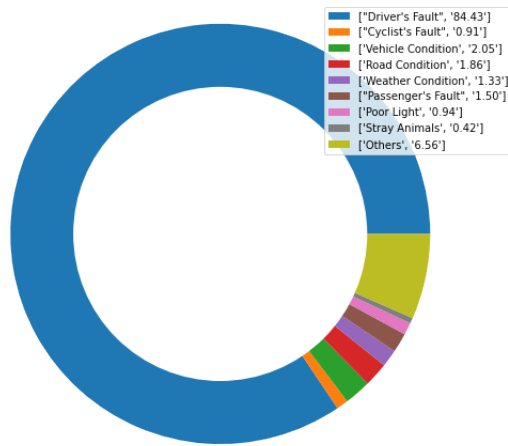
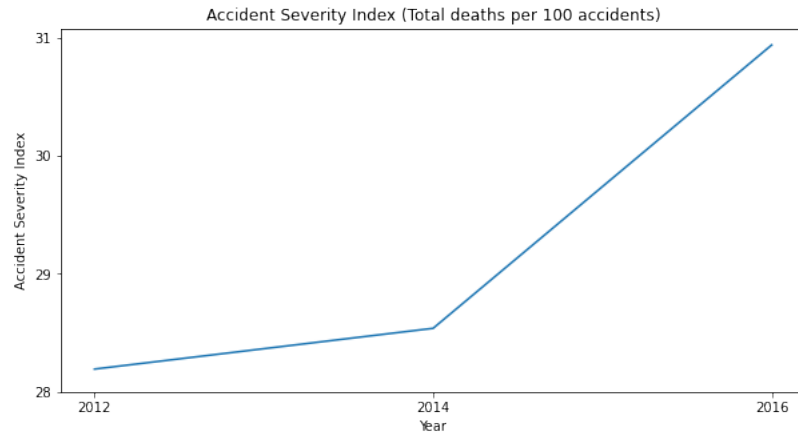
this first part of project.

In second module we are trying to plot Gaussian graph or normal graph. for this we extracted a small part of the first dataset to data regarding five states in india and tried to plot normal graph between cases ,injuries and deaths.

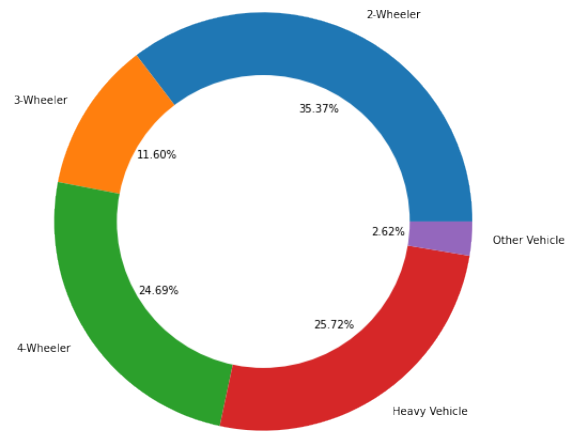
2.5 Result



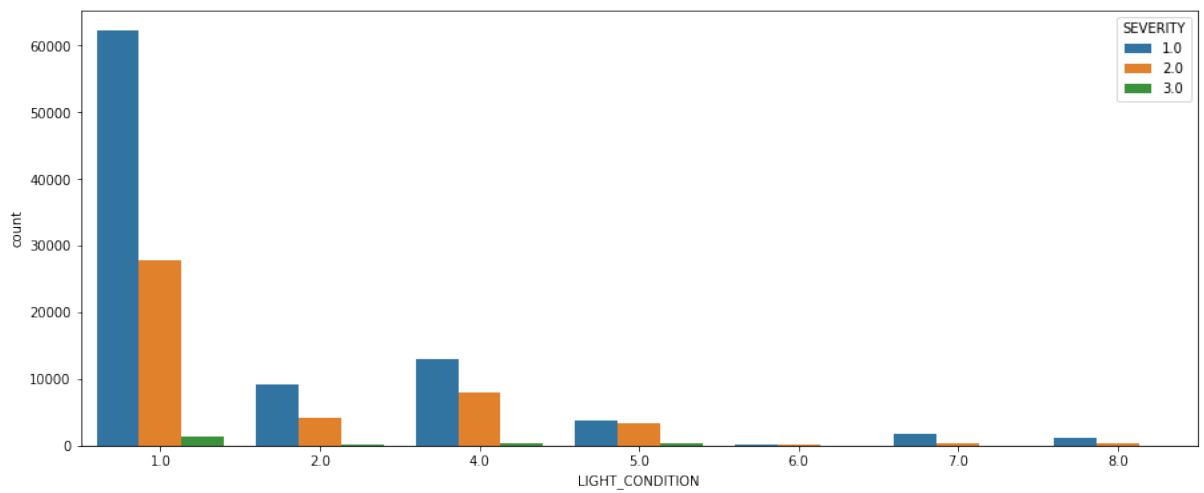
2 Software and Tools



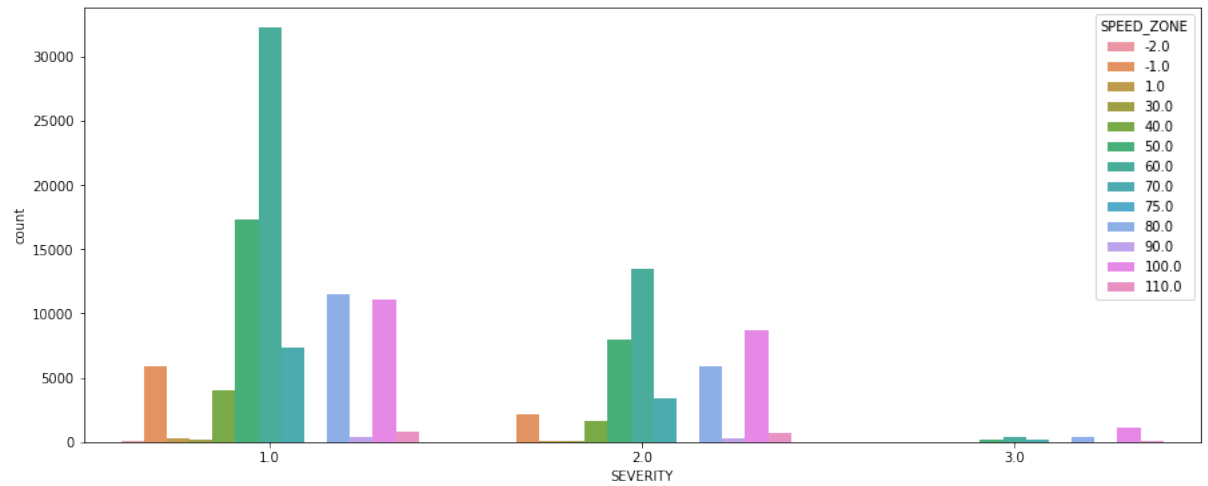
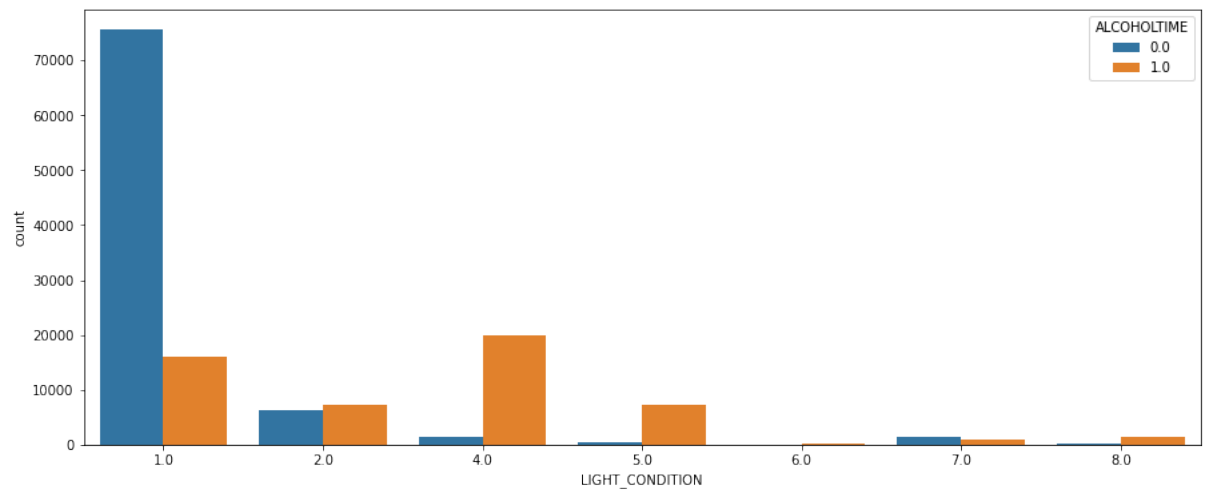
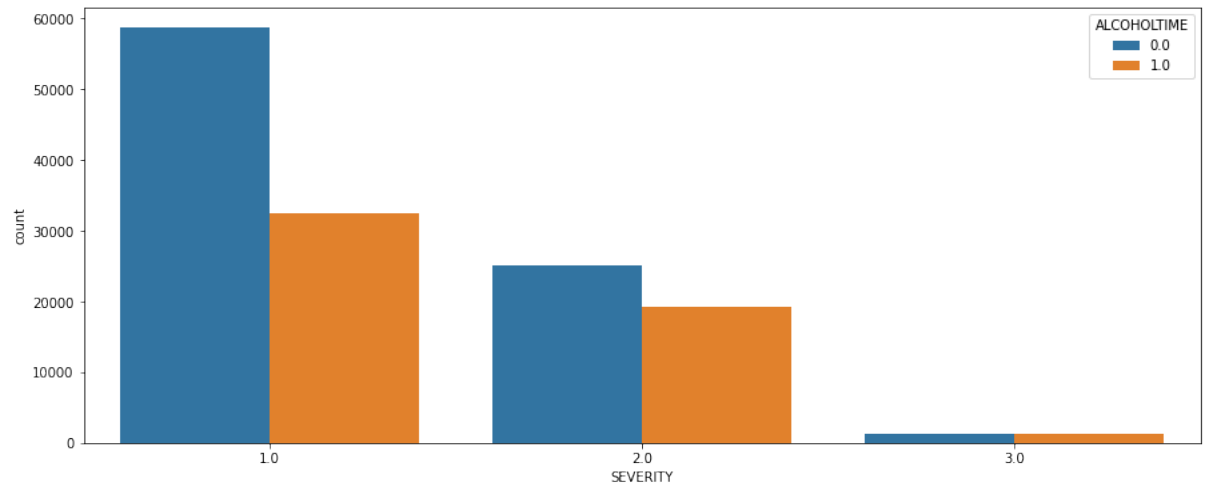
Accidents in 2014



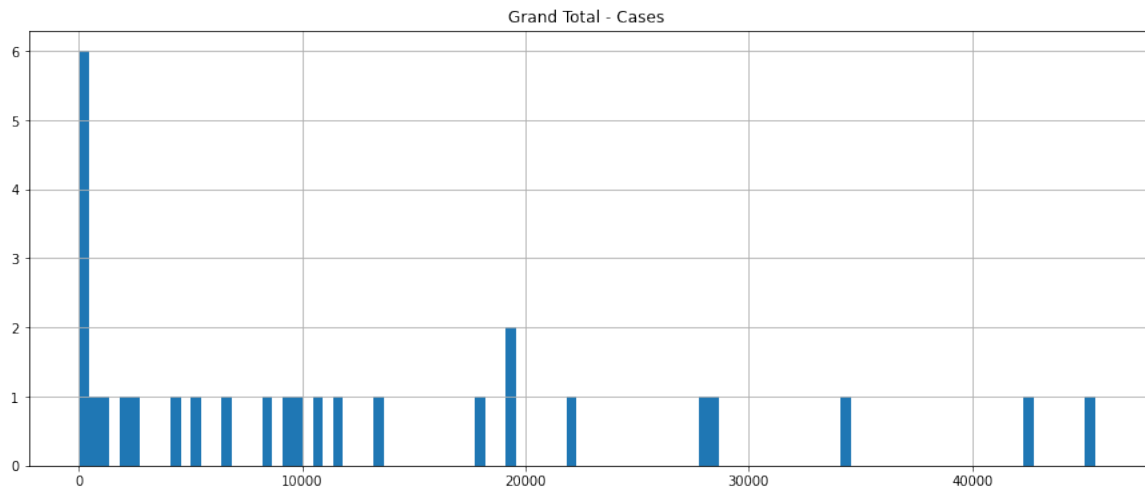
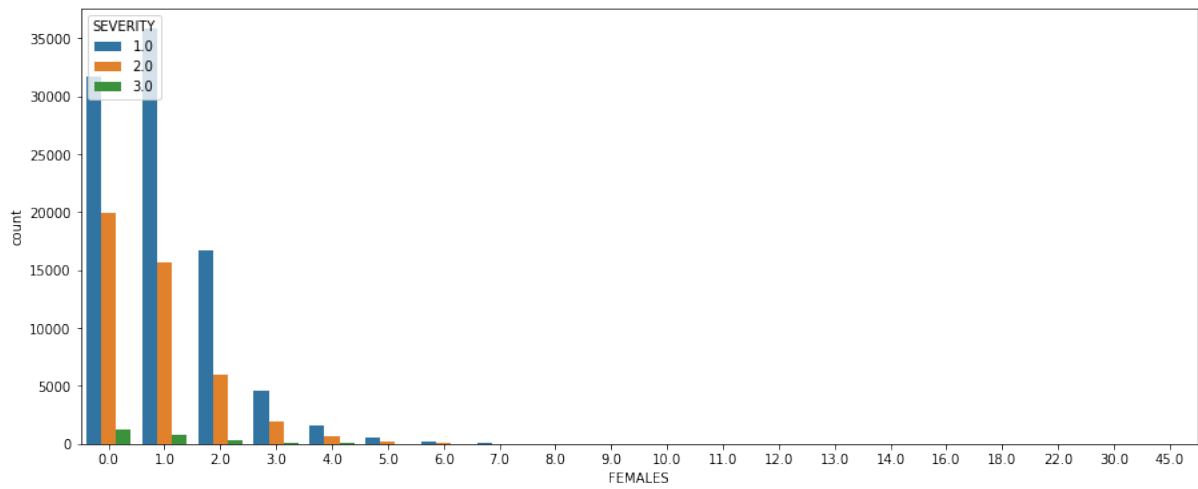
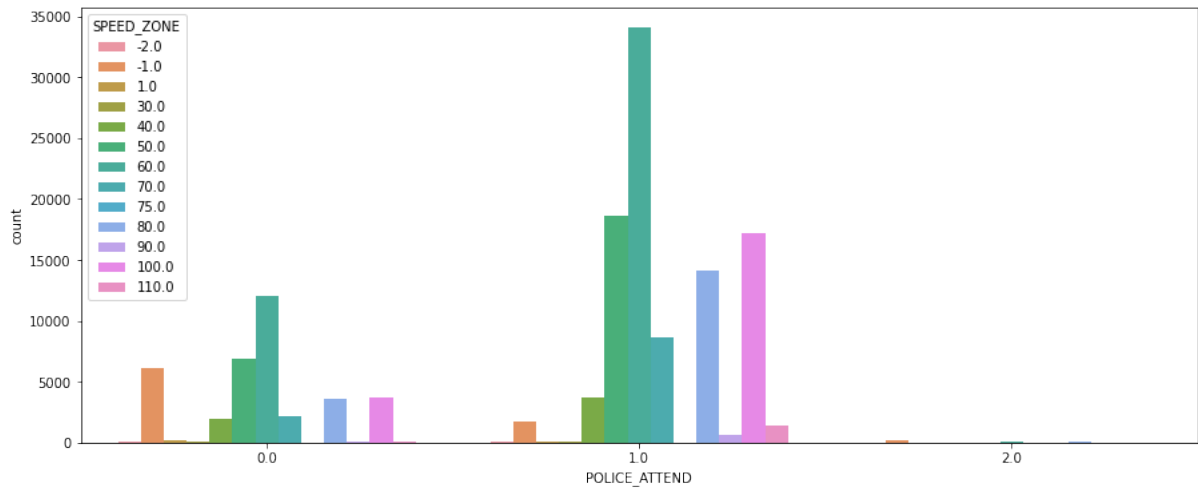
Types of Vehicles Involved in Accidents in 2016



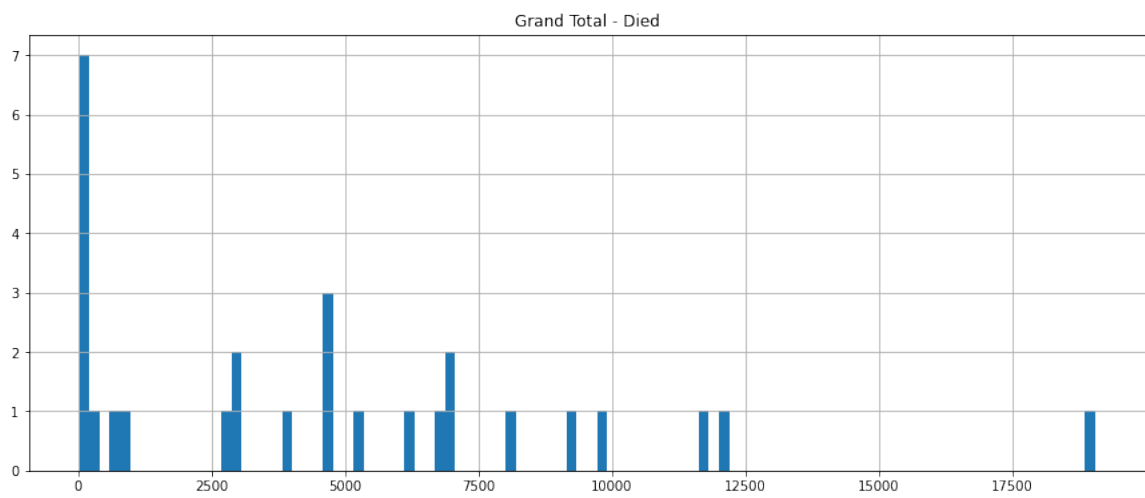
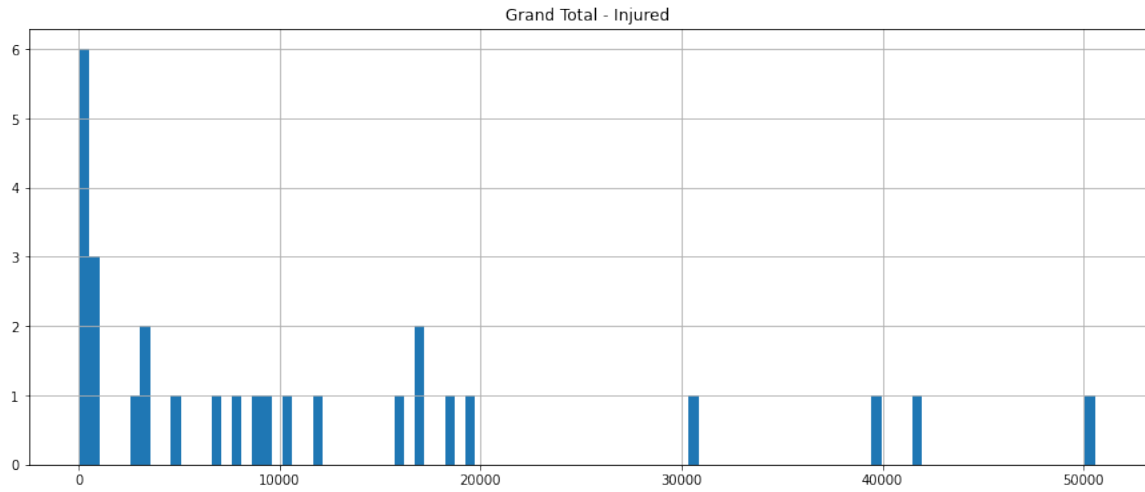
2 Software and Tools



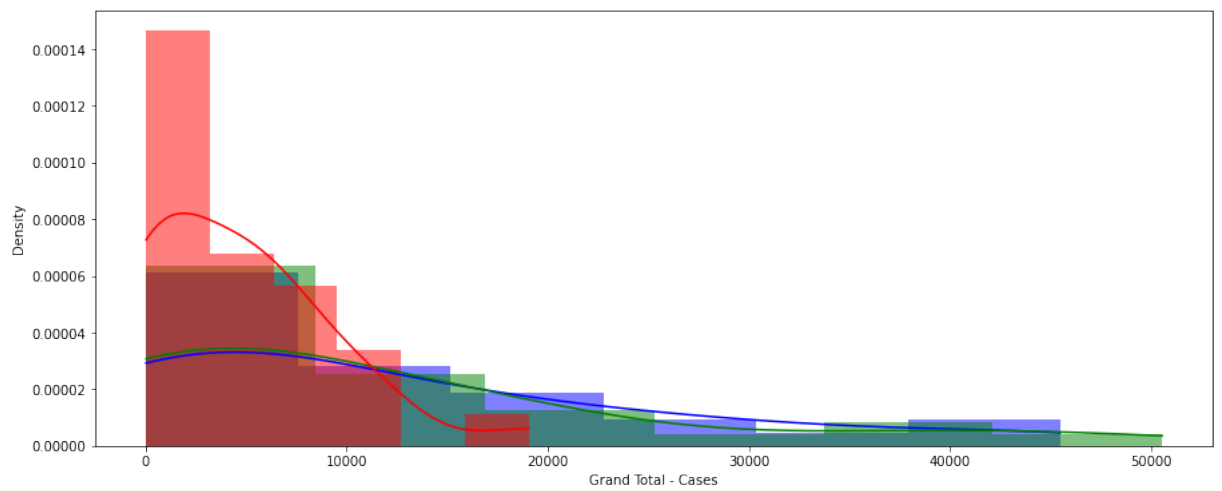
2 Software and Tools



2 Software and Tools



Gaussian plot between total cases.injuries and deaths.



3 Conclusion

Through this project we are able to analysis a set of data set easily and pot the graph for the same without much effort. Graphs are easily understood and can be used for certain study regarding accidents and are helpful in preventing further accidents.

4 References

1. Python Programming by John Zelle.
2. Grokking Algorithms by Adithya Bhargava
3. www.koolplot.com
4. www.cplusplus.com
5. www.geeksforgeeks.com
6. www.quora.com