

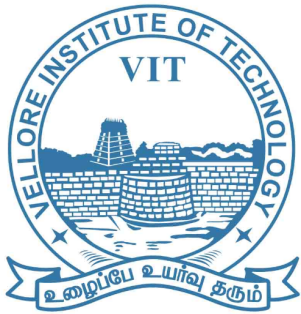
ANALYSIS OF ROAD ACCIDENTS USING DATA VISUALIZATION TECHNIQUES

DATA VISUALIZATION (CSE3020)

J COMPONENT
REVIEW REPORT

TEAM MEMBERS

Student Name	Registration Number
Y D S S Prathyusha	18BCE2051
Siddhant Chourasia	18BCE0275
Vaibhav Singh	18BCE2313



VIT[®]

Vellore Institute of Technology

(Deemed to be University under section 3 of UGC Act, 1956)

Submitted to

Prof. Meenakshi S.P

LIST OF TOPICS

Chapter No	Topic	Page No
1	INTRODUCTION	
	1.1 Abstract	3
	1.2 Introduction	3
	1.3 Objectives	3
2	PROBLEM STATEMENT	3
3	DATA COLLECTION	
	3.1 Dataset view	4
	3.2 Description	5
4	TASKS, SYSTEM DESIGN & TOOLS USED	5
5	ACTIONS	7
6	IMPLEMENTATION	7
7	VALIDATION	
	7.1 Domain level	22
	7.2 Task level	22
	7.3 Data level	22
	7.4 Algorithm level	22
8	SUGGESTIONS/ CONCLUSION	23

1.INTRODUCTION

1.1 ABSTRACT

Roadway traffic safety is a major concern for transportation governing agencies as well as ordinary citizens. In order to give safe driving suggestions, careful analysis of roadway traffic data is critical to find out variables that are closely related to fatal accidents. The relationship between fatal rate and other attributes including collision manner, weather, surface condition, causality severity, and driver age were investigated.

Association rules were discovered by Apriori algorithm, classification model was built by Naive Bayes classifier, and clusters were formed by a simple K-means clustering algorithm. Certain safety driving suggestions were made based on statistics, association rules, classification model, and clusters obtained.

1.2 INTRODUCTION

Investigations of the high-risk areas for road traffic crashes (RTCs) are urgently needed to guide improvements in road safety. In this paper we apply statistics analysis and data visualisation algorithms on the FARS Fatal Accident dataset as an attempt to address this problem. In this paper, the relationship between fatality rate and other attributes including collision manner, weather, surface condition, causality severity, and driver age were investigated.

1.3 OBJECTIVE

- Our Objective is to arrive at a conclusion regarding the problem of Roadway traffic safety and to give safe driving suggestions.
- .Finding Numbers of accidents happening in rural and urban areas.
- Finding the trend in the number of accidents that occur each year.
- .Finding the most dangerous time to drive.
- Finding What fraction of accidents caused minor injuries , major injuries and deaths.
- Finding how fast the number of car accidents drop off with age.
- Finding the ratio of men and women who get injured in accidents.

2. PROBLEM STATEMENT

Roadway traffic safety is a major concern for transportation governing agencies as well as ordinary citizens. In order to give safe driving suggestions, careful analysis of roadway traffic data is critical to find out variables that are closely related to fatal accidents. India is a country

having a high usage of vehicles. The vehicle consumption has drastically increased in the last 40 years from 6 million to 230 million vehicles. Due to the increasing rate of **9%** vehicles per year, the occurrence of road accidents has increased exponentially which in turn has hampered the road security of the people in India.

3. DATA COLLECTION

3.1 DATASET VIEW

Attribute	Attribute type	Details
Accident_Index	Ordinal	Unique identifier of the accident.
Date	Ordinal	Date of the accident.
Day_of_Week	Categorical	Day of the accident.
Time	Ordinal	Time of the accident, in 24 hours format.
Longitude	Quantitative	Longitude coordinate of the accident.
Latitude	Quantitative	Latitude coordinate of the accident.
Number_of_Casualties	Quantitative	Number of casualties resulted from the accident.
Number_of_Vehicles	Quantitative	Number of vehicles involved in the accident.
Accident_Severity	Quantitative	3 levels of severity (Fatal, Serious and Slight)
Speed_limit	Quantitative	Speed limit imposed on the stretch of road, if applicable.
Road_Type	Categorical	6 types of road (Single carriageway, Dual carriageway, Roundabout, One way street, Slip road, Unknown)
Pedestrian_Crossing-Human_Control	Categorical	3 categories (None within 50 metres, Control by other authorised person, Control by school crossing patrol)
Pedestrian_Crossing-Physical_Facilities	Categorical	6 categories (No physical crossing within 50 meters, Pedestrian phase at traffic signal junction, non-junction pedestrian crossing, Zebra crossing, Central refuge, Footbridge or subway)
Light_Conditions	Categorical	5 categories (Street light present, Street light present and lit, No street lighting, Street lighting unknown, Street lights present but unlit)
Weather_Conditions	Categorical	9 categories (Fine without high winds, Raining without high winds, Raining with high winds, Fine with high winds, Snowing without high winds, Fog or mist, Snowing with high winds, Unknown, Others)
Road_Surface_Conditions	Categorical	5 categories (Dry, Flood, Frost, Snow, Wet)
Special_Conditions_at_Site	Categorical	8 categories (Auto traffic signal partly defective, Auto traffic signal out, Mud, None, Roadworks, O1 or diesel, Road surface defective, Permanent sign or marking defective or obscured)
Carriageway_Hazards	Categorical	6 categories (None, Other object in carriageway, Any animal (except a ridden horse), Pedestrian in carriageway (not injured), Involvement with previous accident, Dislodged vehicle load in carriageway)

[illegible]

Sl. No	States/UTs	Fault of Dr	Fault of Dr	Fault of Dr	Fault of Cy	Fault of Cy	Fault of Cy	Fault of Dr	Fault of Dr	Fault of Dr	Fault of Pe	Fault of Pe	Fault of Pe	Defect in C	Defect in C	Defect in C	Defect in F	Defect in F	Defect in F	Weather C	Weather C	Weather C	Fa
1	Andhra Pradesh	21359	6743	26287	329	102	371	288	75	425	234	104	160	67	26	85	285	192	442	60	19	70	
2	Arunachal Pradesh	30	19	30	12	5	36	20	14	45	0	0	0	17	8	22	15	9	29	23	13	25	
3	Assam	6895	2429	6281	59	28	53	2	1	1	144	53	129	0	0	0	0	0	0	0	0	0	
4	Bihar	5008	2646	3374	352	151	260	608	254	395	286	144	233	541	302	355	446	252	282	409	210	306	
5	Chhattisgarh	9108	2458	8710	95	41	96	726	253	665	146	41	139	672	229	489	440	142	381	212	60	206	
6	Goa	3795	247	1650	60	4	23	0	0	0	175	29	86	17	0	13	1	0	0	1	0	0	
7	Gujarat	18945	6635	18256	423	82	519	450	170	516	1512	405	1311	161	60	164	203	32	107	194	43	241	
8	Haryana	8208	3186	7633	153	71	43	456	210	202	348	140	193	165	107	96	141	74	56	127	75	68	
9	Himachal Pradesh	2936	1148	5326	1	0	1	22	3	30	0	0	0	74	37	156	25	11	63	0	0	0	
10	Jammu & Kashmir	3614	561	4088	0	0	0	380	62	937	223	35	133	341	38	193	113	11	128	20	2	36	
12	Jharkhand	2961	1354	2367	46	38	39	358	209	225	305	193	242	285	156	220	223	137	231	203	94	230	
13	Karnataka	35409	8508	45635	522	101	718	3054	648	4489	335	77	867	626	113	810	178	43	183	679	202	307	
14	Kerala	36282	4049	41096	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
15	Madhya Pradesh	41275	6740	44090	286	5	162	2215	346	2240	152	10	99	1664	370	1690	1446	225	1337	539	83	448	
16	Maharashtra	48237	11113	34294	95	44	86	7270	531	2186	334	111	268	497	98	455	381	79	324	102	41	92	

1	ACCIDENT	ACCIDENT	ACCIDENT	ALCOHOL	ACCIDENT	DAY_OF_V	HIT	RUN	LIGHT	CO	POLICE	A	SEVERITY	SPEED	ZO	LONGITUDE	LATITUDE	LGA	NAM	TOTAL	PE	INJ	OR	F	FATALITY	SERIOUS	INJURY	NONINJ	U	MALES	FEMALES	BICYCLIST	
2	T2010001	#####	14.20.00	1	1	6	0	1	1	1	1	60	146.4239	-38.2431	1	7	5	0	0	5	2	0	7	0	0	0	0	0	0	0	0	0	
3	T2010001	#####	19.00.00	1	1	4	0	2	0	2	50	144.9353	-37.7454	3	2	1	0	0	1	0	1	2	0	0	0	0	0	0	0	0	0	0	
4	T2010001	#####	15.00.00	1	2	6	0	1	0	1	0	60	144.962	-37.7681	3	3	1	0	0	1	0	1	2	1	2	1	2	1	2	1	2	1	
5	T2010001	#####	16.10.00	1	1	6	0	1	1	1	1	60	145.0764	-37.863	14	3	1	0	0	1	0	1	2	1	2	1	2	0	0	0	0	0	
6	T2010001	#####	11.00.00	0	3	5	0	1	0	1	0	60	146.0804	-37.9173	2	1	1	0	0	0	1	0	0	1	0	1	0	0	0	0	0	0	0
7	T2010001	#####	17.05.00	1	1	6	0	1	1	1	1	50	144.973	-37.7646	3	3	1	0	0	0	1	2	1	0	1	0	1	0	1	0	1	0	1
8	T2010001	#####	10.20.00	0	4	6	0	1	0	2	-1	146.0013	-37.3617	4	1	1	0	1	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0
9	T2010001	#####	18.00.00	1	1	6	0	2	1	2	70	145.0291	-37.7408	9	3	2	0	2	0	2	0	1	1	2	1	2	0	0	0	0	0	0	0
10	T2010001	#####	19.15.00	1	1	6	0	2	1	1	60	144.9402	-37.5988	15	2	2	0	0	0	2	0	0	2	0	0	0	0	0	0	0	0	0	0

1	ACCIDENT	TOTAL_PERSONS
2	0 #####	7
3	1 #####	2
4	2 #####	3
5	3 #####	3
6	4 #####	1
7	5 #####	3
8	6 #####	1
9	7 #####	3

3.2 DESCRIPTION

Categorical attributes present – State/Union Territory

Quantitative attributes present- All Other attributes present based on each database

For example

Quantitative attributes present in
“Details_of_road_accident_deaths_by_situation_state_2014.csv”

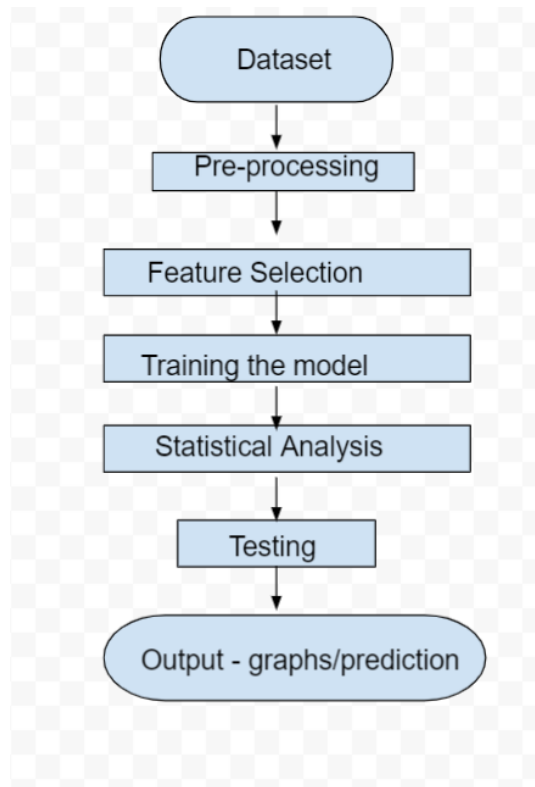
Year, CrimeHeadOffenders (Driver/Pedestrian), Died_MaleOffenders (Driver/Pedestrian), Died_FemaleOffenders (Driver/Pedestrian), Died_TransgenderOffenders (Driver/Pedestrian) , Died_TotalVictims, Died_MaleVictims, Died_FemaleVictims, Died_Transgender, Victims Died_Total

4. TASKS

- Download datasets
- Data preparation
- Modelling

- Visualization

SYSTEM DESIGN



TOOLS/ IMPORTANT LIBRARIES USED

Numpy -NumPy is a Python library used for working with arrays. It also has functions for working in the domain of linear algebra, fourier transform, and matrices.

Pandas - Pandas is a software library written for the Python programming language for data manipulation and analysis. In particular, it offers data structures and operations for manipulating numerical tables and time series.

Matplotlib- Matplotlib is a plotting library for the Python programming language and its numerical mathematics extension NumPy. It provides an object-oriented API for embedding plots into applications.

Seaborn- Seaborn is a Python data visualization library based on matplotlib. It provides a high-level interface for drawing attractive and informative statistical graphics.

Plotly- The plotly Python library is an interactive, open-source plotting library that supports over 40 unique chart types covering a wide range of statistical, financial, geographic, scientific, and 3-dimensional use-cases.

OrderedDict- OrderedDict is a dict subclass that preserves the order in which key-value pairs, commonly known as items, are inserted into the dictionary.

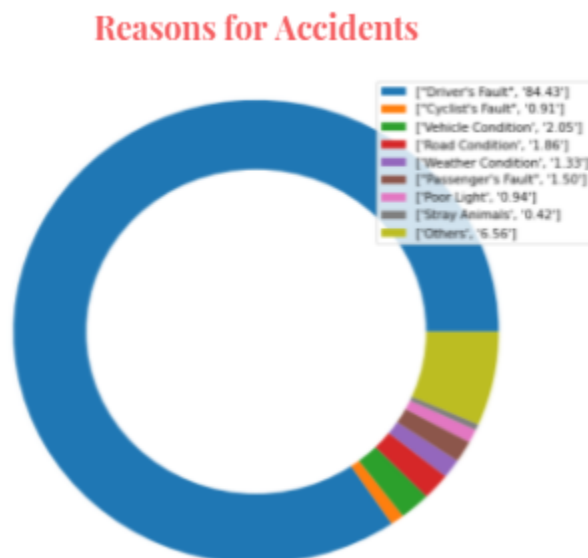
Datetime- It is a Python library used for Time access and conversions.

Bar_chart_race- Make animated bar chart races in Python with matplotlib. [Img.](#)

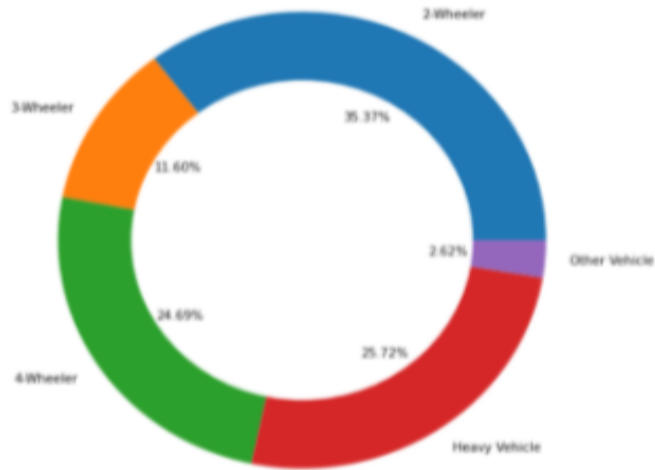
5. ACTIONS

- Data pre-processing
- Feature selection
- Training the model
- Statistical analysis
- Testing
- Output/ graphs prediction

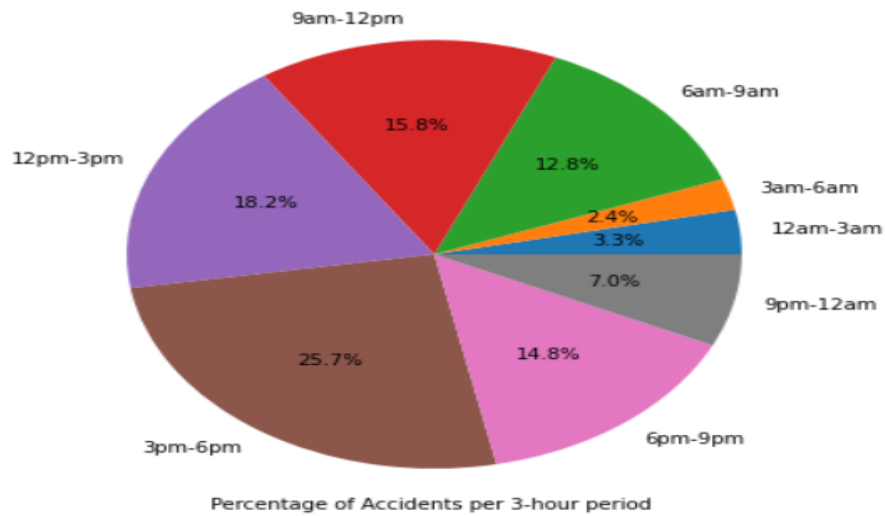
6. IMPLEMENTATION



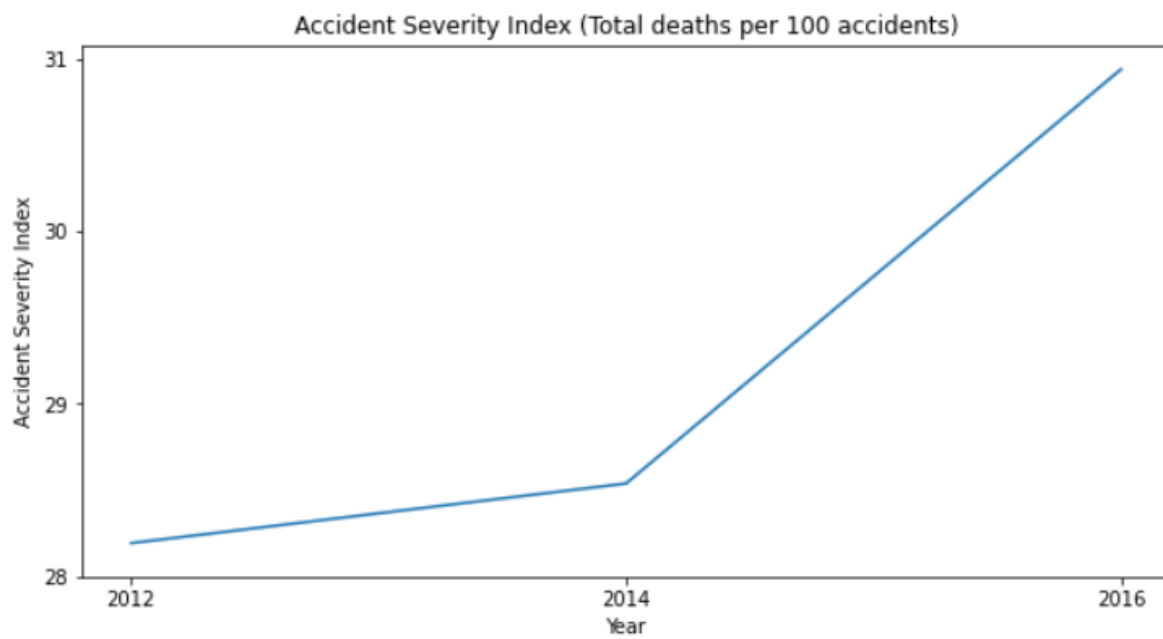
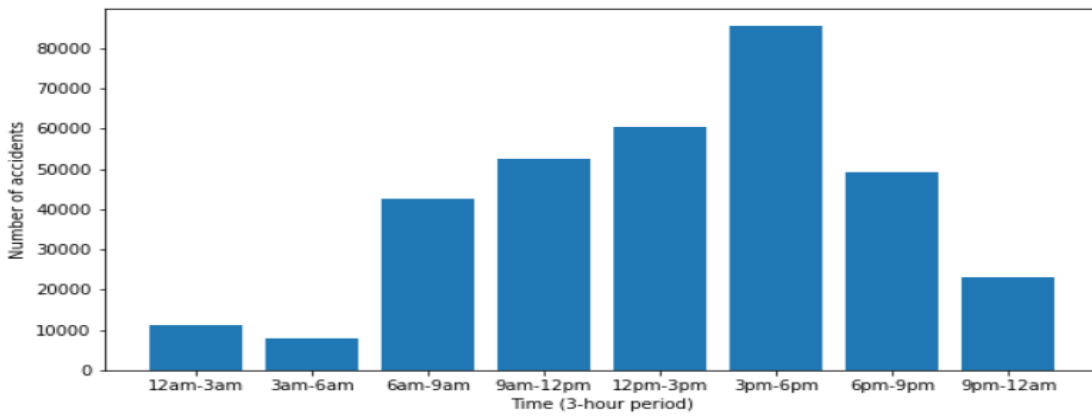
Types of vehicle involved in Accidents



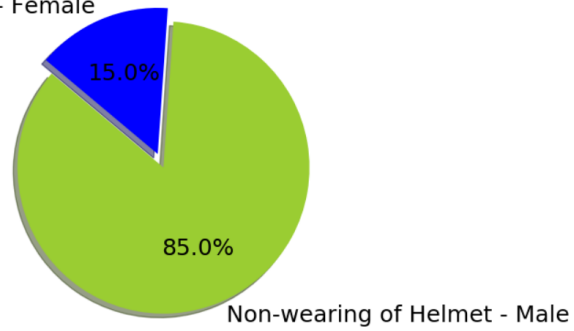
Percentage of accidents per 3-hour period



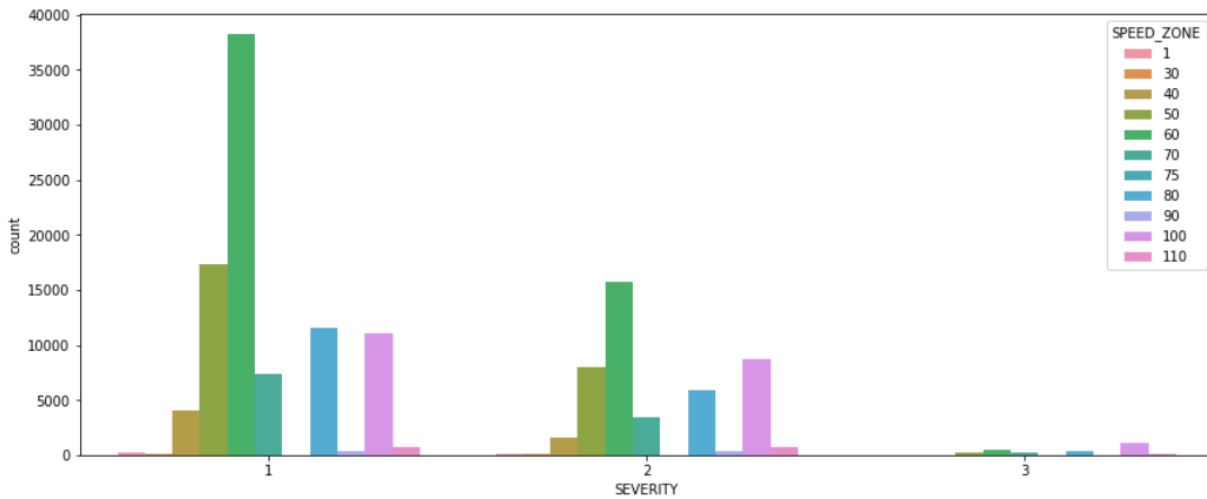
Number of Accidents happening in particular time interval



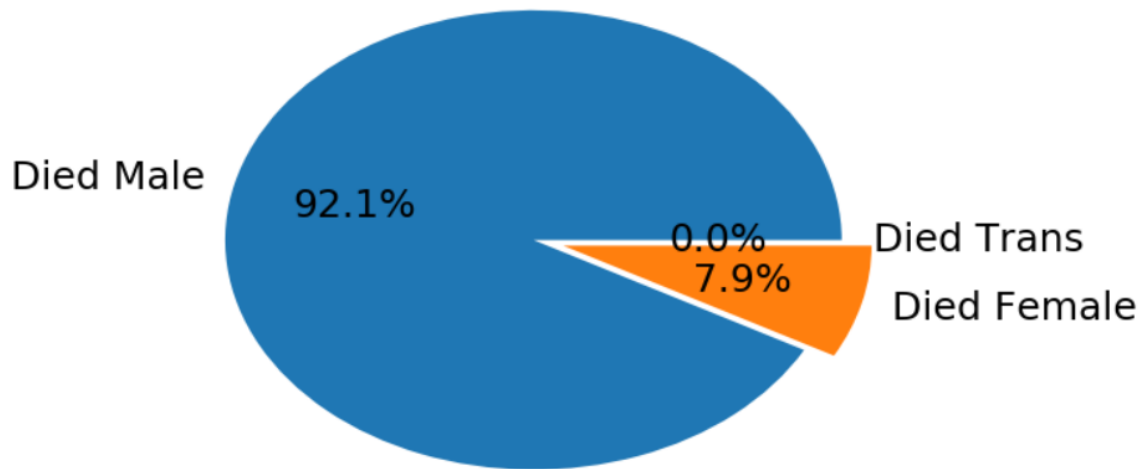
Percentage of Deaths occurring due to non-wearing of helmets between male and female.
Non-wearing of Helmet - Female



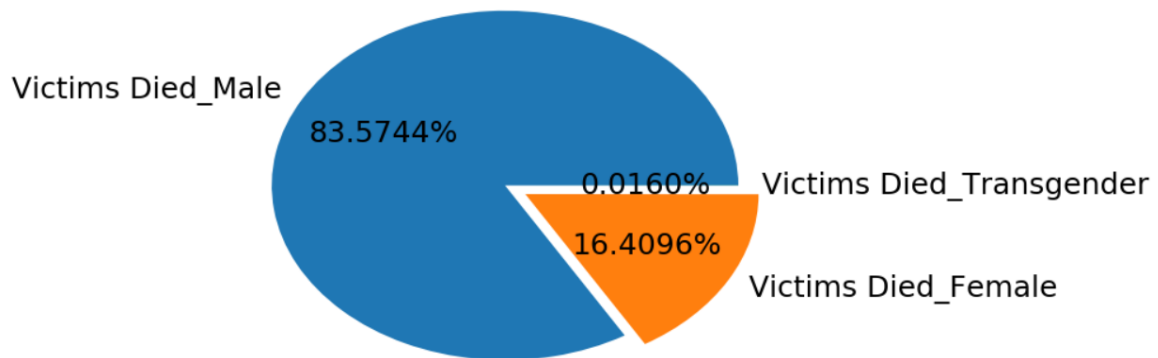
Number of Accidents at various severity level with respect to speed zone



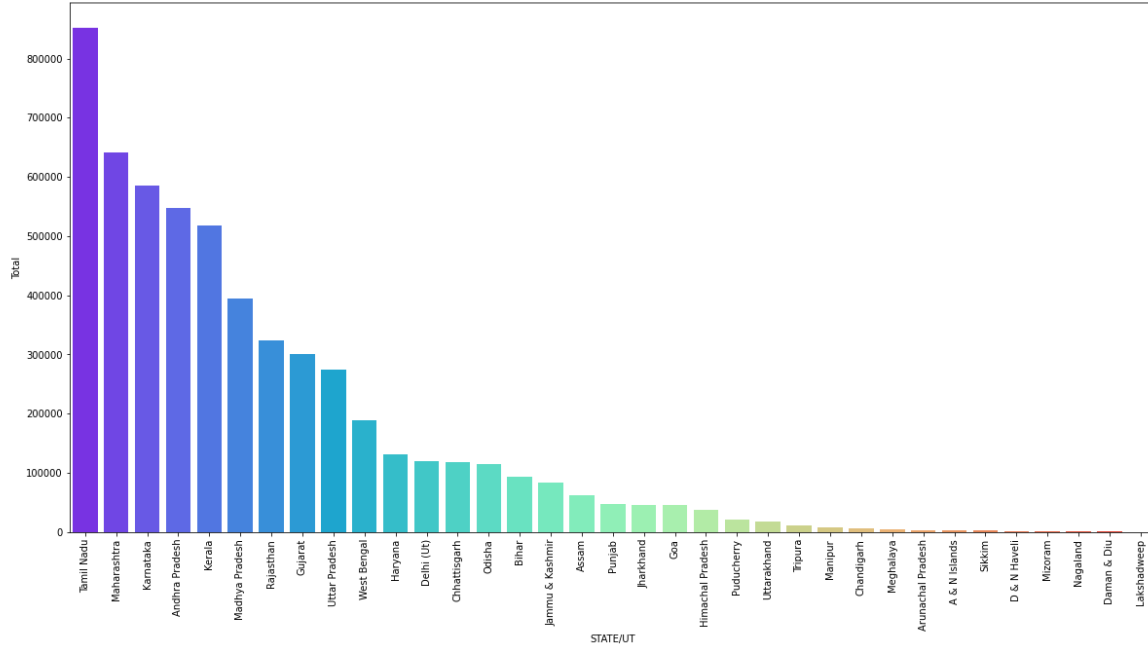
Number of offenders died in accidents with respect to gender



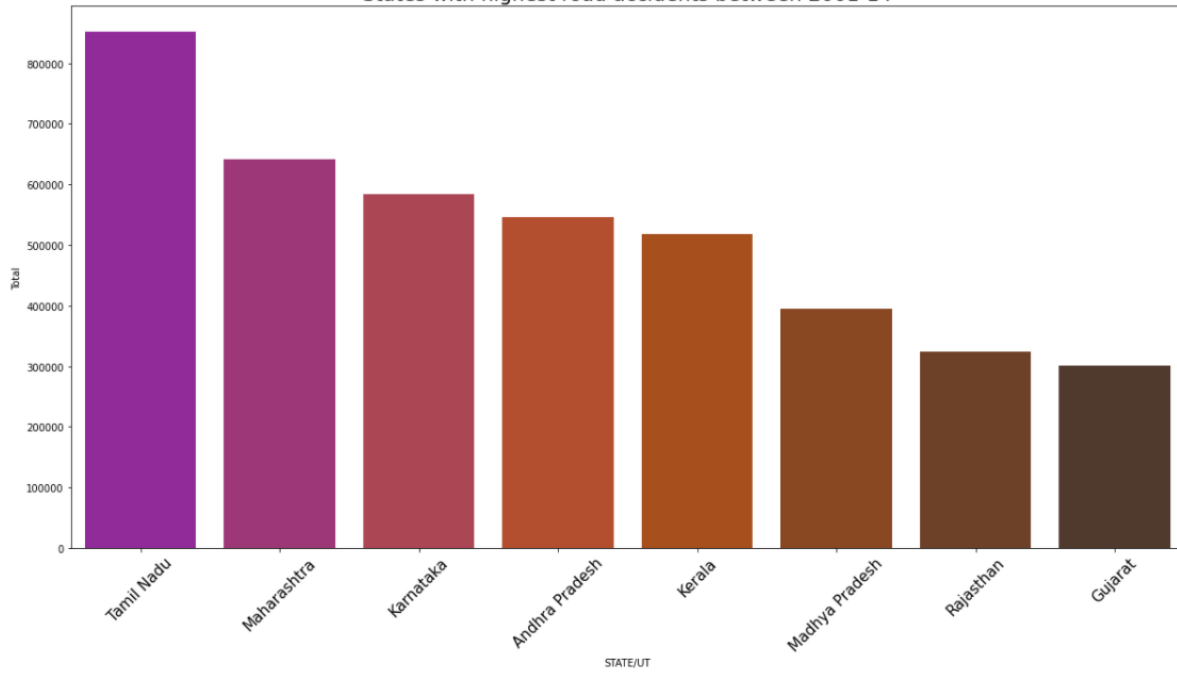
Number of victims died with respect to gender

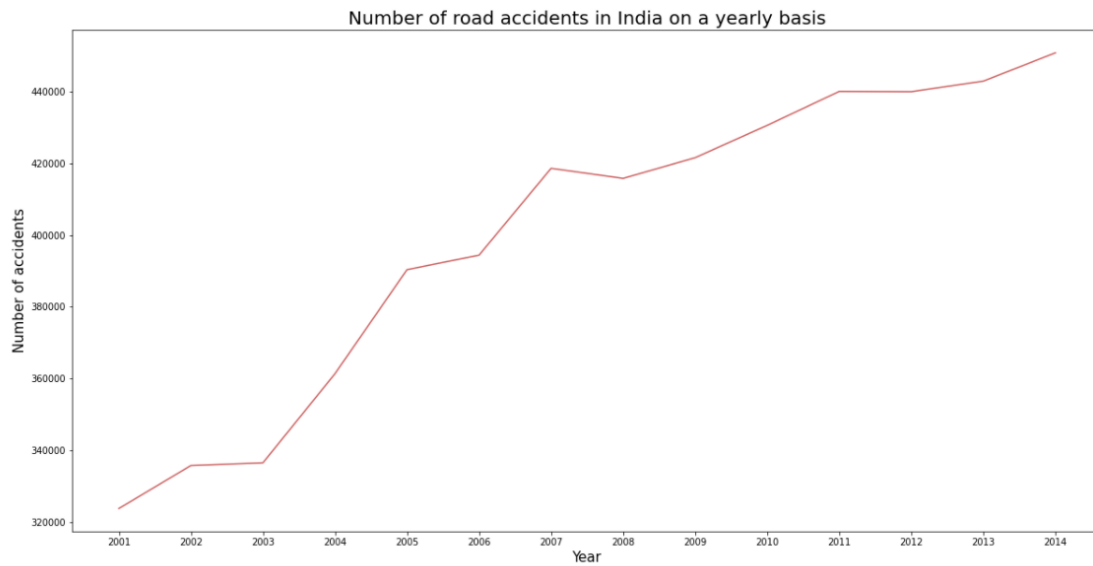


Cases of road accidents in each state/UT from 2001-14

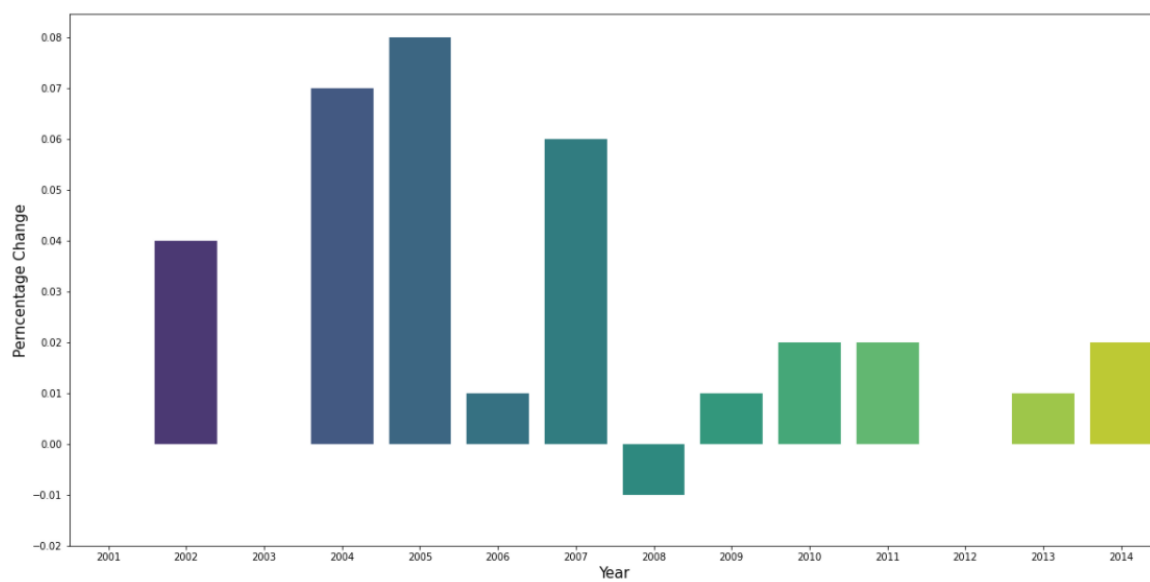


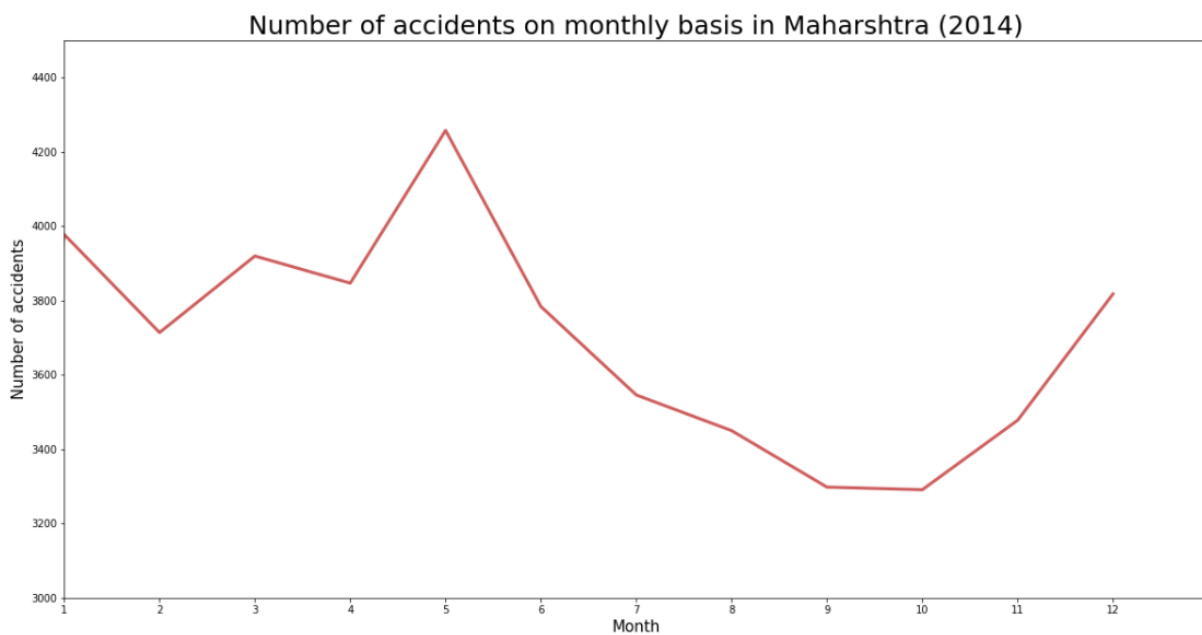
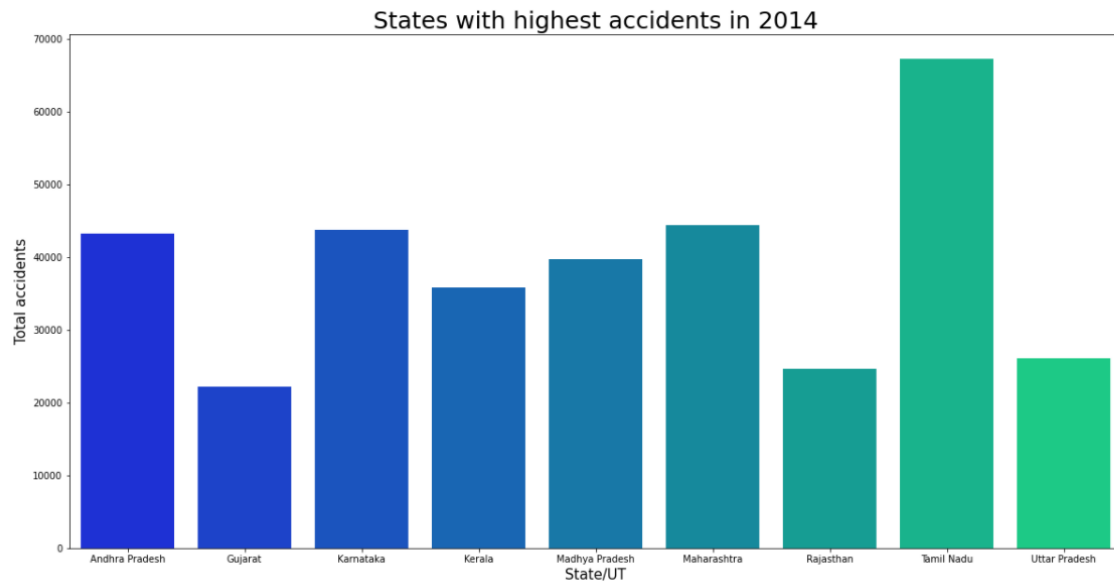
States with highest road accidents between 2001-14

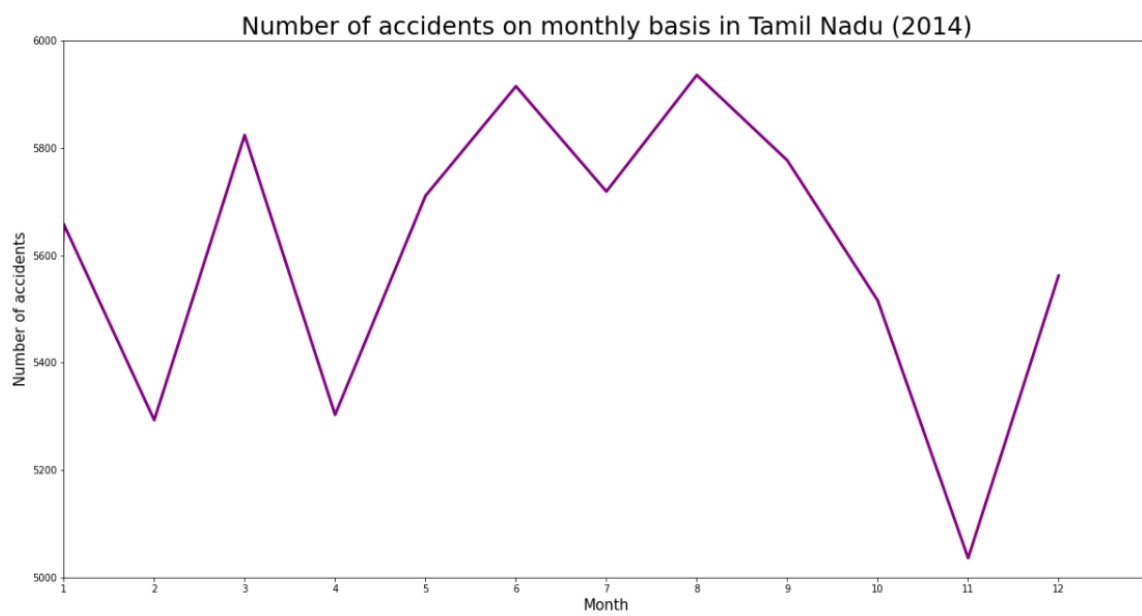
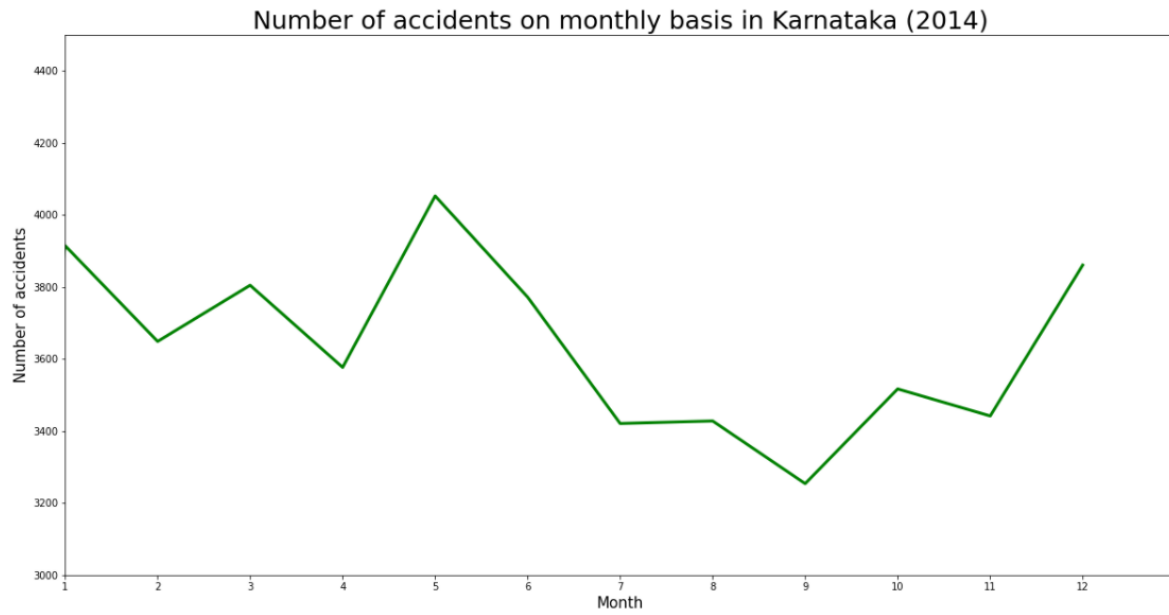


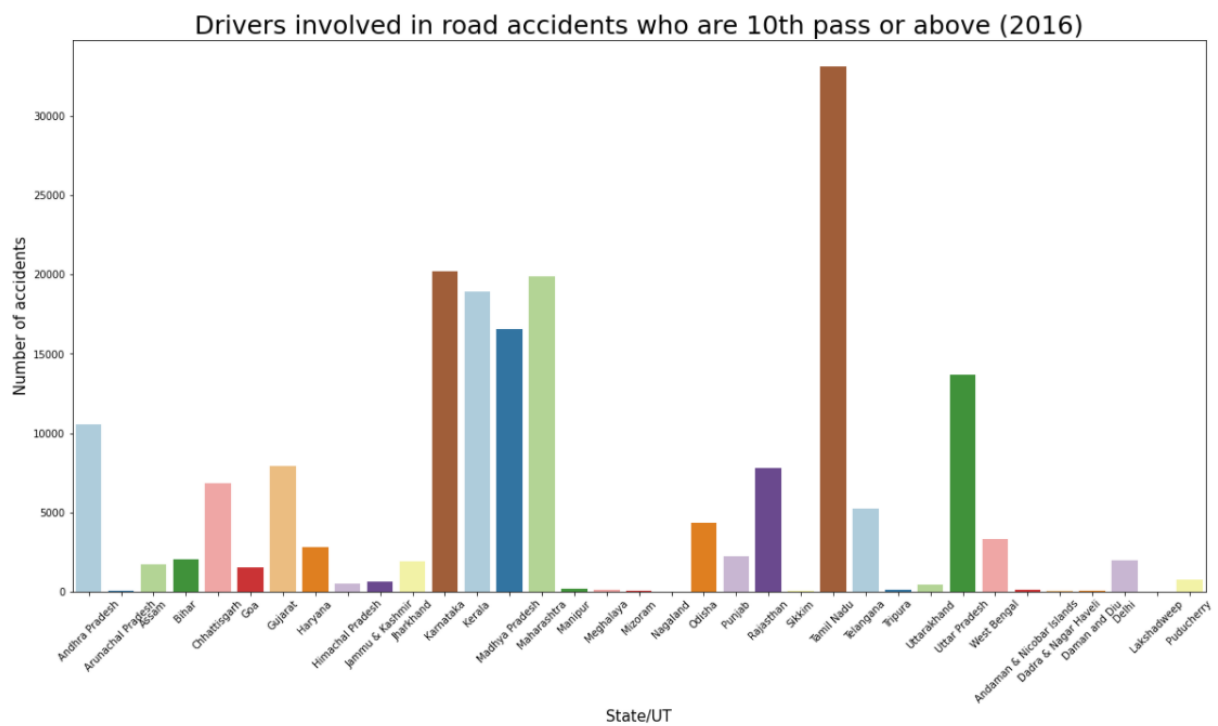
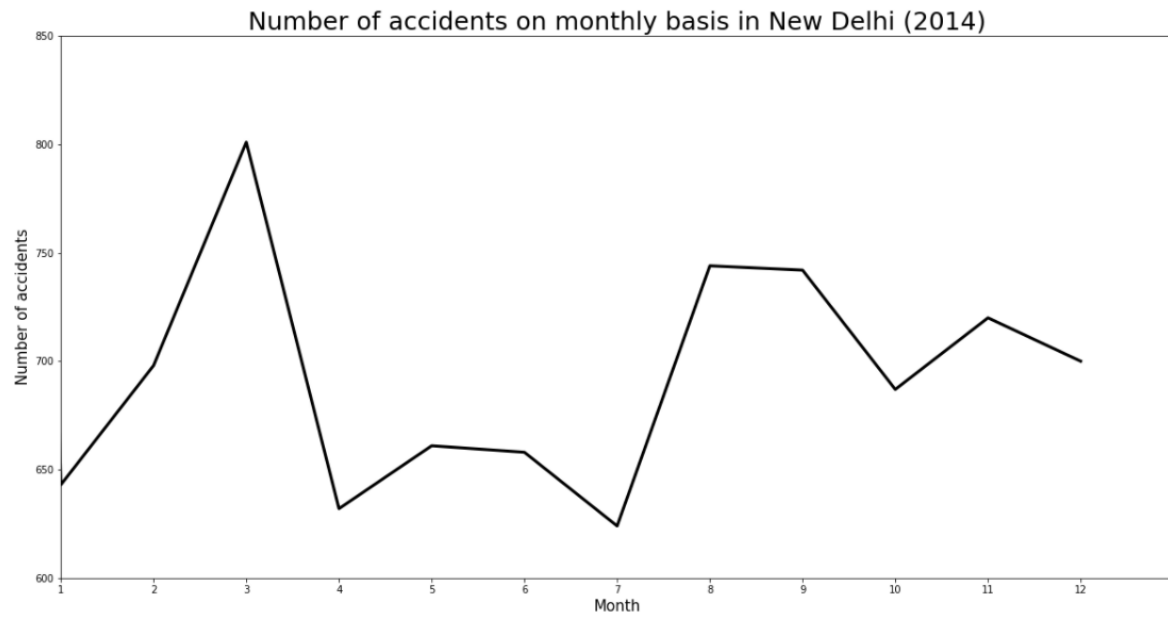


Percentage change in the accidents with previous year

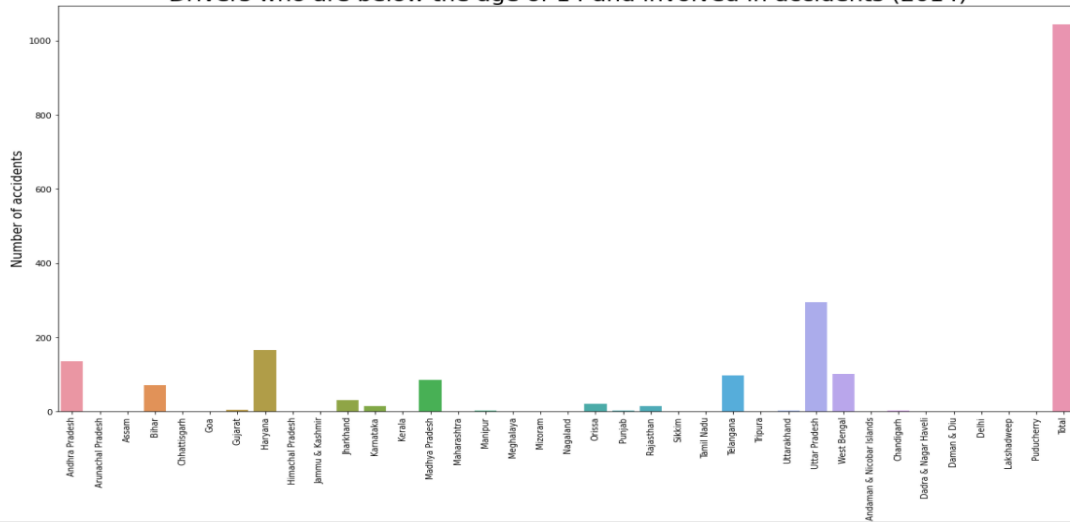




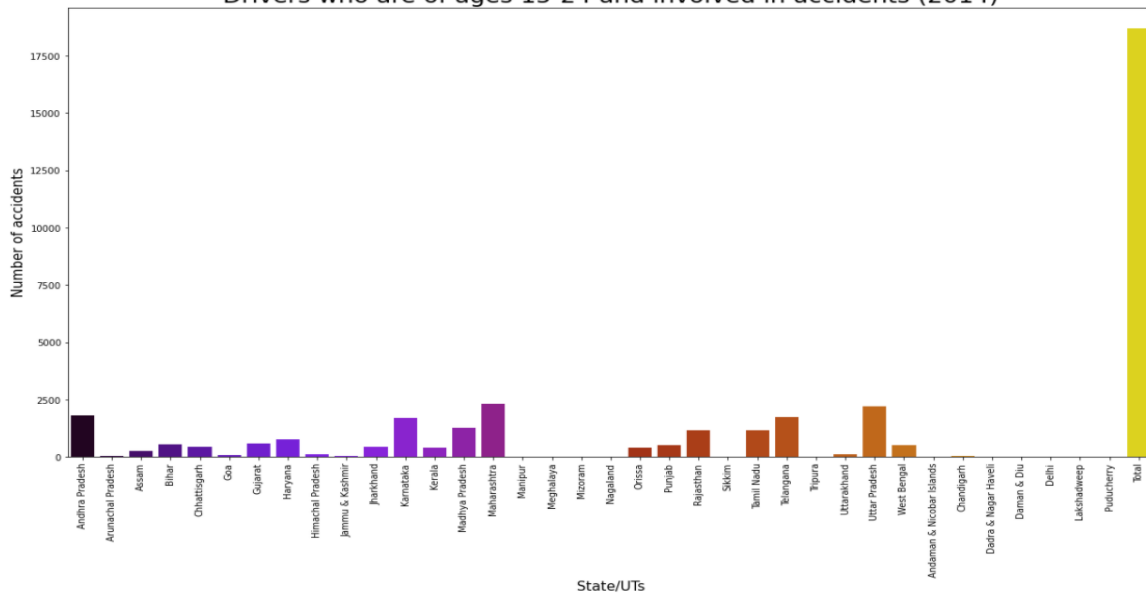


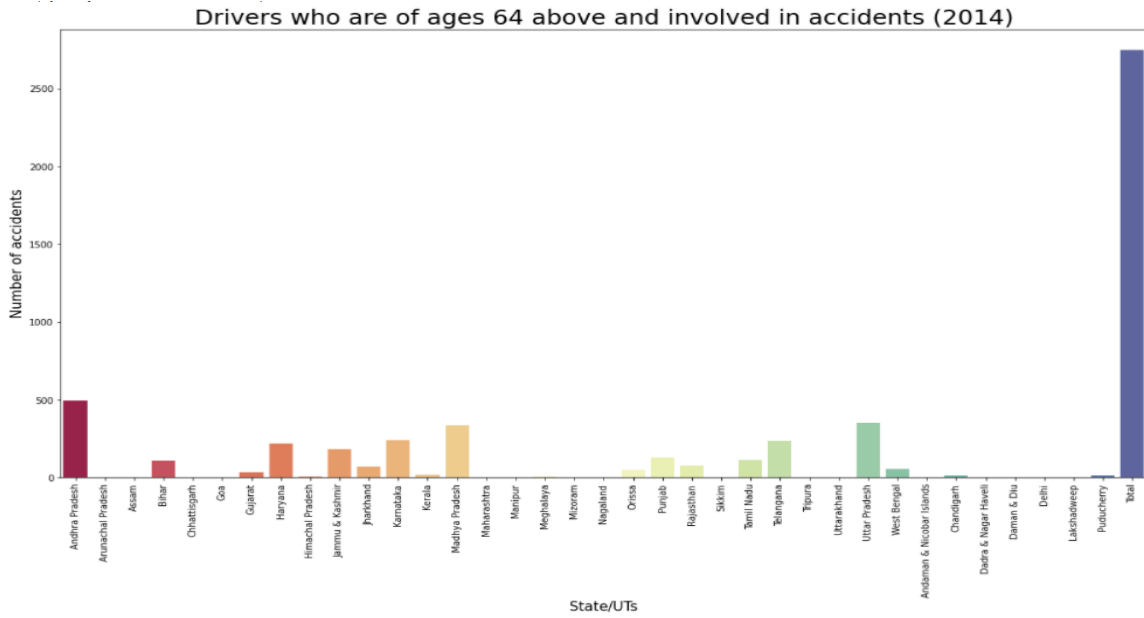
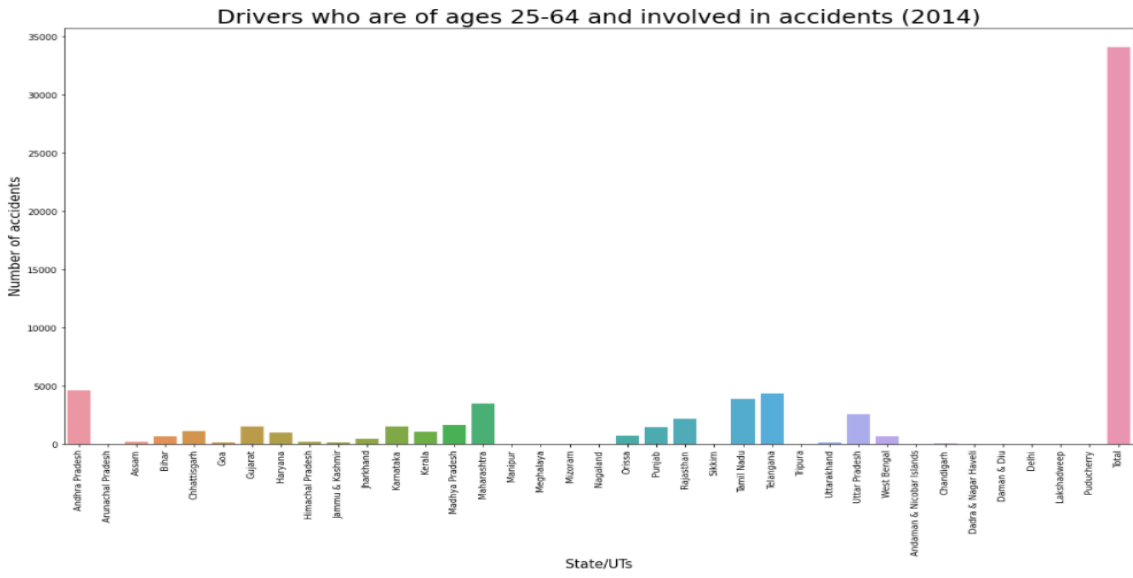


Drivers who are below the age of 14 and involved in accidents (2014)

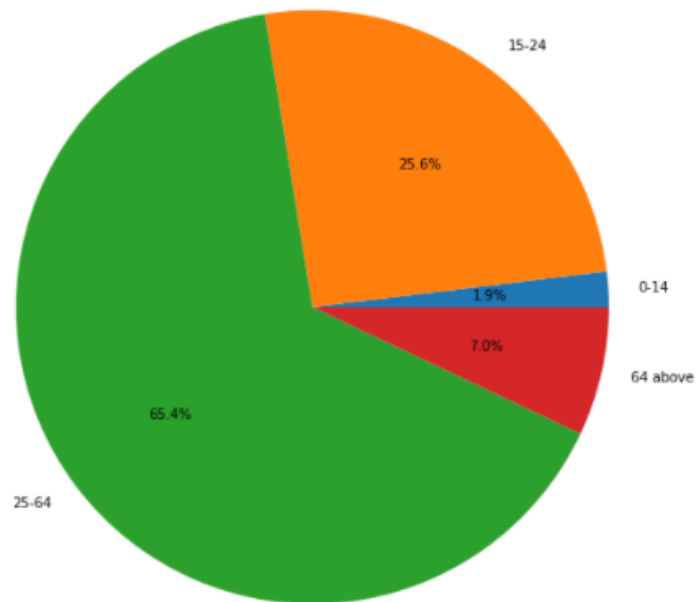


Drivers who are of ages 15-24 and involved in accidents (2014)

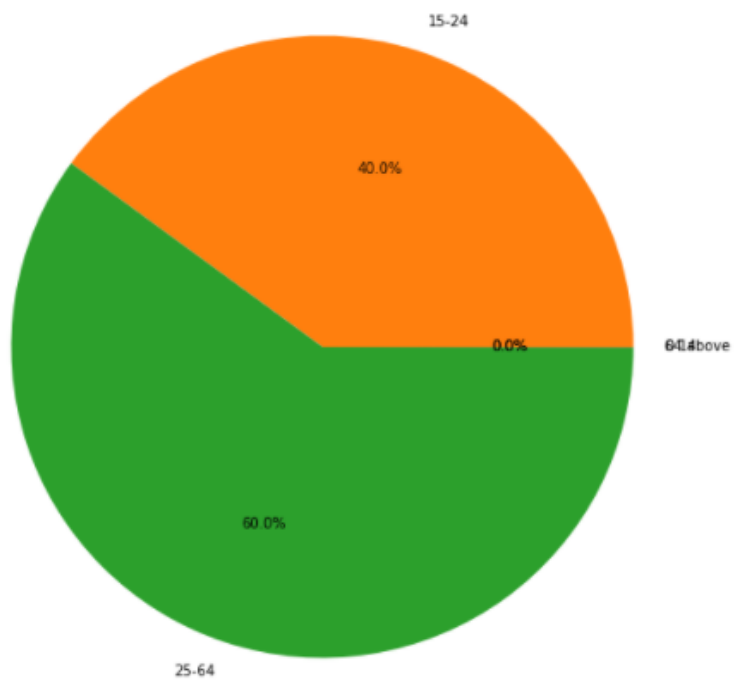




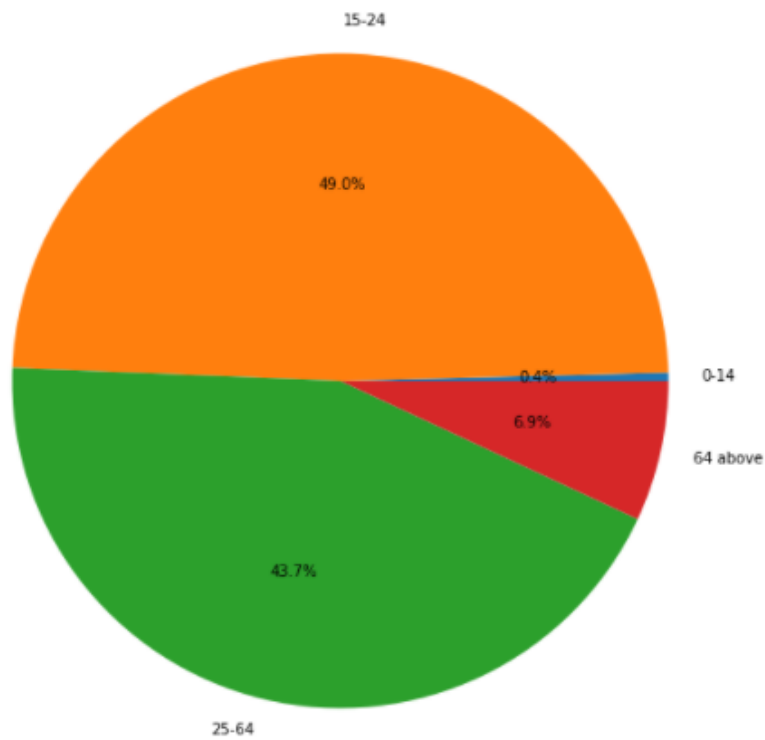
Road accidents in Andhra Pradesh by age groups (2014)



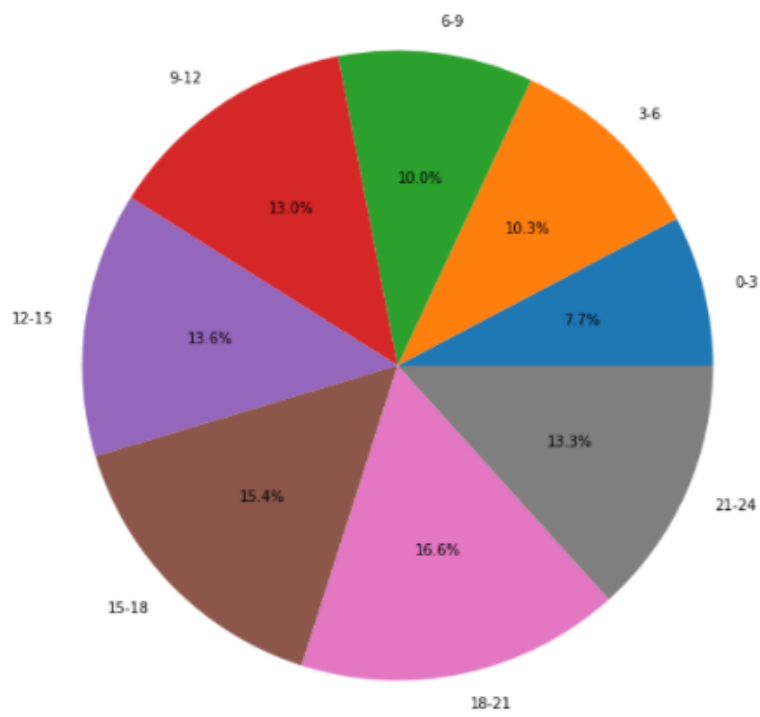
Road accidents in Maharashtra by age groups (2014)



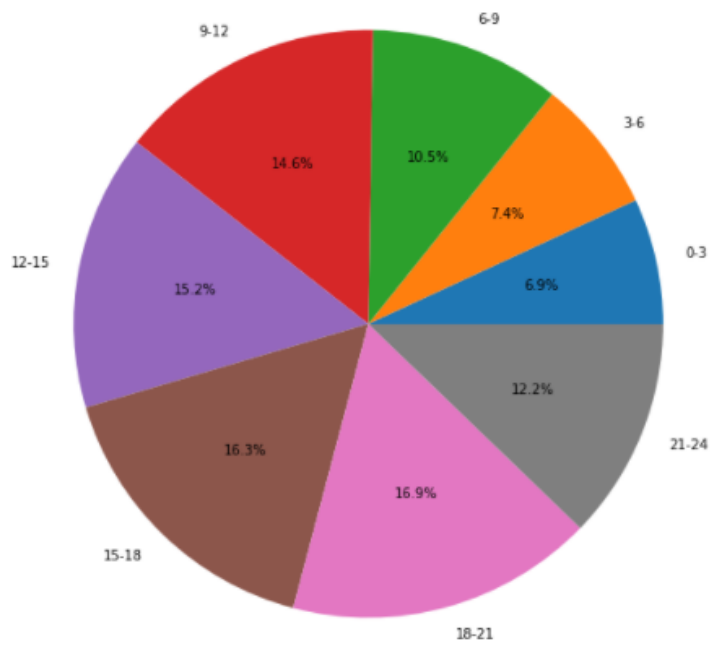
Road accidents in Karnataka by age groups (2014)



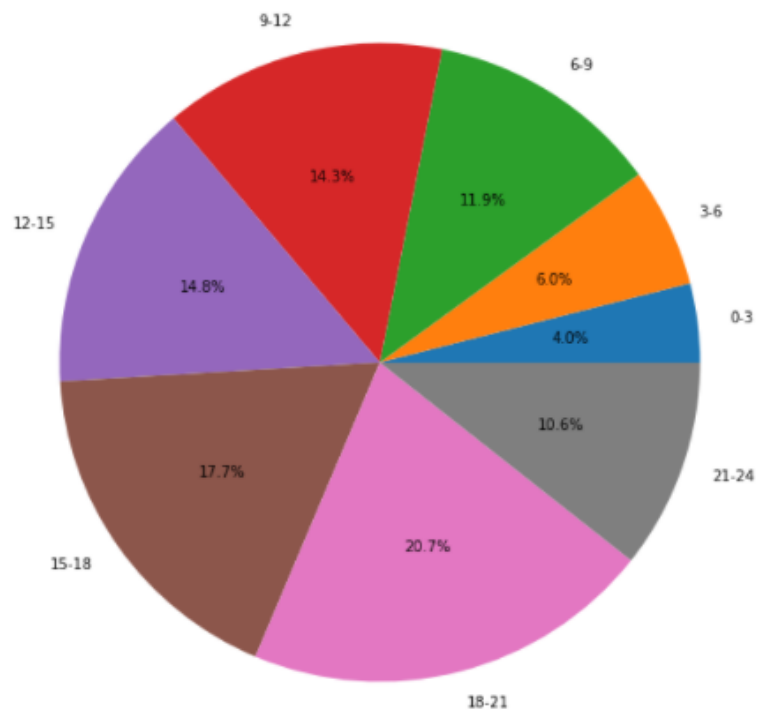
Road accidents in Andhra Pradesh by time of accident in hrs (2014)



Road accidents in Karnataka by time of accident in hrs (2014)



Road accidents in Tamil Nadu by time of accident in hrs (2014)



7. VALIDATION

7.1 DOMAIN LEVEL

- Our primary aim is to analyse the data obtained by applying statistics analysis and data visualisation algorithms on various Accident datasets.
- Certain safety driving suggestions were made based on statistics, association rules, classification model, and clusters obtained.
- These are made by analysing data and comparing trends to find out why and how such factors can affect road accidents.

7.2 TASK LEVEL

- The main motivation of our project revolves around the fact that in India approximately 465033 people lose their lives every year because of road accidents.
- India's young, productive population, aged 18-45 years, is involved in 70% of road accidents.
- So, through this project we aim to analyse trends and information /data related to road accidents in India and give a detailed analysis of these accidents which may be very helpful in controlling these accidents.
- This project analysis makes people cautious of the way accidents happen so that they can take precautions while driving.

7.3 DATA LEVEL

- Reasons for accidents.
- Percentage of accidents in particular periods.
- Types of vehicles involved.
- Severity index.
- Reason for accidents.
- Road accidents of all states and union territories from 2001-14.
- Road accidents on a yearly basis.
- Monthly analysis on states having the highest number of accidents.
- Analysis based on drivers age.

7.4 ALGORITHM LEVEL

- Language used : python
- IDE used - spyder and google collab

8. SUGGESTIONS/ CONCLUSIONS

- Whoever is driving should make sure that he is in a proper state to drive.
- Most of the accidents were happening in the time interval of 3pm-6pm, so people should be careful for various reasons like change in light condition etc..
- Apart from the year 2008, in all other years there is either no change or increase in the percentage of accidents . Despite so many warnings and laws issued regarding Road accidents like no triple riding, don't drink and drive, carry the prescribed weight on vehicles, everybody finds a way to violate these rules. But people should understand that at the end it's going to risk their lives and the lives of people travelling with them.
- The highest number of accidents are recorded from age groups 15-24 and 25-64. There can be various reasons for this like : Alcoholism, Bike racing, Reckless driving, or the urge to work or college on time etc...
- We have also seen that in the months of the rainy season the accidents actually peak down because people are extra careful.
- How many guidelines or awareness drives the government or any organization start , they are of no use until the people consider them seriously. Value their lives and other's lives.
- Even if people are following every precaution there is always a loophole in problems like this. So all we can do is be careful, be sober.