VIT UNIVERSITY
(Estd. u/s 3 of UGC Act 1956)
www.vit.ac.in
Vellore ▪ Chennai
CHENNAI CAMPUS
Vandalur - Kelambakkam Road, Chennai - 600127

ROAD ACCIDENTS ANALYSIS

(J Component)

A Project Report submitted as part of the course.

Exploratory Data Analysis (CSE3040)

to

Ms. SUBHRA RANI PATRA

School of Computer Science and Engineering (SCOPE)

BY

ARUN VENKAT S J	19MIA1076
JOHN CHACKO	19MIA1097
LOKESH KANNA P	19MIA1014
SANJAY A ANTONY	19MIA1092

WINTER 2020 - 21

VIT UNIVERSITY, CHENNAI
VANDALUR KELAMBAKAM ROAD
CHENNAI-600127

ABSTRACT

Road accidents are a human tragedy. They involve high human suffering and monetary costs in terms of untimely deaths, injuries and loss of potential income. Although we have undertaken many initiatives and are implementing various road safety improvement program the overall situation as revealed by data is far from satisfactory. The main aim of this paper is to analyse the road accidents in India at national, state, and metropolitan city level. Moreover, road accidents are relatively higher in extreme weather and during working hours. Analysis of road accident scenario at state and city level shows that there is a huge variation in fatality risk across states and cities. Fatality risk in 16 out of 35 states and union territories is higher than all India average. Although, burden of road accidents in India is marginally lower in its metropolitan cities, almost 50% of the cities face higher fatality risk than their mofussil counterparts. In general, while in many developed and developing countries including China, road safety situation is generally improving, India faces a worsening situation. Without increased efforts and new initiatives, the total number of road traffic deaths in India is likely to cross the mark of 250,000 by the year 2025. There is thus an urgent need to recognize the worsening situation in road deaths and injuries and to take appropriate action. We studied accidental records of various states, identified the black-spots of accidents and then analysed the geometric features of those spots whose observation is given in this paper. The 2004 World Health Report shows that of the 1.2 million people killed in road crash worldwide, 85% are in developing countries. This study tries to analyse the traffic accidents, and develop a preventive strategies and possible counter measures for the route selected. This thesis has main functions. The aim is to provide users with an understanding of the major causes of traffic accidents and present using several Statistical tools. Road traffic accidents are amenable to remedial action. Many a country have curbed the menace of road accidents by adopting a multipronged approach to road safety that encompasses broad range of measures, such as, traffic management, design and quality of road infrastructure, application of intelligent transport system, safer vehicles, law enforcement, effective and quick accident response and care etc. The Government alone cannot tackle road safety problems. There is a need for active involvement of all stake-holders to promote policy reform and implementation of road safety measures. Addressing road safety is comprehensive manner underscores the need to involve multiple agencies and sectors like health, transport and police. The present study provides the magnitude and various dimensions of road accident in India. The analysis on road accidents in this study will help to create awareness, guidelines and assist in informed decision making on road safety.

INTRODUCTION

Traffic accidents are global disasters that occurs frequently in almost every country. Most of the Traffic accidents causes some injuries to the people while more fatal accidents cause deaths. According to statistics every minute one serious road accident occurs in the country and on an average every hour 16 people die on Indian roads. Fatalities and injuries resulting from road traffic accidents are a major and growing public health problem in India.

According to the statistics in the report on Road Accidents in India, "the country recorded at least 4,80,652 accidents in 2016, leading to 1,50,785 deaths. The number also suggests that at least 413 people died every day in 1,317 road accidents. It also said that India witnessed 17 deaths and 55 road accidents every hour in 2016". Every week nearly 2,650 people get killed and 9,000 get injured due to traffic accidents. We can see from the statistics that the number of road accidents are very high and this is becoming an increasing huge problem for us. Road accidents can be caused because of various factors, for example, hardware breakdown as well as the activities of the driver such as talking on mobile phones, speeding, drunk driving, aggressive behaviours like tailgating or unsafe lane changes and so on. The factors which cause accidents can be limitless. The consequences of traffic accidents depend upon a number of factors such as the impact, number of vehicles involved and if vehicle occupants were protected by safety belts and/or air bags. Thus, many traffic accidents can be avoided or their impact can be reduced by taking effective measures.

A large amount of data can be extracted from these accidents data. The data collected from these accidents are large as it is for any country but this is even for a country like India. India is a country with such a large population that the number of traffic accidents also increases and so does the amount of data from these accidents.

Traffic accidents have now earned India a dubious distinction; with nearly 140,000 deaths annually, the country has overtaken China to top the world in road fatalities. India is the only country in the world which faces more than 15 fatalities and 53 injuries every hour as a consequence of road crashes. While in many developed and developing countries , the situation is generally improving, India faces a worsening situation. If the trend continues, the total number of road traffic deaths in India would increase by 100% between 2017 and 2027. The analysis shows that during the last ten years, road accidental fatalities in India have increased at the rate of 5% per year while the population of the country has increased only at the rate of 1.4% per year.

Fatality risk in India is not only quadruple than that in some of the developed countries such as United Kingdom and Sweden but also still increasing rapidly. It is also found that the distribution of road accidental deaths and injuries varies according to age, gender, month and time. Among people of all age groups, people of economically active age group of 30-59 years is the most vulnerable. However, if we compare gender-wise fatalities and accidents, we found that the males accounted for 85.2% of all fatalities and 82.1% of all injuries in 2016.

Analyses of this data will help us to extract valuable information about the impact of these accidents in different areas. By applying various queries on the data we can find out the short comings of the various protocols in different regions and can find a solution to fix it so that the intensity and the impact of the traffic accidents can be lowered.

Moreover, road accidents are relatively higher in May-June and December-January which shows that extreme weather influences the occurrence of road accidents. Accidents remain relatively constant and high during 9 AM - 9 PM and variable but low during mid-night and early hours of the day. However, this does not imply that daytime driving is more risky than nighttime driving. The study also tries to find out cause-wise distribution of road accidents. There are several factors responsible for accidents but drivers' fault is found to be the most important one; drivers' fault accounted for 78% of total accidents.

The Commission for Global Road Safety believes that the urgent priority is to halt this appalling and avoidable rise in road injury and then begin to achieve year on year reductions. The world could prevent 5 million deaths and 50 million serious injuries by 2020 by dramatically scaling up investment in road safety, at global, regional and national levels.

Each year nearly 1.3 million people die as a result of a road traffic collision, more than 3000 deaths each day and more than half of these people are not travelling in a car. Twenty to fifty million more people sustain non-fatal injuries from a collision, and these injuries are an important cause of disability worldwide. Ninety percent of road traffic deaths occur in low and middle-income countries, which claim less than half the world's registered vehicle fleet. Road traffic injuries are among the three leading causes of death for people between 5 and 44 years of age. Unless immediate and effective action is taken, road traffic injuries are predicted to become the fifth leading cause of death in the world, resulting in an estimated 2.4 million deaths each year. This is, in part, a result of rapid increases in motorization without sufficient improvement in road safety strategies and land use planning.

LITERATURE SURVEY

In the industrialized world, accidents are one if not the leading cause of work-related accidental deaths. Many different types of people drive as part of their work, and having to be on the road puts these people at risk. Despite this, relatively few studies have looked at the characteristics and risk factors of Road accidents.

A dozen specialized electronic databases of documents published between 1995 and 2008 were searched for a literature review. From this search, together with an earlier Web search (Google), a total of 162 documents were identified, and these were analysed in detail. The analysis looked mainly at-risk factors identified and described in the documents. These risk factors were classified at five levels: 1) the driver of the vehicle and the passengers; 2) the immediate physical environment (the vehicle); 3) the external physical environment (the road); 4) the organizational work environment (the company); the political environment (laws and regulations). A last section was reserved for accidents that took place at road construction sites. Level 1 included the factors most often considered in the studies found. The characteristics of the driver (age, sex, education, etc.) and what he does (behaviour, alcohol/drug use, etc.) are elements to be taken into consideration in developing a prevention strategy, but factors at the other four levels must be considered as well.

Some of the Literature that helped to have important conclusion on Road accidents: "The Economic and Societal Impact of Motor Vehicle Crashes, 2010," by Lawrence Blincoe, Ted R. Miller, Ph.D., Eduard Zaloshnja, Ph.D., Bruce A. Lawrence, Ph.D., DOT HS 812 013, Washington, D.C. Charles, Geoffrey (11 March 1969). "Cars And Drivers Accident prevention instead of blame". The Times. Quoting from JJ Leeming in Accidents and their prevention: "Blame for accidents seems to me to be at best irrelevant and at worst actively harmful." McKernan, Megan (13 May 2015). "AAA Tests Shine High-Beam on Headlight Limitations". NewsRoom.AAA.com. AAA Automotive Research Centre. AAA's test results suggest that halogen headlights, found in over 80 percent of vehicles on the road today, may fail to safely illuminate unlit roadways at speeds as low as 40 mph. high-beam settings on halogen headlights may only provide enough light to safely stop at speeds of up to 48 mph, leaving drivers vulnerable at highway speeds. Assured Clear Distance Ahead Law & Legal Definition". US Legal, Inc. Retrieved 27 August 2013. ACDA or "assured clear distance ahead" requires a driver to keep his motor vehicle under control so that he can stop in the distance in which he can clearly see. *Cooper, Peter J. (Summer 1997) "The relationship between speeding behaviour (as measured by violation convictions) and crash involvement". Leibowitz, Herschel W.; Owens, D. Alfred; Tyrrell, Richard A.*

(1998). "The assured clear distance ahead rule: implications for night time traffic safety and the law".

In our Project we have collected the following datasets of Road accident statistics to analyse the trend and major reason for accidents

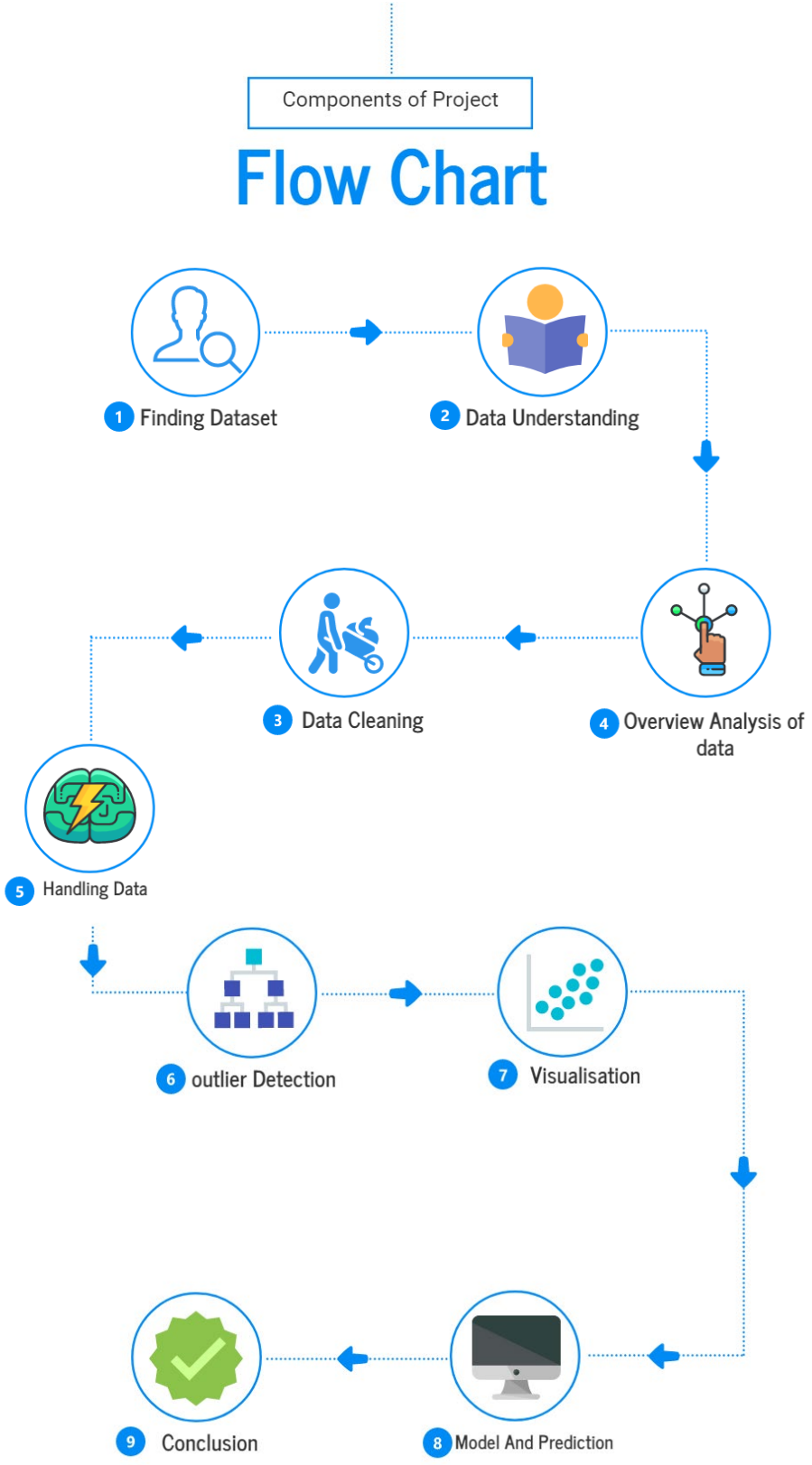
State/UT-wise Total number of Fatal Road Accidents, Total Road Accidents, Persons Killed and Injured on Other Roads from 2014 to 2017

- Weather conditions and no. of deaths and injuries due to road accidents (2014-17)
- Road conditions and no. of deaths and injuries due to road accidents (2014-17)
- State and UT wise no. of people killed in road accidents (2014-17)
- State and UT wise no. of people injured in road accidents (2014-17)

PROPOSED WORK

1. Dataset finding
2. Dataset understanding
3. Overview analysis of the data
4. Handling missing values
5. Outlier detection and removal
6. Visualizations
    a. Zone wise no. of accidents where people were killed
    b. Zone wise no. of accidents where people were injured
    c. Southern zone states: No. of accidents where people were killed
    d. Southern zone states: No. of accidents where people were injured
    e. Weather conditions – No. of People Killed in Road accidents (South zone)
    f. Road Conditions – No. of People Killed in Road Accidents (South Zone)
    g. Road Conditions – No. of People Injured in Road Accidents (South Zone)
    h. Northern Zone States: No. of Road Accidents where people were Killed
    i. Northern Zone States: No. of Road Accidents where people were Injured
    j. Weather conditions – No. of People Killed in Road accidents (North zone)
    k. Weather conditions – No. of People Injured in Road accidents (North zone)
    l. Road Conditions – No. of People Killed in Road Accidents (North Zone)
    m. Road Conditions – No. of People Injured in Road Accidents (North Zone)
7. Model and prediction
    a. Linear Regression
    b. K Means
    c. Random Forest
    d. Decision Tree
    e. Accuracy score Graph

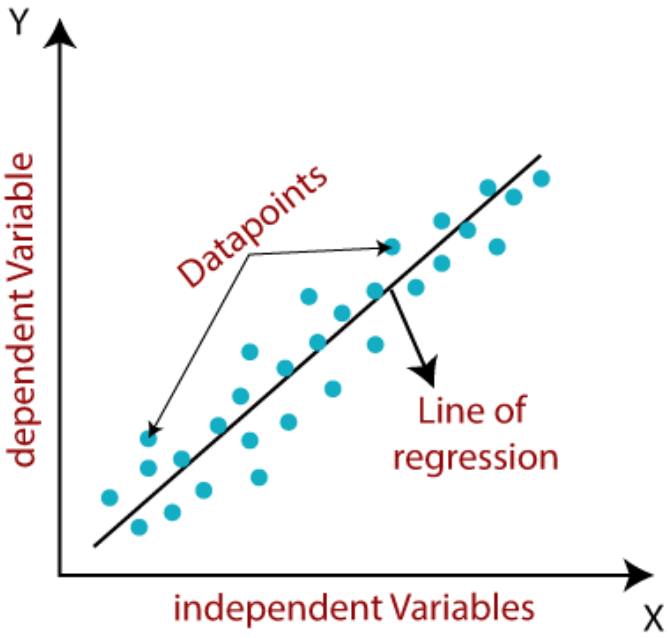
SYSTEM DESIGN OF PROPOSED WORK



WORKING MODULES

- Linear Regression
- K Means
- Random Forest
- Decision Tree
- Accuracy score graph

Linear Regression



Linear regression is an algorithm used to predict, or visualize, a relationship between two different features/variables. In linear regression tasks, there are two kinds of variables being examined: the dependent variable and the independent variable. The independent variable is the variable that stands by itself, not impacted by the other variable. As the independent variable is adjusted, the levels of the dependent variable will fluctuate. The dependent variable is the variable that is being studied, and it is what the regression model solves for/attempts to predict. In linear regression tasks, every observation/instance is comprised of both the dependent variable value and the independent variable value.

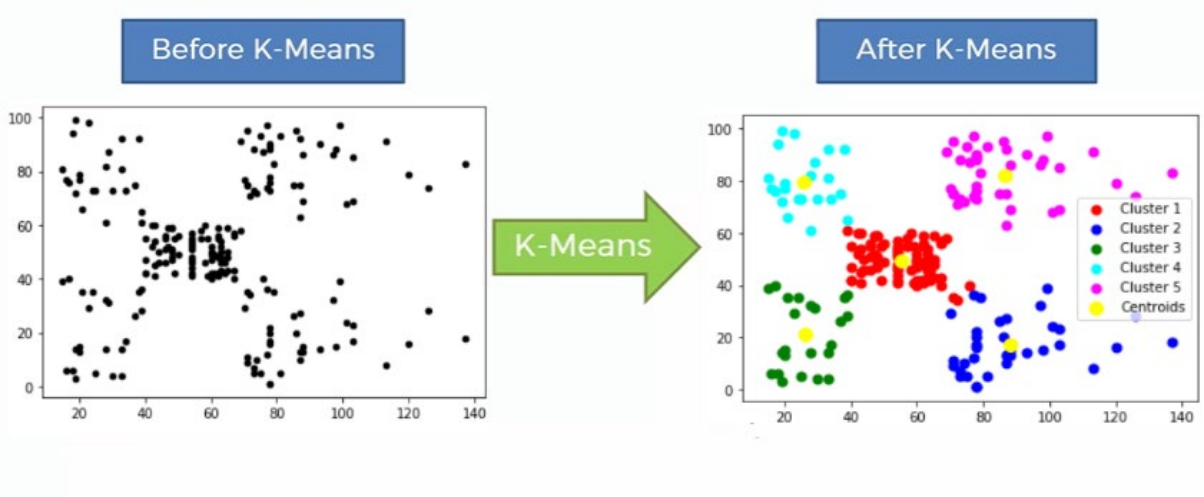
The function of a regression model is to determine a linear function between the X and Y variables that best describes the relationship between the two variables. In linear regression, it's assumed that Y can be calculated from some combination of the input variables. The relationship between the input variables (X) and the target variables (Y) can be portrayed by drawing a line through the points in the graph. The line represents the function that best describes the

relationship between X and Y. The goal is to find an optimal “regression line”, or the line/function that best fits the data.

Lines are typically represented by the equation: Y = m*X + b. X refers to the dependent variable while Y is the independent variable. Meanwhile, m is the slope of the line, as defined by the “rise” over the “run”.

K Means

It is a type of unsupervised algorithm which solves the clustering problem. Its procedure follows a simple and easy way to classify a given data set through a certain number of clusters (assume k clusters). Data points inside a cluster are homogeneous and heterogeneous to peer groups.



How K-means forms cluster:

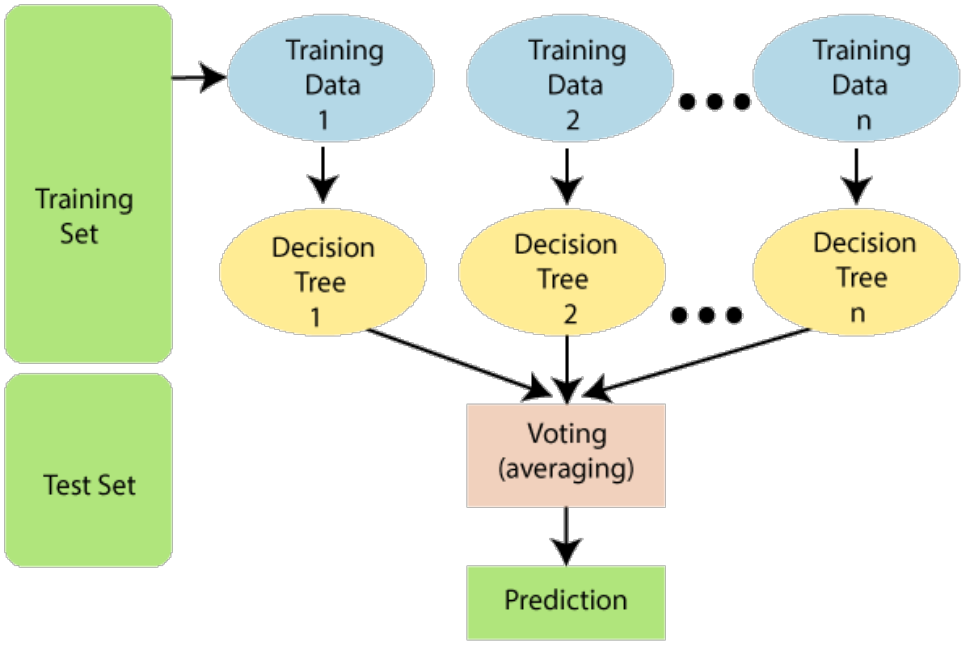
1. K-means picks k number of points for each cluster known as centroids.
2. Each data point forms a cluster with the closest centroids i.e. k clusters.
3. Finds the centroid of each cluster based on existing cluster members. Here we have new centroids.
4. As we have new centroids, repeat step 2 and 3. Find the closest distance for each data point from new centroids and get associated with new k-clusters. Repeat this process until convergence occurs i.e. centroids does not change.

In K-means, we have clusters and each cluster has its own centroid. Sum of square of difference between centroid and the data points within a cluster constitutes within sum of square value for that cluster. Also, when the sum of square values for all the clusters are added, it becomes total within sum of square value for the cluster solution.

We know that as the number of cluster increases, this value keeps on decreasing but if you plot the result you may see that the sum of squared distance decreases sharply up to some value of k, and then much more slowly after that. Here, we can find the optimum number of cluster.

Random Forest

Random Forest is a powerful and versatile supervised machine learning algorithm that grows and combines multiple decision trees to create a "forest." It can be used for both classification and regression problems in R and Python.



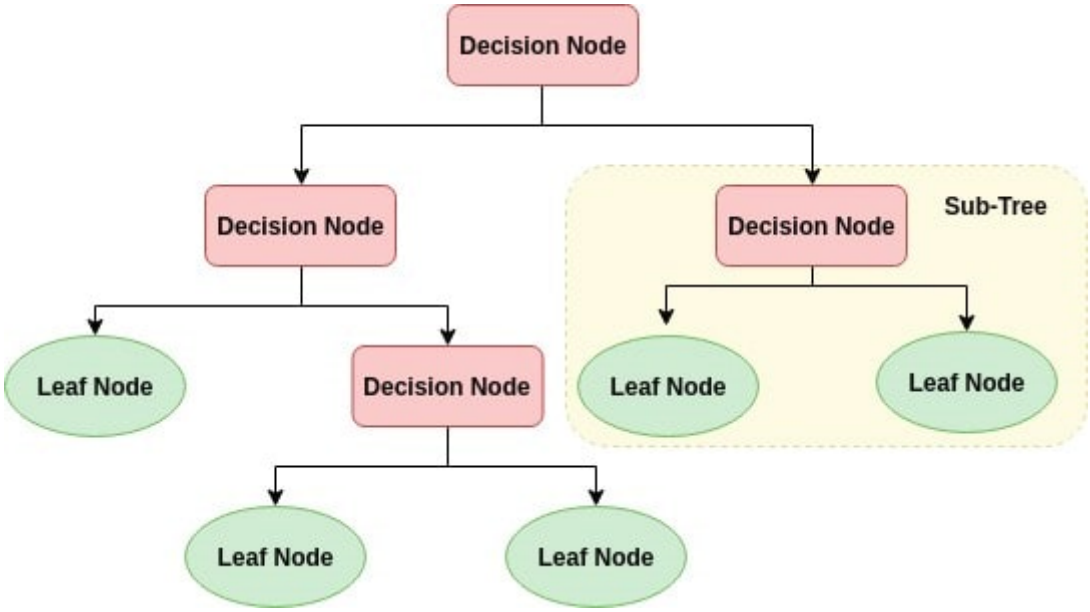
Random Forest is a trademark term for an ensemble of decision trees. In Random Forest, we've collection of decision trees (so known as "Forest"). To classify a new object based on attributes, each tree gives a classification and we say the tree "votes" for that class. The forest chooses the classification having the most votes (over all the trees in the forest).

Each tree is planted & grown as follows:

1. If the number of cases in the training set is N, then sample of N cases is taken at random but *with replacement*. This sample will be the training set for growing the tree.
2. If there are M input variables, a number m<<M is specified such that at each node, m variables are selected at random out of the M and the best split on these m is used to split the node. The value of m is held constant during the forest growing.
3. Each tree is grown to the largest extent possible. There is no pruning.

Decision Tree

It is a type of supervised learning algorithm that is mostly used for classification problems. Surprisingly, it works for both categorical and continuous dependent variables. In this algorithm, we split the population into two or more homogeneous sets. This is done based on most significant attributes/ independent variables to make as distinct groups as possible. It's probably much easier to understand how a decision tree works through an example.



Imagine that our dataset consists of the numbers at the top of the figure to the left. We have two 1s and five 0s (1s and 0s are our classes) and desire to separate the classes using their features. The features are color (red vs. blue) and whether the observation is underlined or not.

Color seems like a pretty obvious feature to split by as all but one of the 0s are blue. So we can use the question, "Is it red?" to split our first node. You can think of a node in a tree as the point where the path splits into two — observations that meet the criteria go down the Yes branch and ones that don't go down the No branch.

The No branch (the blues) is all 0s now so we are done there, but our Yes branch can still be split further. Now we can use the second feature and ask, "Is it underlined?" to make a second split.

The two 1s that are underlined go down the Yes subbranch and the 0 that is not underlined goes down the right subbranch and we are all done. Our decision tree was able to use the two features to split up the data perfectly.

Accuracy score graph

Accuracy score means how accurate our model is. F1 score is used to measure test accuracy. The F1 score is generally a harmonic mean between recall and precision. It can tell how robust your classifier is and how many instances it classifies accurately. Lower recall and higher precision can give you accurate results. Still, it misses a large number of instances that cannot be classified. The best score depends on the type of predictive modeling problem. For the classification problem, the best score of the model is 100% accuracy. For the regression problem, the best score of the model is 0% error.

RESULTS AND DISCUSSION

- Total percentage of people killed in each year.
    - 23.90% people were killed in Road accidents in 2013.
    - 24.00% people were killed in Road accidents in 2014.
    - 25.80% people were killed in Road accidents in 2015.
    - 25.31% people were killed in Road accidents in 2016

- Total percentage of people killed in each year.
    - 25.19% people were injured in Road accidents in 2013.
    - 25.53% people were injured in Road accidents in 2014.
    - 25.24% people were injured in Road accidents in 2015.
    - 23.04% people were injured in Road accidents in 2016

- Accuracy in Logistic Regression – 99.40%
- Accuracy in Random Forest – 44.25%
- Accuracy in DTree – 47.87%

PROGRAM CODE AND OUTPUTS

EDA - Road Accidents

May 31, 2021

[1]:
```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
```

[2]:
```
import matplotlib as mpl
import sklearn
import csv
import os
import xlrd
from collections import defaultdict
import math as m
from jupyterthemes import jtplot
```

[3]:
```
from plotly.offline import init_notebook_mode, iplot
import plotly.figure_factory as ff
import plotly.graph_objs as go
from sklearn.preprocessing import LabelEncoder
from collections import Counter
from sklearn.preprocessing import StandardScaler
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.tree import DecisionTreeClassifier
from sklearn.metrics import classification_report,confusion_matrix
from sklearn.cluster import KMeans
from sklearn.ensemble import RandomForestClassifier
from sklearn import metrics
from sklearn.metrics import mean_squared_error
from sklearn import tree
import plotly.graph_objs as go
import statistics as st
from sklearn.datasets import load_digits
from sklearn.decomposition import PCA
import squarify
from sklearn import tree
import plotly.graph_objs as go
```

1 reading__data

[4]:
```
killed_df =  pd.read_csv("Road_Accidents_2017-Annuxure_Tables_3.csv")
injured_df =  pd.read_csv("Road_Accidents_2017-Annuxure_Tables_4.csv")

weather_df =  pd.
 ↪read_csv("Acc_Classified_according_to_Type_of_Weather_Condition_2014_and_2016.
 ↪csv")
roadcond_df =  pd.read_csv("Acc_clf_acco_to_Road_Cond_2014_and_2016.csv")
```

[ ]:
```

```

[5]:
```
killed_df.describe().T
```

[5]:

	count	mean	std	min
State/UT-wise Total Number of Persons Killed in…	37.0	7549.783784	22784.381050	0.0
State/UT-wise Total Number of Persons Killed in…	37.0	7899.081081	23843.582649	0.0
State/UT-wise Total Number of Persons Killed in…	37.0	8150.540541	24622.961299	1.0
State/UT-wise Total Number of Persons Killed in…	37.0	7995.297297	24156.877674	0.0
Share of States/UTs in Total Number of Persons …	37.0	5.405405	16.314801	0.0
Share of States/UTs in Total Number of Persons …	37.0	5.400000	16.318258	0.0
Share of States/UTs in Total Number of Persons …	37.0	5.402703	16.330620	0.0
Share of States/UTs in Total Number of Persons …	37.0	5.405405	16.332346	0.0
Total Number of Persons Killed in Road Accident…	36.0	10.041667	4.722673	0.0
Total Number of Persons Killed in Road Accident…	36.0	10.333333	4.934427	0.0
Total Number of Persons Killed in Road Accident…	36.0	10.563889	5.255210	1.2
Total Number of Persons Killed in Road Accident…	36.0	10.083333	4.960328	0.0
Total Number of Persons Killed in Road Accident…	37.0	7.021622		
Total Number of Persons Killed in Road Accident…	37.0	6.602703		
Total Number of Persons Killed in Road Accident…	37.0	6.064865		
Total Number of Persons Killed in Road Accident…	37.0	314.051351		
Total Number of Persons Killed in Road Accident…	37.0	330.075676		
Total Number of Persons Killed in Road Accident…	37.0	330.383784		

Total Number of Persons Killed in Road Accident…	3.715070	0.0
Total Number of Persons Killed in Road Accident…	3.526210	0.0
Total Number of Persons Killed in Road Accident…	3.508261	0.6
Total Number of Persons Killed in Road Accident…	233.799622	0.0
Total Number of Persons Killed in Road Accident…	250.954976	0.0
Total Number of Persons Killed in Road Accident…	258.756731	12.7

	25%	50% \
State/UT-wise Total Number of Persons Killed in…	141.000	2522.00
State/UT-wise Total Number of Persons Killed in…	139.000	2397.00
State/UT-wise Total Number of Persons Killed in…	150.000	2572.00
State/UT-wise Total Number of Persons Killed in…	136.000	2783.00
Share of States/UTs in Total Number of Persons …	0.100	1.80
Share of States/UTs in Total Number of Persons …	0.100	1.60
Share of States/UTs in Total Number of Persons …	0.100	1.70
Share of States/UTs in Total Number of Persons …	0.100	1.90
Total Number of Persons Killed in Road Accident…	7.425	9.30
Total Number of Persons Killed in Road Accident…	7.350	9.95
Total Number of Persons Killed in Road Accident…	7.375	10.65
Total Number of Persons Killed in Road Accident…	6.400	9.70
Total Number of Persons Killed in Road Accident…	4.600	7.30
Total Number of Persons Killed in Road Accident…	4.300	6.90
Total Number of Persons Killed in Road Accident…	3.600	5.60
Total Number of Persons Killed in Road Accident…	138.800	253.70
Total Number of Persons Killed in Road Accident…	145.000	263.10
Total Number of Persons Killed in Road Accident…	113.900	237.40

	75%	max
State/UT-wise Total Number of Persons Killed in…	6906.000	139671.0
State/UT-wise Total Number of Persons Killed in…	7110.000	146133.0
State/UT-wise Total Number of Persons Killed in…	7219.000	150785.0
State/UT-wise Total Number of Persons Killed in…	6596.000	147913.0
Share of States/UTs in Total Number of Persons …	4.900	100.0
Share of States/UTs in Total Number of Persons …	4.900	100.0
Share of States/UTs in Total Number of Persons …	4.800	100.0
Share of States/UTs in Total Number of Persons …	4.500	100.0
Total Number of Persons Killed in Road Accident…	13.375	22.1
Total Number of Persons Killed in Road Accident…	13.525	22.7
Total Number of Persons Killed in Road Accident…	13.400	24.8
Total Number of Persons Killed in Road Accident…	13.025	23.2
Total Number of Persons Killed in Road Accident…	8.800	15.3
Total Number of Persons Killed in Road Accident…	8.400	16.2
Total Number of Persons Killed in Road Accident…	8.100	17.3
Total Number of Persons Killed in Road Accident…	444.000	1050.3
Total Number of Persons Killed in Road Accident…	463.500	1054.1
Total Number of Persons Killed in Road Accident…	457.400	1036.3

[6]:
```
injured_df.describe().T
```

[6]:

	count	mean	std	min	25%	50%
State/UT-wise Total Number of Persons Injured i…	37.0	26674.270270	81101.786738	1.0	352.000	6499.00
State/UT-wise Total Number of Persons Injured i…	37.0	27042.108108	82222.705770	3.0	359.000	6835.00
State/UT-wise Total Number of Persons Injured i…	37.0	26736.432432	81380.111582	0.0	391.000	5764.00
State/UT-wise Total Number of Persons Injured i…	37.0	25458.108108	77452.603468	1.0	479.000	6014.00
Share of States/UTs in Total Number of Persons …	37.0	5.405405	16.435783	0.0		
Share of States/UTs in Total Number of Persons …	37.0	5.408108	16.433999	0.0		
Share of States/UTs in Total Number of Persons …	37.0	5.410811	16.450967	0.0		
Share of States/UTs in Total Number of Persons …	37.0	5.408108	16.442482	0.0		
Total Number of Persons Injured in Road Acciden…	36.0	41.019444	31.078636	1.3		
Total Number of Persons Injured in Road Acciden…	36.0	41.205556	32.660845	3.1		
Total Number of Persons Injured in Road Acciden…	36.0	40.450000	33.392313	0.0		
Total Number of Persons Injured in Road Acciden…	36.0	38.350000	32.010797	1.2		
Total Number of Persons injured in Road Acciden…	37.0	26.370270	19.665803	0.8		
Total Number of Persons injured in Road Acciden…	37.0	24.132432	17.607622	2.1		
Total Number of Persons injured in Road Acciden…	37.0	21.113514	15.169618	0.0		
Total Number of Persons injured in Road Acciden…	37.0	1168.478378	972.679941	48.1		
Total Number of Persons injured in Road Acciden…	37.0	1189.013514	1057.068114	19.9		
Total Number of Persons injured in Road Acciden…	37.0	1197.070270	1197.160351	0.0		

```
Share of States/UTs in Total Number of Persons …    0.100     1.30
Share of States/UTs in Total Number of Persons …    0.100     1.40
Share of States/UTs in Total Number of Persons …    0.100     1.20
Share of States/UTs in Total Number of Persons …    0.100     1.30
Total Number of Persons Injured in Road Acciden…   15.775    33.95
Total Number of Persons Injured in Road Acciden…   15.675    33.50
Total Number of Persons Injured in Road Acciden…   15.950    31.35
Total Number of Persons Injured in Road Acciden…   15.000    26.40
Total Number of Persons injured in Road Acciden…   12.400    20.40
Total Number of Persons injured in Road Acciden…   10.700    19.50
Total Number of Persons injured in Road Acciden…    9.800    17.70
Total Number of Persons injured in Road Acciden…  383.000  1055.80
Total Number of Persons injured in Road Acciden…  398.400  1053.90
Total Number of Persons injured in Road Acciden…  319.100  1033.70
```

	75%	max
State/UT-wise Total Number of Persons Injured i…	22337.000	493474.0
State/UT-wise Total Number of Persons Injured i…	22948.000	500279.0
State/UT-wise Total Number of Persons Injured i…	24103.000	494624.0
State/UT-wise Total Number of Persons Injured i…	22071.000	470975.0
Share of States/UTs in Total Number of Persons …	4.500	100.0
Share of States/UTs in Total Number of Persons …	4.600	100.0
Share of States/UTs in Total Number of Persons …	4.900	100.0
Share of States/UTs in Total Number of Persons …	4.700	100.0
Total Number of Persons Injured in Road Acciden…	53.725	116.6
Total Number of Persons Injured in Road Acciden…	54.775	123.3
Total Number of Persons Injured in Road Acciden…	52.225	123.6
Total Number of Persons Injured in Road Acciden…	50.700	119.0
Total Number of Persons injured in Road Acciden…	34.000	88.9
Total Number of Persons injured in Road Acciden…	35.400	78.0
Total Number of Persons injured in Road Acciden…	29.100	56.3
Total Number of Persons injured in Road Acciden…	1814.600	4380.9
Total Number of Persons injured in Road Acciden…	1770.300	5051.9
Total Number of Persons injured in Road Acciden…	1723.400	5627.7

[ ]:

2 Finding variance

[7]:
```
killed_df.var()
```

[7]:
```
State/UT-wise Total Number of Persons Killed in Road Accidents during - 2014
5.191280e+08
State/UT-wise Total Number of Persons Killed in Road Accidents during - 2015
5.685164e+08
State/UT-wise Total Number of Persons Killed in Road Accidents during - 2016
```

```
6.062902e+08
State/UT-wise Total Number of Persons Killed in Road Accidents during - 2017
5.835547e+08
Share of States/UTs in Total Number of Persons Killed in Road Accidents - 2014
2.661727e+02
Share of States/UTs in Total Number of Persons Killed in Road Accidents - 2015
2.662856e+02
Share of States/UTs in Total Number of Persons Killed in Road Accidents - 2016
2.666892e+02
Share of States/UTs in Total Number of Persons Killed in Road Accidents - 2017
2.667455e+02
Total Number of Persons Killed in Road Accidents Per Lakh Population - 2014
2.230364e+01
Total Number of Persons Killed in Road Accidents Per Lakh Population - 2015
2.434857e+01
Total Number of Persons Killed in Road Accidents Per Lakh Population - 2016
2.761723e+01
Total Number of Persons Killed in Road Accidents Per Lakh Population - 2017
2.460486e+01
Total Number of Persons Killed in Road Accidents per 10,000 Vehicles - 2014
1.380174e+01
Total Number of Persons Killed in Road Accidents per 10,000 Vehicles - 2015
1.243416e+01
Total Number of Persons Killed in Road Accidents per 10,000 Vehicles - 2016
1.230790e+01
Total Number of Persons Killed in Road Accidents per 10,000 Km of Roads - 2014
5.466226e+04
Total Number of Persons Killed in Road Accidents per 10,000 Km of Roads - 2015
6.297840e+04
Total Number of Persons Killed in Road Accidents per 10,000 Km of Roads - 2016
6.695505e+04
dtype: float64
```

[8]:
```
injured_df.var()
```

[8]:
```
State/UT-wise Total Number of Persons Injured in Road Accidents during - 2014
6.577500e+09
State/UT-wise Total Number of Persons Injured in Road Accidents during - 2015
6.760573e+09
State/UT-wise Total Number of Persons Injured in Road Accidents during - 2016
6.622723e+09
State/UT-wise Total Number of Persons Injured in Road Accidents during - 2017
5.998906e+09
Share of States/UTs in Total Number of Persons Injured in Road Accidents - 2014
2.701350e+02
Share of States/UTs in Total Number of Persons Injured in Road Accidents - 2015
2.700763e+02
```

```
Share of States/UTs in Total Number of Persons Injured in Road Accidents - 2016
2.706343e+02
Share of States/UTs in Total Number of Persons Injured in Road Accidents - 2017
2.703552e+02
Total Number of Persons Injured in Road Accidents Per Lakh Population - 2014
9.658816e+02
Total Number of Persons Injured in Road Accidents Per Lakh Population - 2015
1.066731e+03
Total Number of Persons Injured in Road Accidents Per Lakh Population - 2016
1.115047e+03
Total Number of Persons Injured in Road Accidents Per Lakh Population - 2017
1.024691e+03
Total Number of Persons injured in Road Accidents per 10,000 Vehicles - 2014
3.867438e+02
Total Number of Persons injured in Road Accidents per 10,000 Vehicles - 2015
3.100284e+02
Total Number of Persons injured in Road Accidents per 10,000 Vehicles - 2016
2.301173e+02
Total Number of Persons injured in Road Accidents per 10,000 Km of Roads - 2014
9.461063e+05
Total Number of Persons injured in Road Accidents per 10,000 Km of Roads - 2015
1.117393e+06
Total Number of Persons injured in Road Accidents per 10,000 Km of Roads - 2016
1.433193e+06
dtype: float64
```

[9]:
```
weather_df.var()
```

[9]:
```
Fine - Total Acc. - 2014              2.707662e+09
Fine - Persons Killed - 2014          1.884333e+08
Fine - Persons Injured - 2014         2.747724e+09
Mist/fog - Total Acc. - 2014          8.030572e+06
Mist/fog - Persons Killed - 2014      9.393587e+05
                                            ...
Dust Storm - Persons Killed - 2016    1.361887e+05
Dust Storm - Persons Injured - 2016   9.650520e+05
Others - Total Accidents - 2016       7.393995e+07
Others - Persons Killed - 2016        1.052093e+07
Others - Persons Injured - 2016       7.360387e+07
Length: 63, dtype: float64
```

[10]:
```
roadcond_df.var()
```

[10]:
```
Surfaced Roads-Accident - 2014                    3.262928e+09
Surfaced Roads- Killed - 2014                     2.468370e+08
Surfaced Roads-Injured - 2014                     3.127718e+09
Metalled Roads-Accident - 2014                    2.003219e+08
```

```
Metalled Roads- Killed - 2014                              2.162614e+07
                                                                 ...
Earthern Shoulder Edge Drop - Persons Killed - 2016        7.343631e+04
Earthern Shoulder Edge Drop - Persons Injured - 2016       3.482978e+05
Others - Number of Accidents - 2016                        1.080777e+08
Others - Persons Killed - 2016                             1.478221e+07
Others - Persons Injured - 2016                            1.185360e+08
Length: 96, dtype: float64
```

3 Finding Inter Quartile Range (IQR) for the columns

[11]:
```
killed_df.quantile(0.75) - killed_df.quantile(0.25)
```

[11]:
```
State/UT-wise Total Number of Persons Killed in Road Accidents during - 2014
6765.000
State/UT-wise Total Number of Persons Killed in Road Accidents during - 2015
6971.000
State/UT-wise Total Number of Persons Killed in Road Accidents during - 2016
7069.000
State/UT-wise Total Number of Persons Killed in Road Accidents during - 2017
6460.000
Share of States/UTs in Total Number of Persons Killed in Road Accidents - 2014
4.800
Share of States/UTs in Total Number of Persons Killed in Road Accidents - 2015
4.800
Share of States/UTs in Total Number of Persons Killed in Road Accidents - 2016
4.700
Share of States/UTs in Total Number of Persons Killed in Road Accidents - 2017
4.400
Total Number of Persons Killed in Road Accidents Per Lakh Population - 2014
5.950
Total Number of Persons Killed in Road Accidents Per Lakh Population - 2015
6.175
Total Number of Persons Killed in Road Accidents Per Lakh Population - 2016
6.025
Total Number of Persons Killed in Road Accidents Per Lakh Population - 2017
6.625
Total Number of Persons Killed in Road Accidents per 10,000 Vehicles - 2014
4.200
Total Number of Persons Killed in Road Accidents per 10,000 Vehicles - 2015
4.100
Total Number of Persons Killed in Road Accidents per 10,000 Vehicles - 2016
4.500
Total Number of Persons Killed in Road Accidents per 10,000 Km of Roads - 2014
305.200
Total Number of Persons Killed in Road Accidents per 10,000 Km of Roads - 2015
```

```
318.500
Total Number of Persons Killed in Road Accidents per 10,000 Km of Roads - 2016
343.500
dtype: float64
```

[12]:
```
injured_df.quantile(0.75) - injured_df.quantile(0.25)
```

[12]:
```
State/UT-wise Total Number of Persons Injured in Road Accidents during - 2014
21985.000
State/UT-wise Total Number of Persons Injured in Road Accidents during - 2015
22589.000
State/UT-wise Total Number of Persons Injured in Road Accidents during - 2016
23712.000
State/UT-wise Total Number of Persons Injured in Road Accidents during - 2017
21592.000
Share of States/UTs in Total Number of Persons Injured in Road Accidents - 2014
4.400
Share of States/UTs in Total Number of Persons Injured in Road Accidents - 2015
4.500
Share of States/UTs in Total Number of Persons Injured in Road Accidents - 2016
4.800
Share of States/UTs in Total Number of Persons Injured in Road Accidents - 2017
4.600
Total Number of Persons Injured in Road Accidents Per Lakh Population - 2014
37.950
Total Number of Persons Injured in Road Accidents Per Lakh Population - 2015
39.100
Total Number of Persons Injured in Road Accidents Per Lakh Population - 2016
36.275
Total Number of Persons Injured in Road Accidents Per Lakh Population - 2017
35.700
Total Number of Persons injured in Road Accidents per 10,000 Vehicles - 2014
21.600
Total Number of Persons injured in Road Accidents per 10,000 Vehicles - 2015
24.700
Total Number of Persons injured in Road Accidents per 10,000 Vehicles - 2016
19.300
Total Number of Persons injured in Road Accidents per 10,000 Km of Roads - 2014
1431.600
Total Number of Persons injured in Road Accidents per 10,000 Km of Roads - 2015
1371.900
Total Number of Persons injured in Road Accidents per 10,000 Km of Roads - 2016
1404.300
dtype: float64
```

4 Finding Kurtosis values for the columns

[13]:
```
killed_df.kurt()
```

[13]:
```
State/UT-wise Total Number of Persons Killed in Road Accidents during - 2014
33.849101
State/UT-wise Total Number of Persons Killed in Road Accidents during - 2015
33.817799
State/UT-wise Total Number of Persons Killed in Road Accidents during - 2016
33.696072
State/UT-wise Total Number of Persons Killed in Road Accidents during - 2017
33.678742
Share of States/UTs in Total Number of Persons Killed in Road Accidents - 2014
33.831795
Share of States/UTs in Total Number of Persons Killed in Road Accidents - 2015
33.809029
Share of States/UTs in Total Number of Persons Killed in Road Accidents - 2016
33.693309
Share of States/UTs in Total Number of Persons Killed in Road Accidents - 2017
33.673949
Total Number of Persons Killed in Road Accidents Per Lakh Population - 2014
0.046921
Total Number of Persons Killed in Road Accidents Per Lakh Population - 2015
0.044539
Total Number of Persons Killed in Road Accidents Per Lakh Population - 2016
0.220121
Total Number of Persons Killed in Road Accidents Per Lakh Population - 2017
0.239057
Total Number of Persons Killed in Road Accidents per 10,000 Vehicles - 2014
-0.301367
Total Number of Persons Killed in Road Accidents per 10,000 Vehicles - 2015
0.455503
Total Number of Persons Killed in Road Accidents per 10,000 Vehicles - 2016
1.452087
Total Number of Persons Killed in Road Accidents per 10,000 Km of Roads - 2014
1.308746
Total Number of Persons Killed in Road Accidents per 10,000 Km of Roads - 2015
0.561920
Total Number of Persons Killed in Road Accidents per 10,000 Km of Roads - 2016
0.438050
dtype: float64
```

[14]:
```
injured_df.kurt()
```

[14]:
```
State/UT-wise Total Number of Persons Injured in Road Accidents during - 2014
32.764926
State/UT-wise Total Number of Persons Injured in Road Accidents during - 2015
```

```
32.760816
State/UT-wise Total Number of Persons Injured in Road Accidents during - 2016
32.609010
State/UT-wise Total Number of Persons Injured in Road Accidents during - 2017
32.675314
Share of States/UTs in Total Number of Persons Injured in Road Accidents - 2014
32.756963
Share of States/UTs in Total Number of Persons Injured in Road Accidents - 2015
32.768652
Share of States/UTs in Total Number of Persons Injured in Road Accidents - 2016
32.617807
Share of States/UTs in Total Number of Persons Injured in Road Accidents - 2017
32.694527
Total Number of Persons Injured in Road Accidents Per Lakh Population - 2014
0.164145
Total Number of Persons Injured in Road Accidents Per Lakh Population - 2015
0.351507
Total Number of Persons Injured in Road Accidents Per Lakh Population - 2016
0.532669
Total Number of Persons Injured in Road Accidents Per Lakh Population - 2017
0.197017
Total Number of Persons injured in Road Accidents per 10,000 Vehicles - 2014
1.807142
Total Number of Persons injured in Road Accidents per 10,000 Vehicles - 2015
1.372358
Total Number of Persons injured in Road Accidents per 10,000 Vehicles - 2016
-0.019793
Total Number of Persons injured in Road Accidents per 10,000 Km of Roads - 2014
1.829573
Total Number of Persons injured in Road Accidents per 10,000 Km of Roads - 2015
3.506893
Total Number of Persons injured in Road Accidents per 10,000 Km of Roads - 2016
4.483892
dtype: float64
```

5 Searching for null values to do perform data analysis

[15]:
```
killed_df.isnull().sum()
```

[15]:
```
States/UTs
0
State/UT-wise Total Number of Persons Killed in Road Accidents during - 2014
0
State/UT-wise Total Number of Persons Killed in Road Accidents during - 2015
0
State/UT-wise Total Number of Persons Killed in Road Accidents during - 2016
```

```
0
State/UT-wise Total Number of Persons Killed in Road Accidents during - 2017
0
Share of States/UTs in Total Number of Persons Killed in Road Accidents - 2014
0
Share of States/UTs in Total Number of Persons Killed in Road Accidents - 2015
0
Share of States/UTs in Total Number of Persons Killed in Road Accidents - 2016
0
Share of States/UTs in Total Number of Persons Killed in Road Accidents - 2017
0
Total Number of Persons Killed in Road Accidents Per Lakh Population - 2014
1
Total Number of Persons Killed in Road Accidents Per Lakh Population - 2015
1
Total Number of Persons Killed in Road Accidents Per Lakh Population - 2016
1
Total Number of Persons Killed in Road Accidents Per Lakh Population - 2017
1
Total Number of Persons Killed in Road Accidents per 10,000 Vehicles - 2014
0
Total Number of Persons Killed in Road Accidents per 10,000 Vehicles - 2015
0
Total Number of Persons Killed in Road Accidents per 10,000 Vehicles - 2016
0
Total Number of Persons Killed in Road Accidents per 10,000 Km of Roads - 2014
0
Total Number of Persons Killed in Road Accidents per 10,000 Km of Roads - 2015
0
Total Number of Persons Killed in Road Accidents per 10,000 Km of Roads - 2016
0
dtype: int64
```

[16]:
```
injured_df.isnull().sum()
```

[16]:
```
States/UTs
0
State/UT-wise Total Number of Persons Injured in Road Accidents during - 2014
0
State/UT-wise Total Number of Persons Injured in Road Accidents during - 2015
0
State/UT-wise Total Number of Persons Injured in Road Accidents during - 2016
0
State/UT-wise Total Number of Persons Injured in Road Accidents during - 2017
0
Share of States/UTs in Total Number of Persons Injured in Road Accidents - 2014
0
```

```
Share of States/UTs in Total Number of Persons Injured in Road Accidents - 2015
0
Share of States/UTs in Total Number of Persons Injured in Road Accidents - 2016
0
Share of States/UTs in Total Number of Persons Injured in Road Accidents - 2017
0
Total Number of Persons Injured in Road Accidents Per Lakh Population - 2014
1
Total Number of Persons Injured in Road Accidents Per Lakh Population - 2015
1
Total Number of Persons Injured in Road Accidents Per Lakh Population - 2016
1
Total Number of Persons Injured in Road Accidents Per Lakh Population - 2017
1
Total Number of Persons injured in Road Accidents per 10,000 Vehicles - 2014
0
Total Number of Persons injured in Road Accidents per 10,000 Vehicles - 2015
0
Total Number of Persons injured in Road Accidents per 10,000 Vehicles - 2016
0
Total Number of Persons injured in Road Accidents per 10,000 Km of Roads - 2014
0
Total Number of Persons injured in Road Accidents per 10,000 Km of Roads - 2015
0
Total Number of Persons injured in Road Accidents per 10,000 Km of Roads - 2016
0
dtype: int64
```

[ ]:

[ ]:

[ ]:

[ ]:

[ ]:

5.0.1 Dataset : Killed by Accident (Road_Accidents_2017-Annuxure_Tables_3.csv)

[17]:
```
killed_df.head()
```

[17]:
```
          States/UTs  \
0     Andhra Pradesh
1  Arunachal Pradesh
2              Assam
3              Bihar
```

```
4       Chhattisgarh

   State/UT-wise Total Number of Persons Killed in Road Accidents during - 2014
\
0                                                 7908
1                                                  119
2                                                 2522
3                                                 4913
4                                                 4022

   State/UT-wise Total Number of Persons Killed in Road Accidents during - 2015
\
0                                                 8297
1                                                  127
2                                                 2397
3                                                 5421
4                                                 4082

   State/UT-wise Total Number of Persons Killed in Road Accidents during - 2016
\
0                                                 8541
1                                                  149
2                                                 2572
3                                                 4901
4                                                 3908

   State/UT-wise Total Number of Persons Killed in Road Accidents during - 2017
\
0                                                 8060
1                                                  110
2                                                 2783
3                                                 5554
4                                                 4136

   Share of States/UTs in Total Number of Persons Killed in Road Accidents -
2014  \
0                                                  5.7
1                                                  0.1
2                                                  1.8
3                                                  3.5
4                                                  2.9

   Share of States/UTs in Total Number of Persons Killed in Road Accidents -
2015  \
0                                                  5.7
1                                                  0.1
2                                                  1.6
```

```
3                                                 3.7
4                                                 2.8

   Share of States/UTs in Total Number of Persons Killed in Road Accidents -
2016  \
0                                                 5.7
1                                                 0.1
2                                                 1.7
3                                                 3.3
4                                                 2.6

   Share of States/UTs in Total Number of Persons Killed in Road Accidents -
2017  \
0                                                 5.4
1                                                 0.1
2                                                 1.9
3                                                 3.8
4                                                 2.8

   Total Number of Persons Killed in Road Accidents Per Lakh Population - 2014
\
0                                                 9.1
1                                                 9.3
2                                                 8.0
3                                                 4.8
4                                                15.9

   Total Number of Persons Killed in Road Accidents Per Lakh Population - 2015
\
0                                                 9.5
1                                                 9.8
2                                                 7.5
3                                                 5.3
4                                                16.0

   Total Number of Persons Killed in Road Accidents Per Lakh Population - 2016
\
0                                                 9.7
1                                                11.3
2                                                 7.9
3                                                 4.7
4                                                15.1

   Total Number of Persons Killed in Road Accidents Per Lakh Population - 2017
\
0                                                 9.1
1                                                 8.3
```

```
2                                                 8.5
3                                                 5.3
4                                                15.8

   Total Number of Persons Killed in Road Accidents per 10,000 Vehicles - 2014
\
0                                                11.3
1                                                 7.9
2                                                11.4
3                                                11.8
4                                                10.4

   Total Number of Persons Killed in Road Accidents per 10,000 Vehicles - 2015
\
0                                                10.5
1                                                 8.4
2                                                 9.6
3                                                11.3
4                                                 9.5

   Total Number of Persons Killed in Road Accidents per 10,000 Vehicles - 2016
\
0                                                 9.8
1                                                 5.6
2                                                 9.1
3                                                 8.9
4                                                 8.1

   Total Number of Persons Killed in Road Accidents per 10,000 Km of Roads -
2014  \
0                                               444.0
1                                                48.6
2                                                80.4
3                                               234.2
4                                               425.3

   Total Number of Persons Killed in Road Accidents per 10,000 Km of Roads -
2015  \
0                                               463.5
1                                                50.1
2                                                73.4
3                                               263.1
4                                               418.5

   Total Number of Persons Killed in Road Accidents per 10,000 Km of Roads -
2016
0                                               489.8
```

```
1                                                 48.5
2                                                 78.1
3                                                237.4
4                                                407.9
```

5.0.2 Dataset : Killed by Accident (Road_Accidents_2017-Annuxure_Tables_4.csv)

[18]:
```
injured_df.head()
```

[18]:
```
          States/UTs  \
0    Andhra Pradesh
1  Arunachal Pradesh
2              Assam
3              Bihar
4       Chhattisgarh

   State/UT-wise Total Number of Persons Injured in Road Accidents during - 2014
\
0                                              29931
1                                                308
2                                               6499
3                                               6640
4                                              13157

   State/UT-wise Total Number of Persons Injured in Road Accidents during - 2015
\
0                                              29439
1                                                359
2                                               7068
3                                               6835
4                                              13426

   State/UT-wise Total Number of Persons Injured in Road Accidents during - 2016
\
0                                              30051
1                                                391
2                                               6127
3                                               5651
4                                              12955

   State/UT-wise Total Number of Persons Injured in Road Accidents during - 2017
\
0                                              27475
1                                                316
2                                               6163
3                                               6014
4                                              12550
```

	Share of States/UTs in Total Number of Persons Injured in Road Accidents - 2014
0	6.1
1	0.1
2	1.3
3	1.3
4	2.7

	Share of States/UTs in Total Number of Persons Injured in Road Accidents - 2015
0	5.9
1	0.1
2	1.4
3	1.4
4	2.7

	Share of States/UTs in Total Number of Persons Injured in Road Accidents - 2016
0	6.1
1	0.1
2	1.2
3	1.1
4	2.6

	Share of States/UTs in Total Number of Persons Injured in Road Accidents - 2017
0	5.8
1	0.1
2	1.3
3	1.3
4	2.7

	Total Number of Persons Injured in Road Accidents Per Lakh Population - 2014
0	34.4
1	24.0
2	20.5
3	6.5
4	52.1

	Total Number of Persons Injured in Road Accidents Per Lakh Population - 2015
0	33.6
1	27.6
2	22.0
3	6.7

```
4                                                52.5

   Total Number of Persons Injured in Road Accidents Per Lakh Population - 2016
\
0                                                34.0
1                                                29.8
2                                                18.9
3                                                 5.4
4                                                50.1

   Total Number of Persons Injured in Road Accidents Per Lakh Population - 2017
\
0                                                30.9
1                                                23.8
2                                                18.8
3                                                 5.7
4                                                47.9

   Total Number of Persons injured in Road Accidents per 10,000 Vehicles - 2014
\
0                                                42.7
1                                                20.4
2                                                29.3
3                                                15.9
4                                                34.0

   Total Number of Persons injured in Road Accidents per 10,000 Vehicles - 2015
\
0                                                37.3
1                                                23.7
2                                                28.2
3                                                14.3
4                                                31.1

   Total Number of Persons injured in Road Accidents per 10,000 Vehicles - 2016
\
0                                                34.4
1                                                14.8
2                                                21.7
3                                                10.3
4                                                26.9

   Total Number of Persons injured in Road Accidents per 10,000 Km of Roads -
2014  \
0                                              1680.6
1                                               125.9
2                                               207.2
```

```
3                                                316.6
4                                               1391.1

   Total Number of Persons injured in Road Accidents per 10,000 Km of Roads -
2015  \
0                                               1644.4
1                                                141.6
2                                                216.5
3                                                331.8
4                                               1376.5

   Total Number of Persons injured in Road Accidents per 10,000 Km of Roads -
2016
0                                               1723.4
1                                                127.4
2                                                185.9
3                                                273.7
4                                               1352.2
```

5.0.3 Dataset : Classfied by Weather (Acc_Classified_according_to_Type_of_Weather_Conditi

[19]:
```
weather_df.head()
```

[19]:
```
  S. No.          State/ UT  Fine - Total Acc. - 2014  \
0      1     Andhra Pradesh                   14591.0
1      2  Arunachal Pradesh                      71.0
2      3              Assam                    3575.0
3      4              Bihar                    2343.0
4      5       Chhattisgarh                    5000.0

   Fine - Persons Killed - 2014  Fine - Persons Injured - 2014  \
0                          4586                        17065.0
1                            30                          110.0
2                          1318                         3216.0
3                          1218                         1626.0
4                          1354                         4584.0

   Mist/fog - Total Acc. - 2014  Mist/fog - Persons Killed - 2014  \
0                         724.0                               219
1                          14.0                                10
2                         494.0                               150
3                        1713.0                               881
4                         382.0                               149

   Mist/fog - Persons Injured - 2014  Cloudy - Total Acc. - 2014  \
0                              925.0                       647.0
1                               26.0                        11.0
```

```
2                              368.0                       285.0
3                             1081.0                       438.0
4                              376.0                       863.0

   Cloudy - Persons Killed - 2014  …  Snowfall - Persons Injured - 2016  \
0                           188.0  …                                  0
1                             8.0  …                                  0
2                           100.0  …                                  0
3                           203.0  …                                  0
4                           216.0  …                                  0

   Hail/Sleet - Total Accidents - 2016  Hail/Sleet - Persons Killed - 2016  \
0                                   92                                  45
1                                   12                                   7
2                                   15                                   8
3                                    0                                   0
4                                    0                                   0

   Hail/Sleet - Persons Injured - 2016  Dust Storm - Total Accidents - 2016  \
0                                  137                                  42
1                                   13                                   0
2                                    5                                  36
3                                    0                                 582
4                                    0                                 132

   Dust Storm - Persons Killed - 2016  Dust Storm - Persons Injured - 2016  \
0                                  34                                  61
1                                   0                                   0
2                                  14                                  19
3                                 318                                 399
4                                  28                                 151

   Others - Total Accidents - 2016  Others - Persons Killed - 2016  \
0                             2839                             1086
1                               44                               22
2                             2188                              763
3                             1820                             1175
4                             3972                             1262

   Others - Persons Injured - 2016
0                             3408
1                               91
2                             1752
3                             1247
4                             3793

[5 rows x 65 columns]
```

5.0.4 Dataset : Classfied by Road Conditions (Acc_clf_acco_to_Road_Cond_2014_and_2016.cs

[20]:
```
roadcond_df.head()
```

[20]:
```
  S. No.          State/ UT  Surfaced Roads-Accident - 2014  \
0      1     Andhra Pradesh                         13846.0
1      2  Arunachal Pradesh                            81.0
2      3              Assam                          4948.0
3      4              Bihar                          5990.0
4      5       Chhattisgarh                          7248.0

   Surfaced Roads- Killed - 2014  Surfaced Roads-Injured - 2014  \
0                           4333                        17323.0
1                             43                          111.0
2                           1631                         5043.0
3                           3184                         3732.0
4                           2111                         7170.0

   Metalled Roads-Accident - 2014  Metalled Roads- Killed - 2014  \
0                          5211.0                           2392
1                            73.0                             37
2                          1296.0                            548
3                          2194.0                           1148
4                          4799.0                           1301

   Metalled Roads-Injured - 2014  Kutcha Roads-Accident - 2014  \
0                         4138.0                        5383.0
1                           97.0                          51.0
2                          824.0                         900.0
3                         1601.0                        1372.0
4                         4543.0                        1774.0

   Kutcha Roads- Killed - 2014  …  Sharp Curve - Persons Injured - 2016  \
0                         1183  …                                  1185
1                           39  …                                    65
2                          343  …                                    22
3                          581  …                                   491
4                          610  …                                   152

   Steep Gradient - Number of Accidents - 2016  \
0                                          144
1                                            0
2                                           44
3                                          117
4                                          109

   Steep Gradient - Persons Killed - 2016  \
```

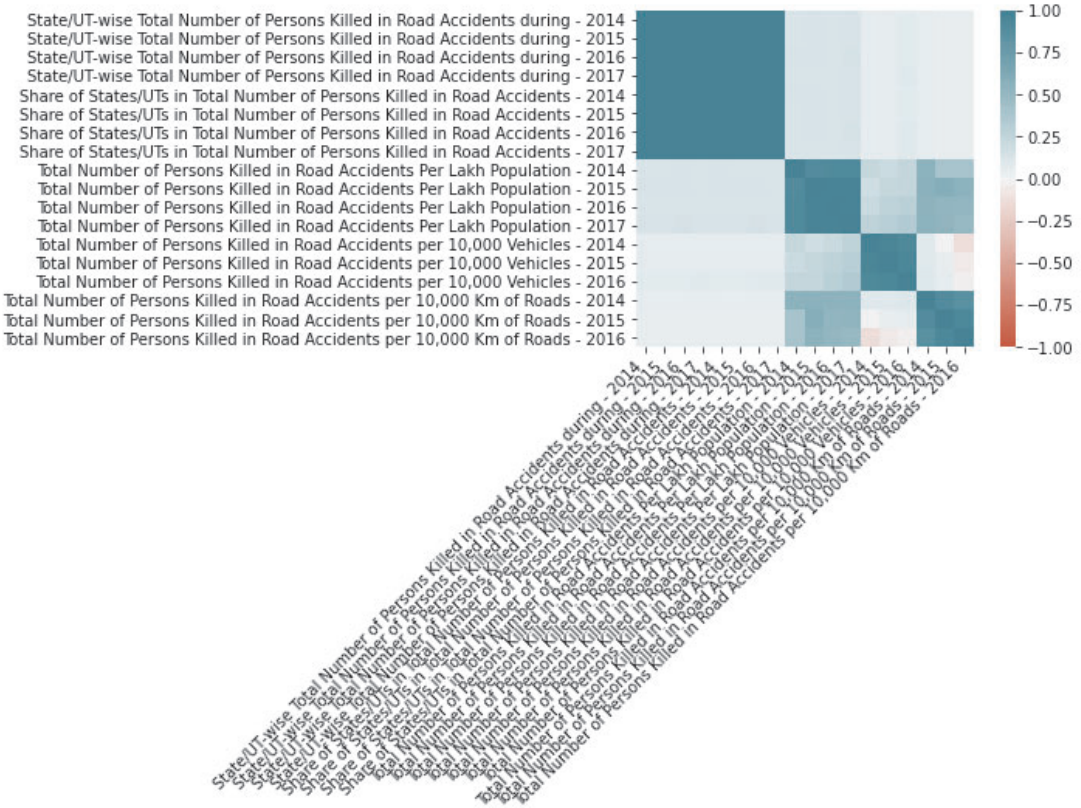
```
0                                    38
1                                     0
2                                     9
3                                    59
4                                    39

   Steep Gradient - Persons Injured - 2016  \
0                                    189
1                                      0
2                                     24
3                                    119
4                                    114

   Earthern Shoulder Edge Drop - Number of Accidents - 2016  \
0                                                  255
1                                                    0
2                                                    2
3                                                  123
4                                                    8

   Earthern Shoulder Edge Drop - Persons Killed - 2016  \
0                                                  145
1                                                    0
2                                                    2
3                                                   65
4                                                    2

   Earthern Shoulder Edge Drop - Persons Injured - 2016  \
0                                                  302
1                                                    0
2                                                    9
3                                                  133
4                                                    8

   Others - Number of Accidents - 2016  Others - Persons Killed - 2016  \
0                                 3651                            1290
1                                   44                              22
2                                 1634                             645
3                                 1328                             773
4                                 1499                             453

   Others - Persons Injured - 2016
0                             4420
1                               91
2                             1153
3                              944
4                             1329
```

```
[5 rows x 98 columns]
```

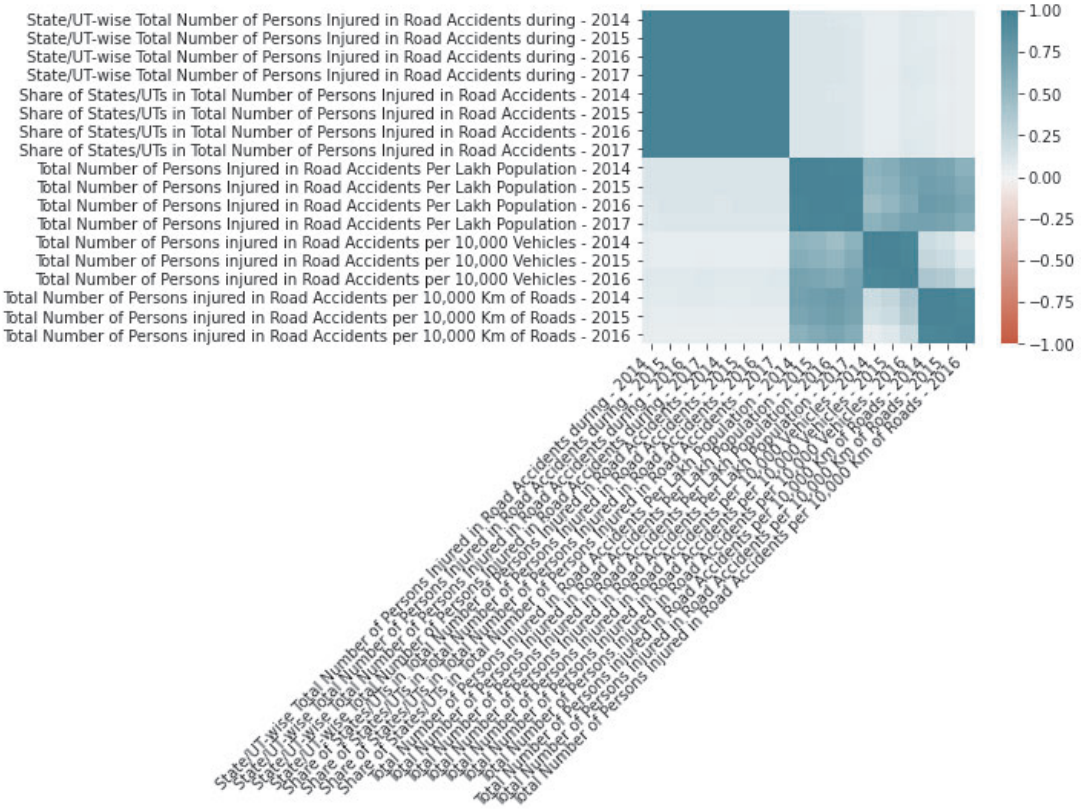
6 Heatmaps and Correlation Matrix Plots

[ ]:

[21]:
```
data1 = pd.read_csv("Road_Accidents_2017-Annuxure_Tables_3.csv")

corr = data1.corr()
ax = sns.heatmap(
    corr,
    vmin=-1, vmax=1, center=0,
    cmap=sns.diverging_palette(20, 220, n=200),
    square=True
)
ax.set_xticklabels(
    ax.get_xticklabels(),
    rotation=45,
    horizontalalignment='right'
);
```



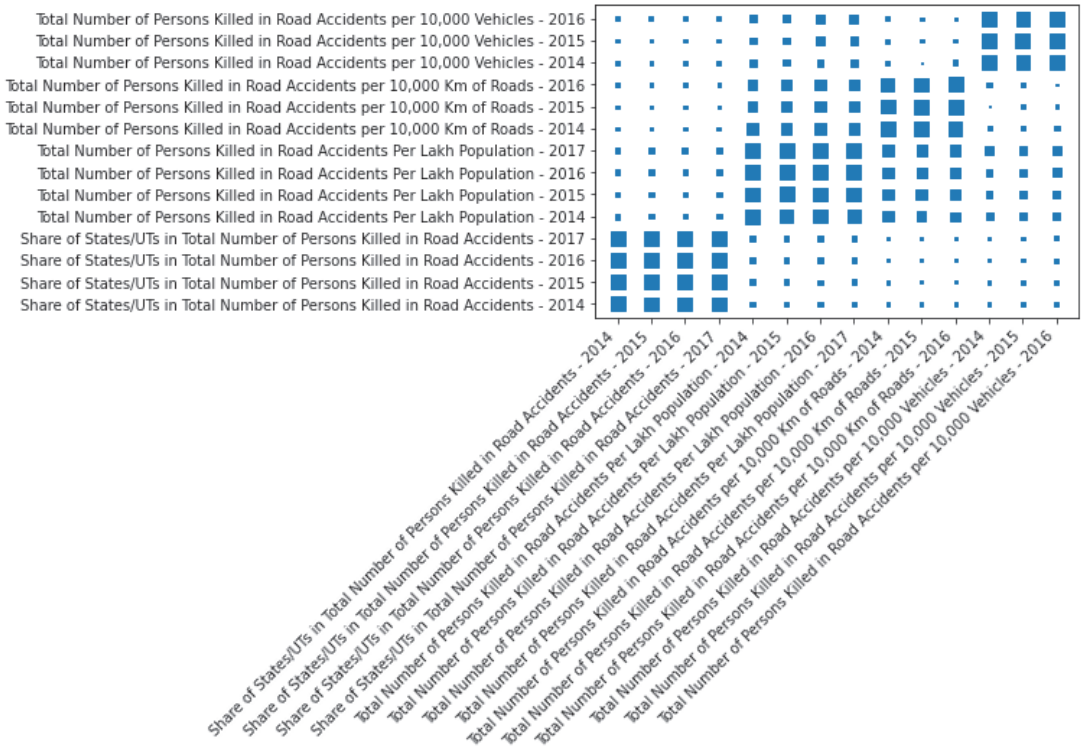
[22]:
```
data2 = pd.read_csv("Road_Accidents_2017-Annuxure_Tables_4.csv")

corr = data2.corr()
ax = sns.heatmap(
    corr,
    vmin=-1, vmax=1, center=0,
    cmap=sns.diverging_palette(20, 220, n=200),
    square=True
)
ax.set_xticklabels(
    ax.get_xticklabels(),
    rotation=45,
    horizontalalignment='right'
);
```



[23]:
```
def heatmap(x, y, size):
    fig, ax = plt.subplots()

    # Mapping from column names to integer coordinates
```

```
    x_labels = [v for v in sorted(x.unique())]
    y_labels = [v for v in sorted(y.unique())]
    x_to_num = {p[1]:p[0] for p in enumerate(x_labels)}
    y_to_num = {p[1]:p[0] for p in enumerate(y_labels)}

    size_scale = 100
    ax.scatter(
        x=x.map(x_to_num), # Use mapping for x
        y=y.map(y_to_num), # Use mapping for y
        s=size * size_scale, # Vector of square sizes, proportional to size␣
↪parameter
        marker='s' # Use square as scatterplot marker
    )

    # Show column labels on the axes
    ax.set_xticks([x_to_num[v] for v in x_labels])
    ax.set_xticklabels(x_labels, rotation=45, horizontalalignment='right')
    ax.set_yticks([y_to_num[v] for v in y_labels])
    ax.set_yticklabels(y_labels)

data = pd.read_csv("Road_Accidents_2017-Annuxure_Tables_3.csv")
columns = ['Share of States/UTs in Total Number of Persons Killed in Road␣
↪Accidents - 2014',
       'Share of States/UTs in Total Number of Persons Killed in Road Accidents␣
↪- 2015',
       'Share of States/UTs in Total Number of Persons Killed in Road Accidents␣
↪- 2016',
       'Share of States/UTs in Total Number of Persons Killed in Road Accidents␣
↪- 2017',
       'Total Number of Persons Killed in Road Accidents Per Lakh Population -␣
↪2014',
       'Total Number of Persons Killed in Road Accidents Per Lakh Population -␣
↪2015',
       'Total Number of Persons Killed in Road Accidents Per Lakh Population -␣
↪2016',
       'Total Number of Persons Killed in Road Accidents Per Lakh Population -␣
↪2017',
       'Total Number of Persons Killed in Road Accidents per 10,000 Vehicles -␣
↪2014',
       'Total Number of Persons Killed in Road Accidents per 10,000 Vehicles -␣
↪2015',
       'Total Number of Persons Killed in Road Accidents per 10,000 Vehicles -␣
↪2016',
       'Total Number of Persons Killed in Road Accidents per 10,000 Km of Roads␣
↪- 2014',
```

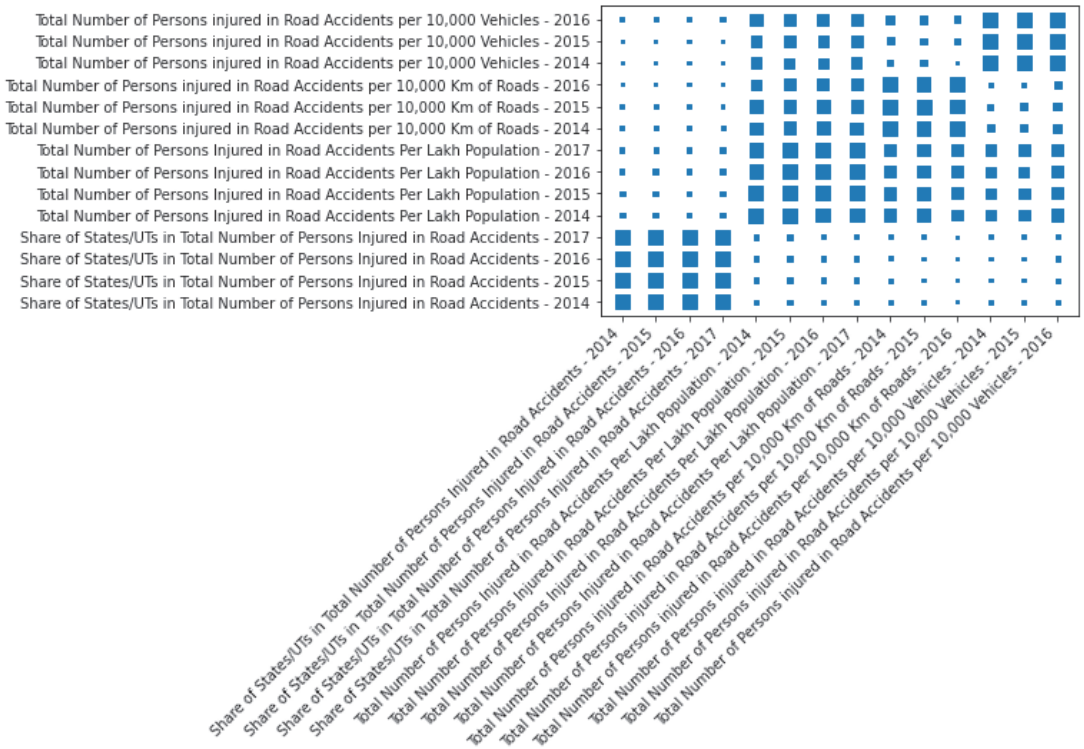
```
      'Total Number of Persons Killed in Road Accidents per 10,000 Km of Roads␣
 ↪- 2015',
      'Total Number of Persons Killed in Road Accidents per 10,000 Km of Roads␣
 ↪- 2016']
corr = data[columns].corr()
corr = pd.melt(corr.reset_index(), id_vars='index') # Unpivot the dataframe, so␣
 ↪we can get pair of arrays for x and y
corr.columns = ['x', 'y', 'value']
heatmap(
    x=corr['x'],
    y=corr['y'],
    size=corr['value'].abs()
)
```



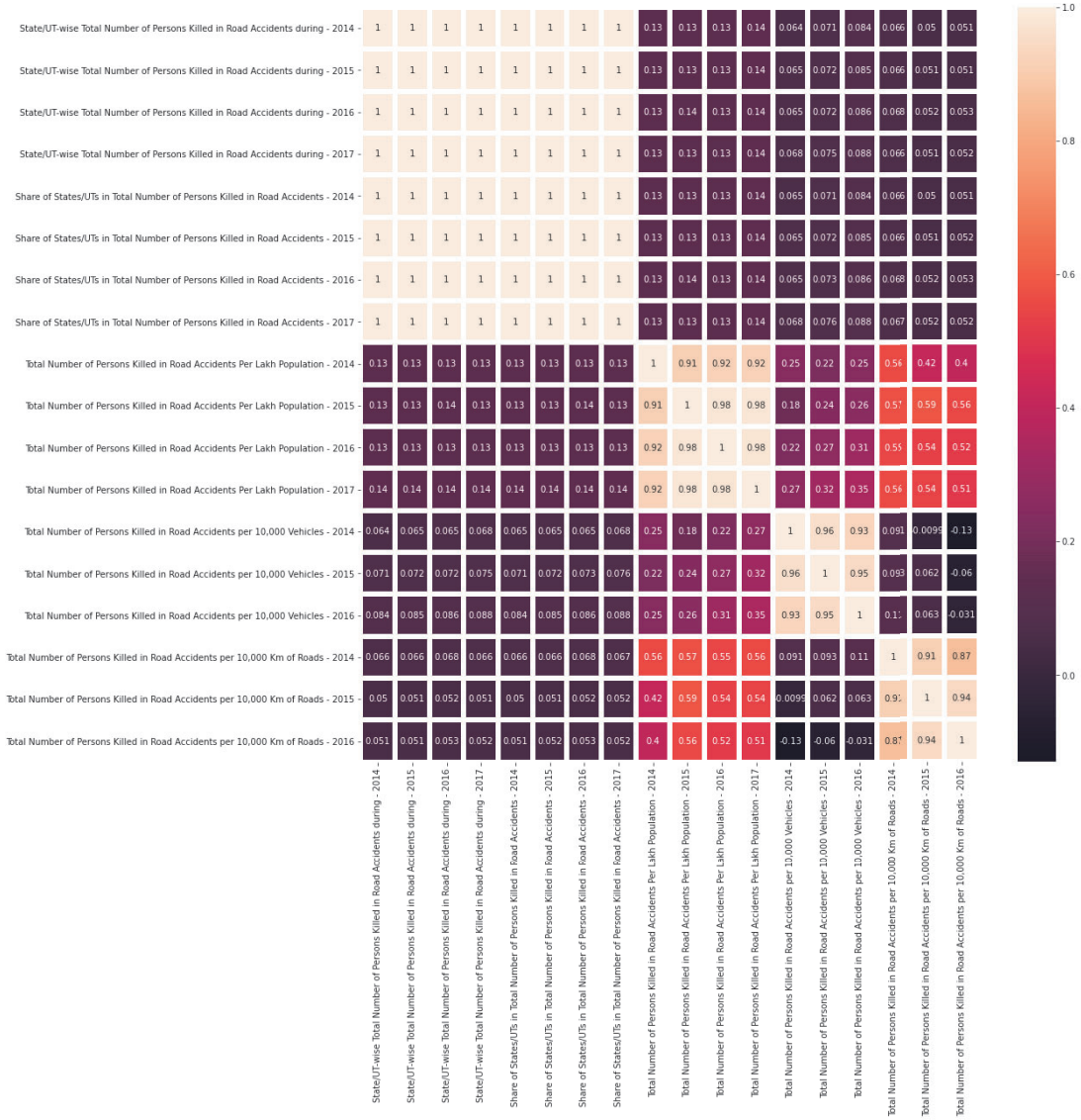
[24]:
```
def heatmap(x, y, size):
    fig, ax = plt.subplots()

    # Mapping from column names to integer coordinates
    x_labels = [v for v in sorted(x.unique())]
    y_labels = [v for v in sorted(y.unique())]
    x_to_num = {p[1]:p[0] for p in enumerate(x_labels)}
    y_to_num = {p[1]:p[0] for p in enumerate(y_labels)}
```

```
    size_scale = 100
    ax.scatter(
        x=x.map(x_to_num), # Use mapping for x
        y=y.map(y_to_num), # Use mapping for y
        s=size * size_scale, # Vector of square sizes, proportional to size␣
↪parameter
        marker='s' # Use square as scatterplot marker
    )

    # Show column labels on the axes
    ax.set_xticks([x_to_num[v] for v in x_labels])
    ax.set_xticklabels(x_labels, rotation=45, horizontalalignment='right')
    ax.set_yticks([y_to_num[v] for v in y_labels])
    ax.set_yticklabels(y_labels)

data = pd.read_csv("Road_Accidents_2017-Annuxure_Tables_4.csv")
columns = ['Share of States/UTs in Total Number of Persons Injured in Road␣
↪Accidents - 2014',
       'Share of States/UTs in Total Number of Persons Injured in Road␣
↪Accidents - 2015',
       'Share of States/UTs in Total Number of Persons Injured in Road␣
↪Accidents - 2016',
       'Share of States/UTs in Total Number of Persons Injured in Road␣
↪Accidents - 2017',
       'Total Number of Persons Injured in Road Accidents Per Lakh Population -␣
↪2014',
       'Total Number of Persons Injured in Road Accidents Per Lakh Population -␣
↪2015',
       'Total Number of Persons Injured in Road Accidents Per Lakh Population -␣
↪2016',
       'Total Number of Persons Injured in Road Accidents Per Lakh Population -␣
↪2017',
       'Total Number of Persons injured in Road Accidents per 10,000 Vehicles -␣
↪2014',
       'Total Number of Persons injured in Road Accidents per 10,000 Vehicles -␣
↪2015',
       'Total Number of Persons injured in Road Accidents per 10,000 Vehicles -␣
↪2016',
       'Total Number of Persons injured in Road Accidents per 10,000 Km of␣
↪Roads - 2014',
       'Total Number of Persons injured in Road Accidents per 10,000 Km of␣
↪Roads - 2015',
       'Total Number of Persons injured in Road Accidents per 10,000 Km of␣
↪Roads - 2016']
corr = data[columns].corr()
```

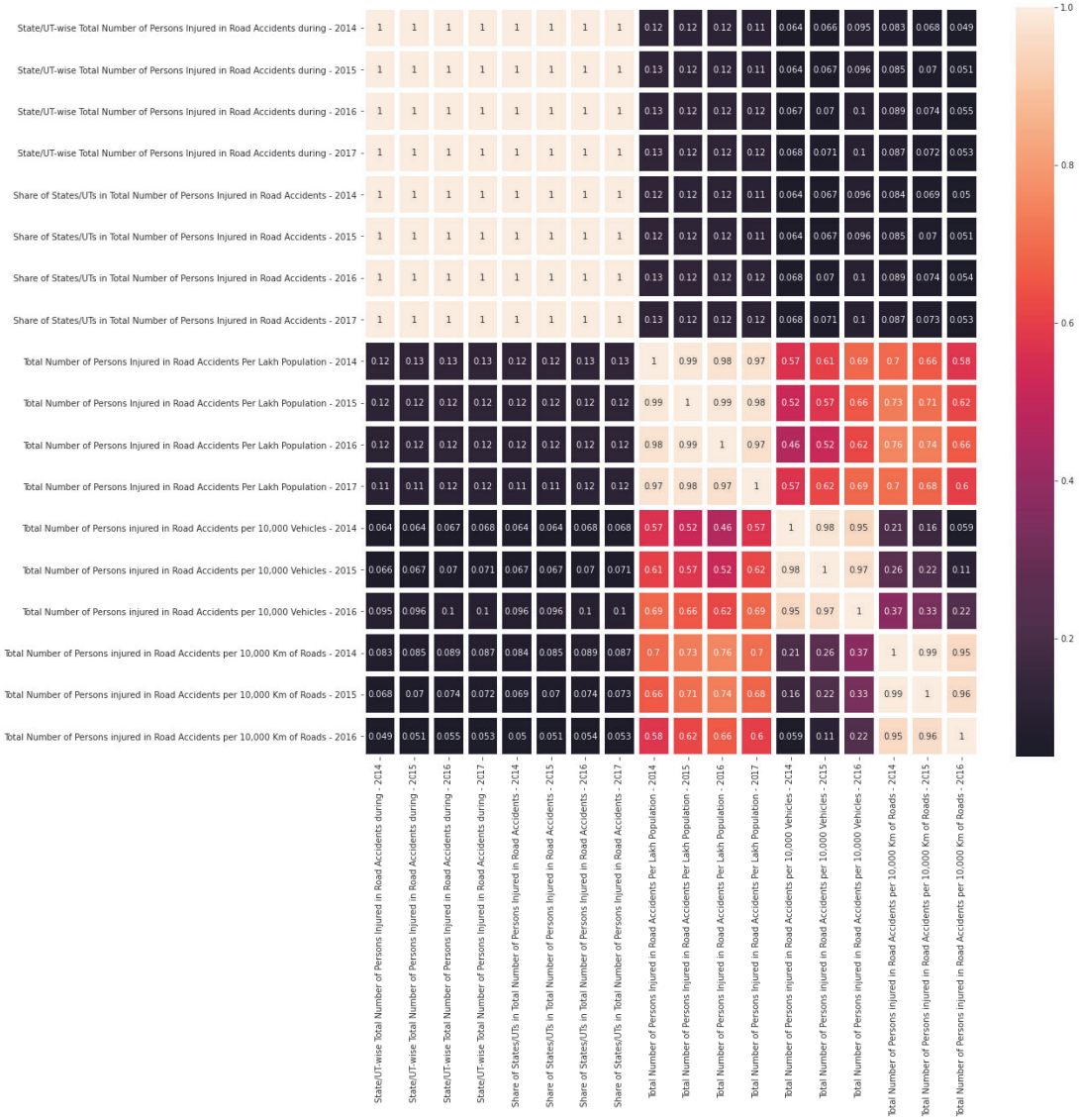
```
corr = pd.melt(corr.reset_index(), id_vars='index') # Unpivot the dataframe, so␣
 ↪we can get pair of arrays for x and y
corr.columns = ['x', 'y', 'value']
heatmap(
    x=corr['x'],
    y=corr['y'],
    size=corr['value'].abs()
)
```



[25]:
```
f,ax = plt.subplots(figsize=(16,16))
sns.heatmap(killed_df.corr(),annot=True,linewidths=5,ax=ax)
plt.show()
```



[26]:
```
f,ax = plt.subplots(figsize=(16,16))
sns.heatmap(injured_df.corr(),annot=True,linewidths=5,ax=ax)
plt.show()
```

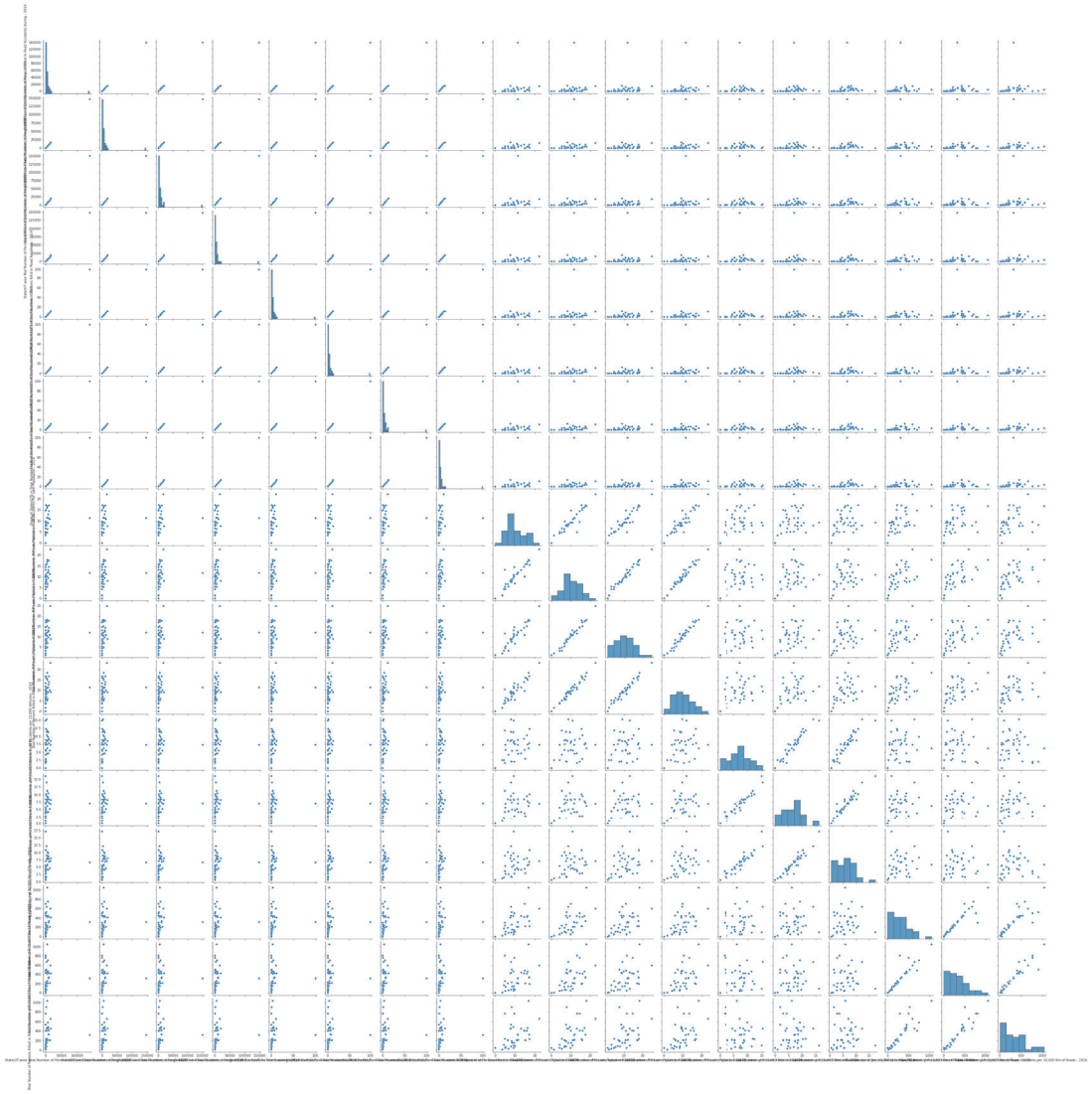


[ ]:

7 Pairplot

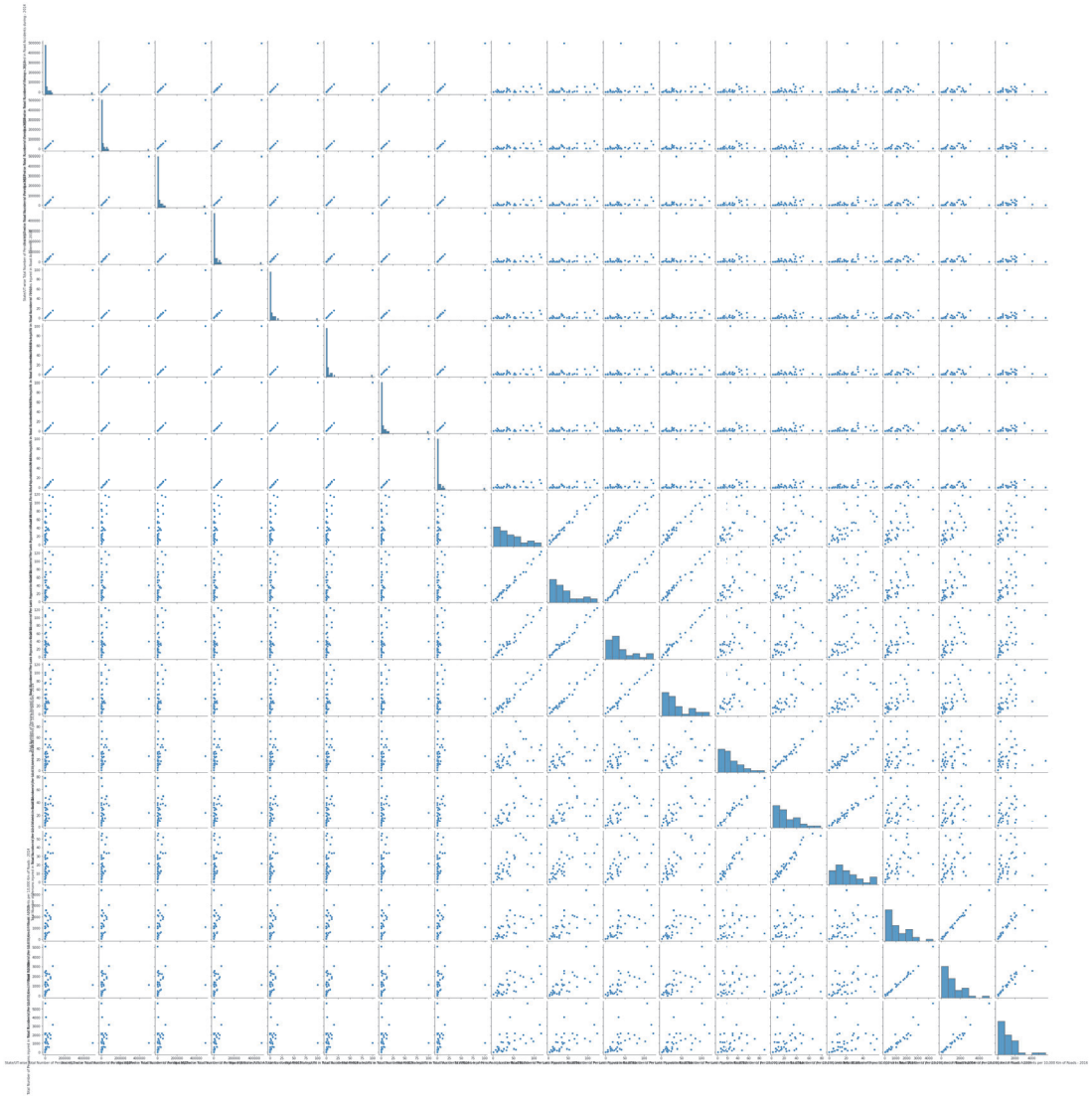
[27]:
```
sns.pairplot(killed_df)
```

[27]: <seaborn.axisgrid.PairGrid at 0x1d000379af0>



[28]:
```
sns.pairplot(injured_df)
```

[28]:
```
<seaborn.axisgrid.PairGrid at 0x1d00f3ad2b0>
```



[ ]:

[ ]:

8 data_cleanup

8.0.1 In data cleaning

- the unwanted columns are dropped
- columns are renamed for easy access
- all states are classified into each zones for easy modelling

[29]:
```
killed_df = killed_df.drop(columns = ['Share of States/UTs in Total Number of␣
↪Persons Killed in Road Accidents - 2014',
      'Share of States/UTs in Total Number of Persons Killed in Road Accidents␣
↪- 2015',
      'Share of States/UTs in Total Number of Persons Killed in Road Accidents␣
↪- 2016',
      'Share of States/UTs in Total Number of Persons Killed in Road Accidents␣
↪- 2017',
      'Total Number of Persons Killed in Road Accidents Per Lakh Population -␣
↪2014',
      'Total Number of Persons Killed in Road Accidents Per Lakh Population -␣
↪2015',
      'Total Number of Persons Killed in Road Accidents Per Lakh Population -␣
↪2016',
      'Total Number of Persons Killed in Road Accidents Per Lakh Population -␣
↪2017',
      'Total Number of Persons Killed in Road Accidents per 10,000 Vehicles -␣
↪2014',
      'Total Number of Persons Killed in Road Accidents per 10,000 Vehicles -␣
↪2015',
      'Total Number of Persons Killed in Road Accidents per 10,000 Vehicles -␣
↪2016',
      'Total Number of Persons Killed in Road Accidents per 10,000 Km of Roads␣
↪- 2014',
      'Total Number of Persons Killed in Road Accidents per 10,000 Km of Roads␣
↪- 2015',
      'Total Number of Persons Killed in Road Accidents per 10,000 Km of Roads␣
↪- 2016'])
injured_df = injured_df.drop(columns = ['Share of States/UTs in Total Number of␣
↪Persons Injured in Road Accidents - 2014',
      'Share of States/UTs in Total Number of Persons Injured in Road␣
↪Accidents - 2015',
      'Share of States/UTs in Total Number of Persons Injured in Road␣
↪Accidents - 2016',
      'Share of States/UTs in Total Number of Persons Injured in Road␣
↪Accidents - 2017',
      'Total Number of Persons Injured in Road Accidents Per Lakh Population -␣
↪2014',
      'Total Number of Persons Injured in Road Accidents Per Lakh Population -␣
↪2015',
      'Total Number of Persons Injured in Road Accidents Per Lakh Population -␣
↪2016',
      'Total Number of Persons Injured in Road Accidents Per Lakh Population -␣
↪2017',
```

```
      'Total Number of Persons injured in Road Accidents per 10,000 Vehicles -␣
 ↪2014',
      'Total Number of Persons injured in Road Accidents per 10,000 Vehicles -␣
 ↪2015',
      'Total Number of Persons injured in Road Accidents per 10,000 Vehicles -␣
 ↪2016',
      'Total Number of Persons injured in Road Accidents per 10,000 Km of␣
 ↪Roads - 2014',
      'Total Number of Persons injured in Road Accidents per 10,000 Km of␣
 ↪Roads - 2015',
      'Total Number of Persons injured in Road Accidents per 10,000 Km of␣
 ↪Roads - 2016'])
```

[30]:
```
killed_df = killed_df.rename(columns = {'State/UT-wise Total Number of Persons␣
 ↪Killed in Road Accidents during - 2014':2014,
                                        'State/UT-wise Total Number of Persons␣
 ↪Killed in Road Accidents during - 2015':2015,
                                        'State/UT-wise Total Number of Persons␣
 ↪Killed in Road Accidents during - 2016':2016,
                                        'State/UT-wise Total Number of Persons␣
 ↪Killed in Road Accidents during - 2017':2017})
injured_df = injured_df.rename(columns = {'State/UT-wise Total Number of␣
 ↪Persons Injured in Road Accidents during - 2014':2014,
                                        'State/UT-wise Total Number of Persons␣
 ↪Injured in Road Accidents during - 2015':2015,
                                        'State/UT-wise Total Number of Persons␣
 ↪Injured in Road Accidents during - 2016':2016,
                                        'State/UT-wise Total Number of Persons␣
 ↪Injured in Road Accidents during - 2017':2017})
roadcond_df =  roadcond_df.rename(columns = {'State/ UT':'States/UTs'})
weather_df =  weather_df.rename(columns = {'State/ UT':'States/UTs'})
```

[ ]:

[ ]:

9 Outlier Analysis : Box Plot

9.0.1 killed

[31]:
```
df1=killed_df
trace1 = go.Box(
y = df1[2014],
name = "State/UT-wise Total Number of Persons Killed in Road Accidents during -␣
 ↪2014",
marker = dict(
```

```
color = "rgb(255,0,255)",
)
)
trace2 = go.Box(
y = df1[2015],
name = "State/UT-wise Total Number of Persons Killed in Road Accidents during -␣
 ↪2015",
marker = dict(
color = "rgb(255,0,255)",
)
)
trace3 = go.Box(
y = df1[2016],
name = "State/UT-wise Total Number of Persons Killed in Road Accidents during -␣
 ↪2016",
marker = dict(
color = "rgb(255,0,255)",
)
)
trace4 = go.Box(
y = df1[2017],
name = "State/UT-wise Total Number of Persons Killed in Road Accidents during -␣
 ↪2017",
marker = dict(
color = "rgb(255,0,255)",
)
)

data = [trace1,trace2,trace3,trace4]

iplot(data)
```

9.0.2 injured

[32]:
```
df1=injured_df
trace1 = go.Box(
y = df1[2014],
name = "State/UT-wise Total Number of Persons Injured in Road Accidents during␣
 ↪- 2014",
marker = dict(
color = "rgb(255,0,255)",
)
)
trace2 = go.Box(
y = df1[2015],
name = "State/UT-wise Total Number of Persons Injured in Road Accidents during␣
 ↪- 2015",
```

```
marker = dict(
color = "rgb(255,0,255)",
)
)
trace3 = go.Box(
y = df1[2016],
name = "State/UT-wise Total Number of Persons Injured in Road Accidents during␣
 ↪- 2016",
marker = dict(
color = "rgb(255,0,255)",
)
)
trace4 = go.Box(
y = df1[2017],
name = "State/UT-wise Total Number of Persons Injured in Road Accidents during␣
 ↪- 2017",
marker = dict(
color = "rgb(255,0,255)",
)
)

data = [trace1,trace2,trace3,trace4]

iplot(data)
```

[ ]:

[ ]:

[ ]:

9.0.3 Dividing States into Zones and Adding a Column

[33]:
```
north_india = ['Jammu & Kashmir', 'Punjab', 'Himachal Pradesh', 'Haryana',␣
 ↪'Uttarakhand', 'Uttar Pradesh', 'Chandigarh', 'Jammu and Kashmir', 'Delhi']
east_india = ['Bihar', 'Odisha', 'Jharkhand', 'West Bengal', 'Orissa']
south_india = ['Andhra Pradesh', 'Karnataka', 'Kerala' ,'Tamil Nadu',␣
 ↪'Telangana']
west_india = ['Rajasthan' , 'Gujarat', 'Goa','Maharashtra','Goa']
central_india = ['Madhya Pradesh', 'Chhattisgarh']
north_east_india = ['Assam', 'Sikkim', 'Nagaland', 'Meghalaya', 'Manipur',␣
 ↪'Mizoram', 'Tripura', 'Arunachal Pradesh']
ut_india = ['Andaman and Nicobar Islands', 'Dadra and Nagar Haveli',␣
 ↪'Puducherry', 'Andaman & Nicobar Islands', 'Dadra & Nagar Haveli', 'Daman &␣
 ↪Diu', 'Lakshadweep', 'A & N Islands', 'D & N Haveli']
```

[34]:
```
def get_zonal_names(row):
    if row['States/UTs'].strip() in north_india:
        val = 'North Zone'
    elif row['States/UTs'].strip()  in south_india:
        val = 'South Zone'
    elif row['States/UTs'].strip()  in east_india:
        val = 'East Zone'
    elif row['States/UTs'].strip()  in west_india:
        val = 'West Zone'
    elif row['States/UTs'].strip()  in central_india:
        val = 'Central Zone'
    elif row['States/UTs'].strip()  in north_east_india:
        val = 'NE Zone'
    elif row['States/UTs'].strip()  in ut_india:
        val = 'Union Terr'
    else:
        val = 'No Value'
    return val
```

[35]:
```
killed_df.drop(killed_df[killed_df['States/UTs'] == 'Total'].index,␣
↪inplace=True)
killed_df['Zones'] = killed_df.apply(get_zonal_names, axis=1)
injured_df.drop(injured_df[injured_df['States/UTs'] == 'Total'].index,␣
↪inplace=True)
injured_df['Zones'] = injured_df.apply(get_zonal_names, axis=1)
roadcond_df.drop(roadcond_df[roadcond_df['States/UTs'] == 'Total'].index,␣
↪inplace=True)
roadcond_df['Zones'] = roadcond_df.apply(get_zonal_names, axis=1)
weather_df.drop(weather_df[weather_df['States/UTs'] == 'Total'].index,␣
↪inplace=True)
weather_df['Zones'] = weather_df.apply(get_zonal_names, axis=1)
```

9.0.4 Separating Features in the Weather and Road Condition Dataframe

[36]:
```
total_col = [col for col in weather_df.columns if 'Total' in col]
killed_col = [col for col in weather_df.columns if 'Killed' in col]
injured_col = [col for col in weather_df.columns if 'Injured' in col]
weather_df_killed = weather_df.drop(columns = total_col+injured_col)
weather_df_injured = weather_df.drop(columns = total_col+killed_col)
```

[37]:
```
total_col = [col for col in roadcond_df.columns if 'Accident' in col]
killed_col = [col for col in roadcond_df.columns if 'Killed' in col]
injured_col = [col for col in roadcond_df.columns if 'Injured' in col]
roadcond_df_killed = roadcond_df.drop(columns = total_col+injured_col)
roadcond_df_injured = roadcond_df.drop(columns = total_col+killed_col)
```

[ ]:
```

```

9.1 Calculating the mean of State wise Total number of Persons Killed in Road Accidents.

[38]:
```
mean13 = np.mean(killed_df[2014])
print("Mean of accidents happened in all states in year 2014: {}".
 ↪format(mean13))
```

```
Mean of accidents happened in all states in year 2014: 3879.75
```

[39]:
```
mean14 = np.mean(killed_df[2015])
print("Mean of accidents happened in all states in year 2015 : {}".
 ↪format(mean14))
```

```
Mean of accidents happened in all states in year 2015 : 4059.25
```

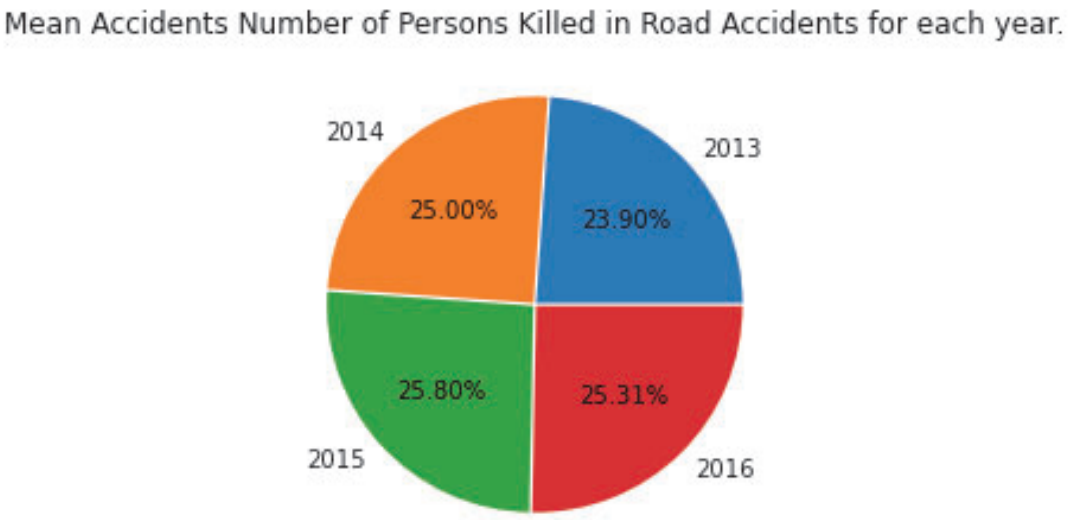
[40]:
```
mean15 = np.mean(killed_df[2016])
print("Mean of accidents happened in all states in year 2016 : {}".
 ↪format(mean15))
```

```
Mean of accidents happened in all states in year 2016 : 4188.472222222223
```

[41]:
```
mean16 = np.mean(killed_df[2017])
print("Mean of accidents happened in all states in 2017 {}".format(mean16))
```

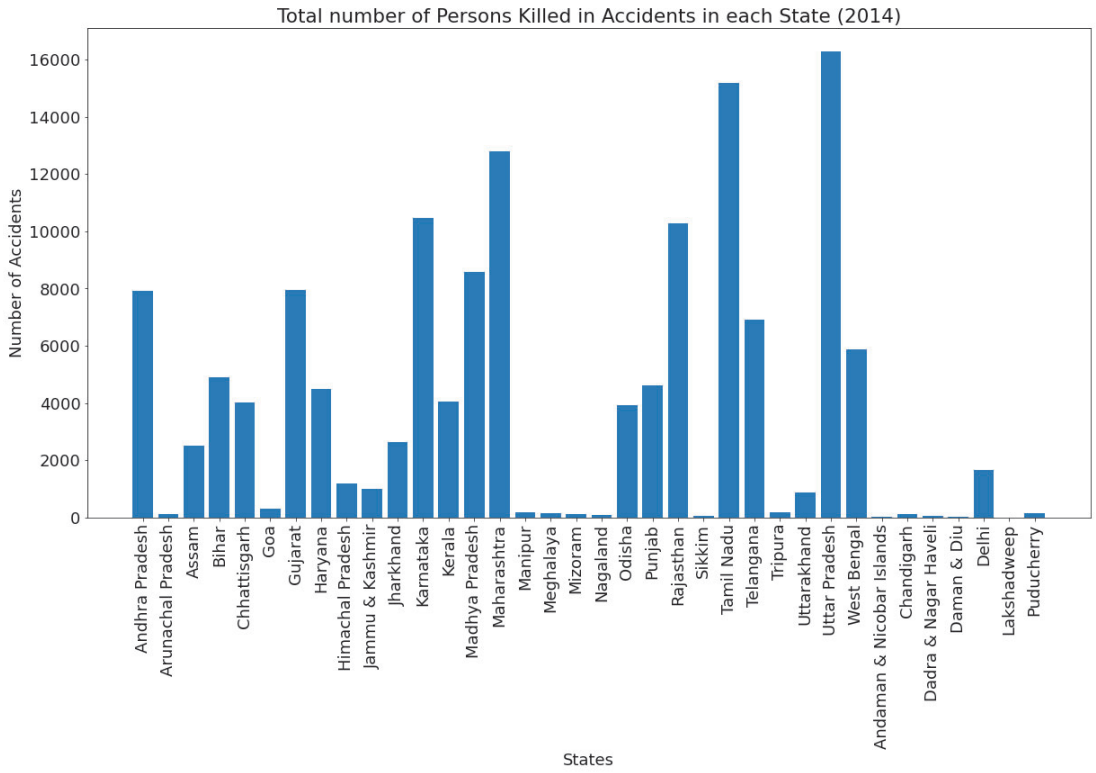
```
Mean of accidents happened in all states in 2017 4108.694444444444
```

[42]:
```
labels = ('2013','2014','2015','2016')
explode = (0.01,0.01,0.01,0.01)
performance = [mean13, mean14, mean15, mean16]
plt.pie(performance, labels = labels, autopct = '%.2f%%', center = (0,0),␣
 ↪explode = explode )
plt.title("Mean Accidents Number of Persons Killed in Road Accidents for each␣
 ↪year.")
plt.show()
```



9.1.1 It is clearly visible that, the percentage of road accidents are almost constant during all years. It indicates that the government is making less efforts to prevent accidents by creating wider, good quality roads or creating new safety rules.

[43]:
```
plt.figure(figsize = (20,10))
plt.rcParams.update({'font.size':18})
y = killed_df[2014]
yd = killed_df['States/UTs']
p =killed_df['States/UTs'].nunique()
d = np.linspace(1,p,p)    # refer notes
plt.bar(d, y, align = 'center')
plt.xticks(d, yd, rotation = 90)
plt.xlabel('States')
plt.ylabel('Number of Accidents')
plt.title('Total number of Persons Killed in Accidents in each State (2014)')
plt.show()
```



Maharashtra and Tamil Nadu have the highest number of accidents. Further investigation needs to be done to understand the case. Arunachal, Manipur, Meghalaya, Mizoram, Nagaland, Tripura have the least number of accidents. They surprisingly all belong to the north-eastern area.

9.1.2 Calculating the mean of State wise Total number of Persons injured in Road Accidents.

[ ]:

[44]:
```
acc13 = np.mean(injured_df[2014])
print("Mean of accidents per lakh population in 2014 : {}".format(acc13))
```

```
Mean of accidents per lakh population in 2014 : 13707.611111111111
```

[45]:
```
acc14 = np.mean(injured_df[2015])
print("Mean of accidents per lakh population in 2015 : {}".format(acc14))
```

```
Mean of accidents per lakh population in 2015 : 13896.638888888889
```

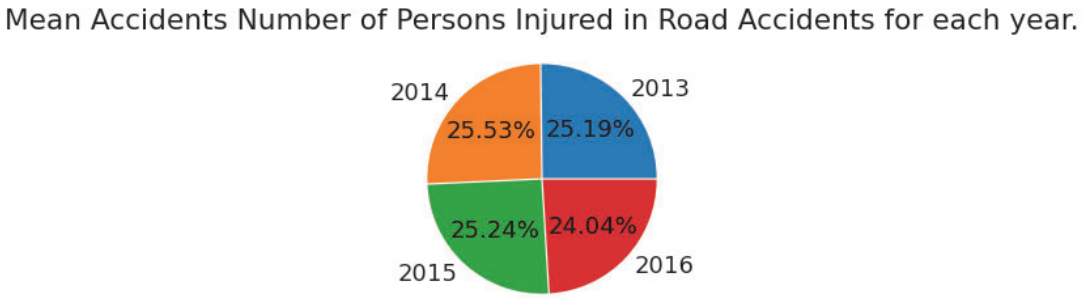
[46]:
```
acc15 = np.mean(injured_df[2016])
print("Mean of accidents per lakh population in 2016 : {}".format(acc15))
```

```
Mean of accidents per lakh population in 2016 : 13739.555555555555
```

[47]:
```
acc16 = np.mean(injured_df[2017])
print("Mean of accidents per lakh population in 2017 : {}".format(acc16))
```

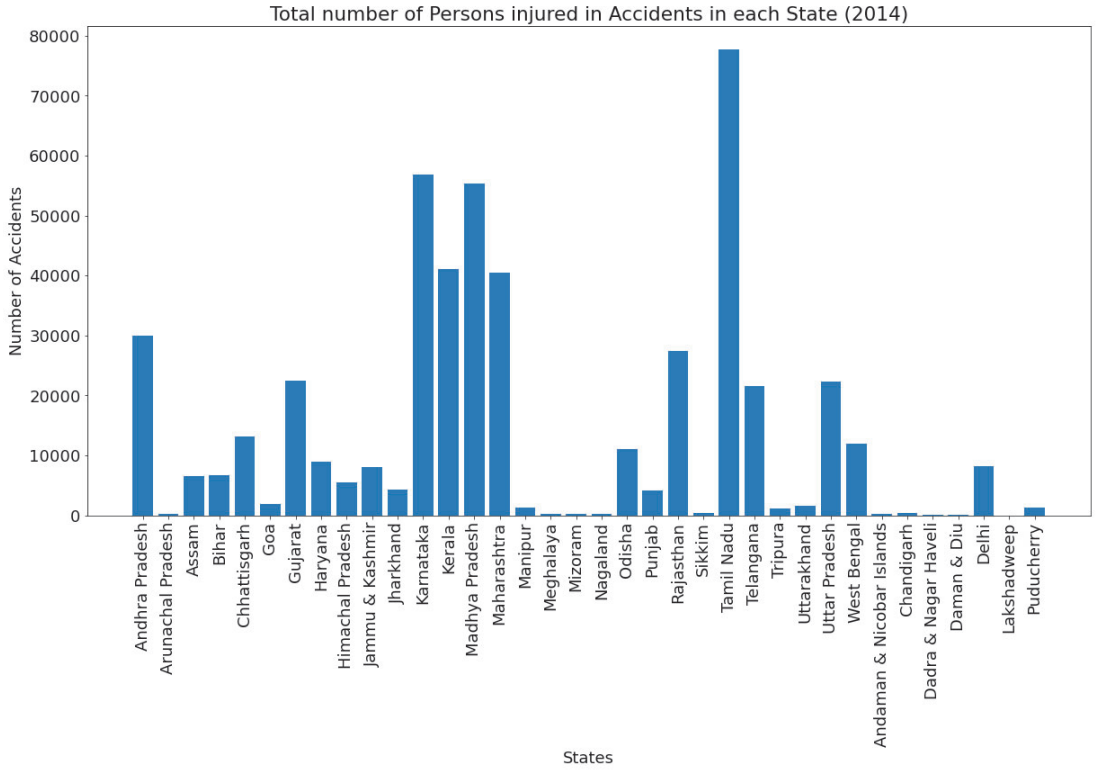
```
Mean of accidents per lakh population in 2017 : 13082.638888888889
```

[48]:
```
labels = ('2013','2014','2015','2016')
explode = (0.01,0.01,0.01,0.01)
performance = [acc13, acc14, acc15, acc16]
plt.pie(performance, labels = labels, autopct = '%.2f%%', center = (0,0),␣
↪explode = explode )
plt.title("Mean Accidents Number of Persons Injured in Road Accidents for each␣
↪year.")
plt.show()
```



9.2 A similar rate is obtained as previous

[49]:
```
plt.figure(figsize = (20,10))
plt.rcParams.update({'font.size':18})
y = injured_df[2014]
yd = injured_df['States/UTs']
p =injured_df['States/UTs'].nunique()
d = np.linspace(1,p,p)    # refer notes
plt.bar(d, y, align = 'center')
plt.xticks(d, yd, rotation = 90)
plt.xlabel('States')
plt.ylabel('Number of Accidents')
plt.title('Total number of Persons injured in Accidents in each State (2014)')
plt.show()
```



Maharashtra and Tamil Nadu have the highest number of accidents. Further investigation needs to be done to understand the case. Arunachal, Manipur, Meghalaya, Mizoram, Nagaland, Tripura have the least number of accidents. They surprisingly all belong to the **north-eastern area.**

Also,

It is clearly visible that, the percentage of road accidents are almost constant during all years. It indicates that the government is making less efforts to prevent accidents by creating wider, good quality roads or creating new safety rules.

[ ]:

9.2.1 Min_Max Scores for Total Number of Persons Killed in Road Accidents Per Lakh Population

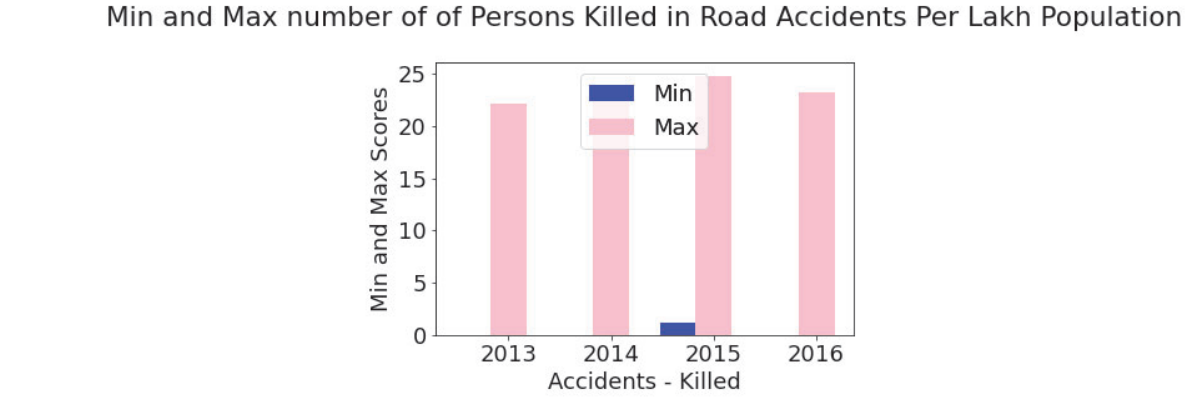
[50]:
```
min13 = np.min(data1['Total Number of Persons Killed in Road Accidents Per Lakh␣
 ↪Population - 2014'])
max13 = np.max(data1['Total Number of Persons Killed in Road Accidents Per Lakh␣
 ↪Population - 2014'])
```

[51]:
```
min14 = np.min(data1['Total Number of Persons Killed in Road Accidents Per Lakh␣
 ↪Population - 2015'])
max14 = np.max(data1['Total Number of Persons Killed in Road Accidents Per Lakh␣
 ↪Population - 2015'])
```

[52]:
```
min15 = np.min(data1['Total Number of Persons Killed in Road Accidents Per Lakh␣
 ↪Population - 2016'])
max15 = np.max(data1['Total Number of Persons Killed in Road Accidents Per Lakh␣
 ↪Population - 2016'])
```

[53]:
```
min16 = np.min(data1['Total Number of Persons Killed in Road Accidents Per Lakh␣
 ↪Population - 2017'])
max16 = np.max(data1['Total Number of Persons Killed in Road Accidents Per Lakh␣
 ↪Population - 2017'])
```

[54]:
```
n = 4
minx = (min13, min14, min15, min16)
maxx = (max13, max14, max15, max16)
index = np.arange(n)
bar_width = 0.35

r1 = plt.bar(index, minx, bar_width, align = 'center', color = 'b', label =␣
 ↪'Min')

r2 = plt.bar(index + bar_width, maxx, bar_width, align = 'center', color␣
 ↪='pink', label = 'Max')

plt.xlabel("Accidents - Killed")
plt.ylabel("Min and Max Scores")
plt.title("Min and Max number of of Persons Killed in Road Accidents Per Lakh␣
 ↪Population\n")
plt.xticks(index + bar_width,('2013','2014','2015','2016'))
plt.legend(loc = 'upper center')
plt.show()
```



[ ]:

9.2.2 Min_Max Scores for Total Number of Persons Injured in Road Accidents Per Lakh Population

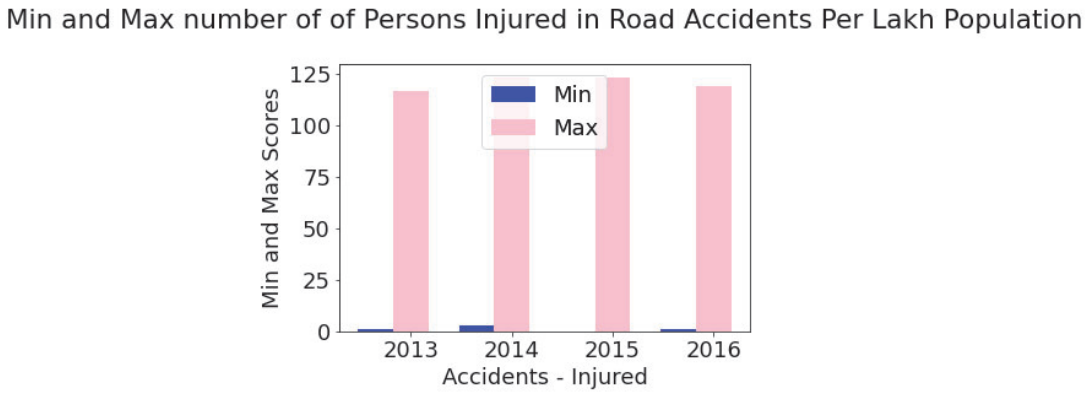
[55]:
```
min23 = np.min(data2['Total Number of Persons Injured in Road Accidents Per␣
 ↪Lakh Population - 2014'])
max23 = np.max(data2['Total Number of Persons Injured in Road Accidents Per␣
 ↪Lakh Population - 2014'])
```

[56]:
```
min24 = np.min(data2['Total Number of Persons Injured in Road Accidents Per␣
 ↪Lakh Population - 2015'])
max24 = np.max(data2['Total Number of Persons Injured in Road Accidents Per␣
 ↪Lakh Population - 2015'])
```

[57]:
```
min25 = np.min(data2['Total Number of Persons Injured in Road Accidents Per␣
 ↪Lakh Population - 2016'])
max25 = np.max(data2['Total Number of Persons Injured in Road Accidents Per␣
 ↪Lakh Population - 2016'])
```

[58]:
```
min26 = np.min(data2['Total Number of Persons Injured in Road Accidents Per␣
 ↪Lakh Population - 2017'])
max26 = np.max(data2['Total Number of Persons Injured in Road Accidents Per␣
 ↪Lakh Population - 2017'])
```

[59]:
```
n = 4
minx = (min23, min24, min25, min26)
maxx = (max23, max24, max25, max26)
index = np.arange(n)
bar_width = 0.35

r1 = plt.bar(index, minx, bar_width, align = 'center', color = 'b', label =␣
 ↪'Min')

r2 = plt.bar(index + bar_width, maxx, bar_width, align = 'center', color␣
 ↪='pink', label = 'Max')

plt.xlabel("Accidents - Injured")
plt.ylabel("Min and Max Scores")
plt.title("Min and Max number of of Persons Injured in Road Accidents Per Lakh␣
 ↪Population\n")
plt.xticks(index + bar_width,('2013','2014','2015','2016'))
plt.legend(loc = 'upper center')
plt.show()
```



[ ]:

[ ]:

[ ]:

[ ]:

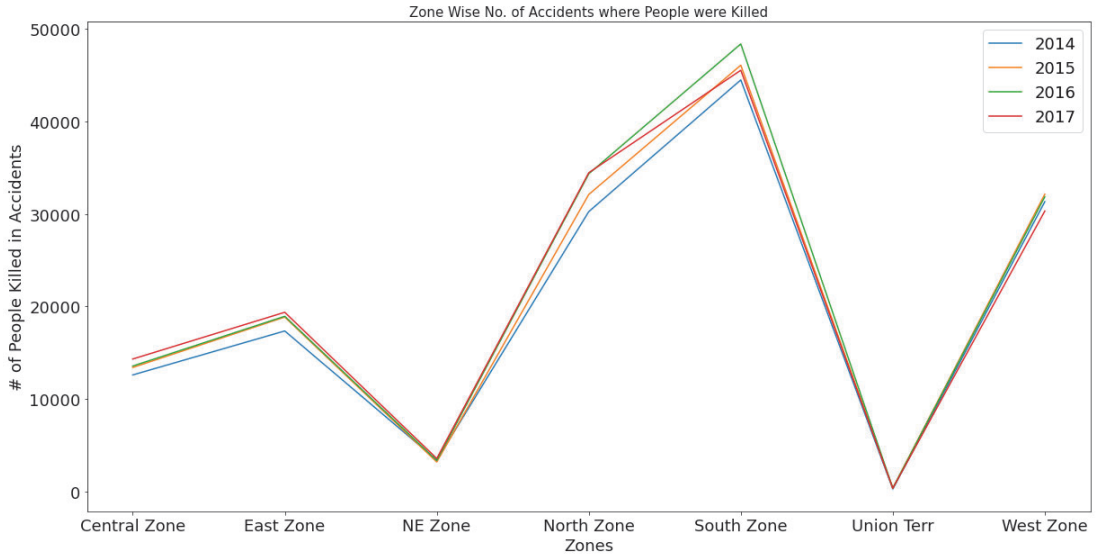
[ ]:

[ ]:

[ ]:

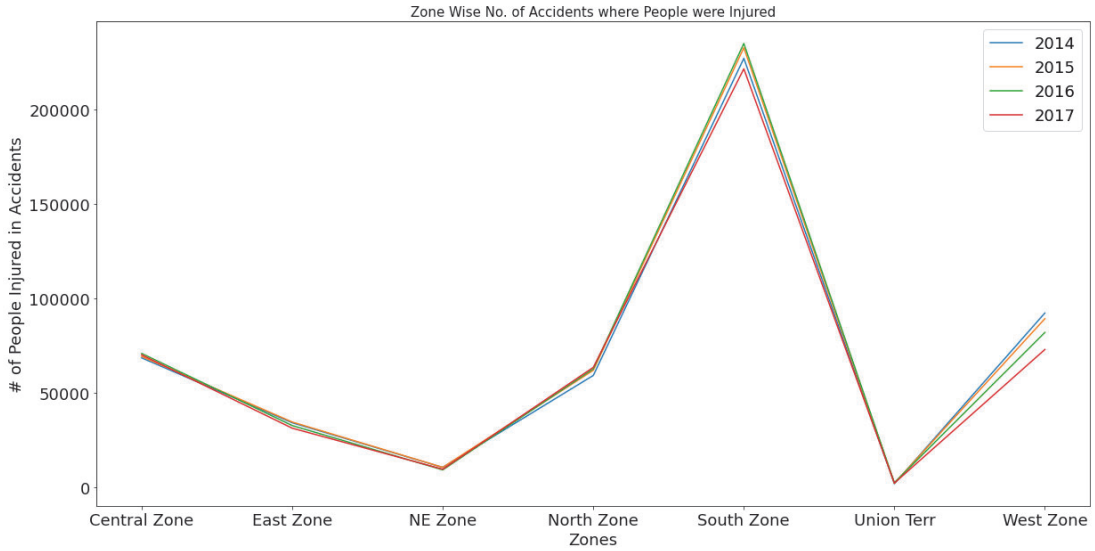
10 Zone Wise No. of Accidents where People were Killed

[60]:
```
df =pd.pivot_table(killed_df, index=['Zones'],values=[2014, 2015, 2016,␣
↪2017],aggfunc=np.sum).reset_index()
df

years = [2014,2015,2016,2017]
fig,ax = plt.subplots(1,1, figsize=(20,10))
for i,year in enumerate(years):
    sns.lineplot(x=df['Zones'],y=df[year], label=years[i])
    plt.ylabel('# of People Killed in Accidents')
    plt.title('Zone Wise No. of Accidents where People were Killed',␣
↪fontsize=15)
```



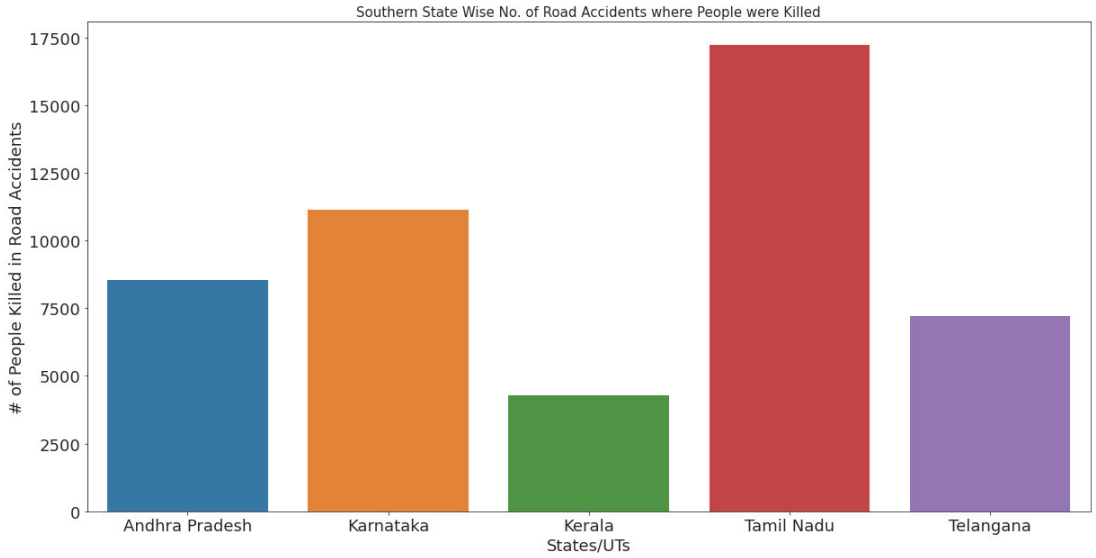
11 Zone Wise No. of Accidents where People were Injured

[61]:
```
df =pd.pivot_table(injured_df, index=['Zones'],values=[2014, 2015, 2016,␣
↪2017],aggfunc=np.sum).reset_index()
df

years = [2014,2015,2016,2017]
fig,ax = plt.subplots(1,1, figsize=(20,10))
for i,year in enumerate(years):
    sns.lineplot(x=df['Zones'],y=df[year], label=years[i])
    plt.ylabel('# of People Injured in Accidents')
    plt.title('Zone Wise No. of Accidents where People were Injured',␣
↪fontsize=15)
```



12 Southern-Zone States: No. of Road Accidents where People were Killed

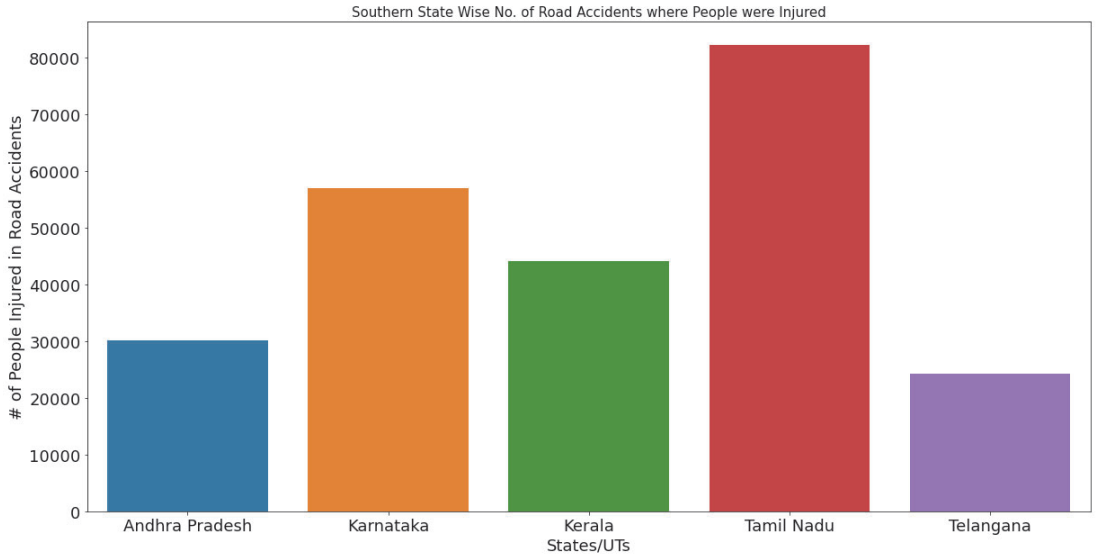
[62]:
```
sub_df = killed_df[killed_df['Zones'] == 'South Zone']
df =pd.pivot_table(sub_df, index=['States/UTs'],values=[2014, 2015, 2016,␣
↪2017],aggfunc=np.sum).reset_index()
df

years = [2014,2015,2016,2017]
fig,ax = plt.subplots(1,1, figsize=(20,10))
for i,year in enumerate(years):
    sns.barplot(x=df['States/UTs'],y=df[year])
    plt.ylabel('# of People Killed in Road Accidents')
    plt.title('Southern State Wise No. of Road Accidents where People were␣
↪Killed', fontsize=15)
```



12.0.1 We see that more deaths due to Road accidents happened in Tamil Nadu.

13 Southern-Zone States: No. of Road Accidents where People were Injured

[63]:
```
sub_df = injured_df[injured_df['Zones'] == 'South Zone']
df =pd.pivot_table(sub_df, index=['States/UTs'],values=[2014, 2015, 2016,␣
 ↪2017],aggfunc=np.sum).reset_index()
df

years = [2014,2015,2016,2017]
fig,ax = plt.subplots(1,1, figsize=(20,10))
for i,year in enumerate(years):
    sns.barplot(x=df['States/UTs'],y=df[year])
    plt.ylabel('# of People Injured in Road Accidents')
    plt.title('Southern State Wise No. of Road Accidents where People were␣
 ↪Injured', fontsize=15)
```



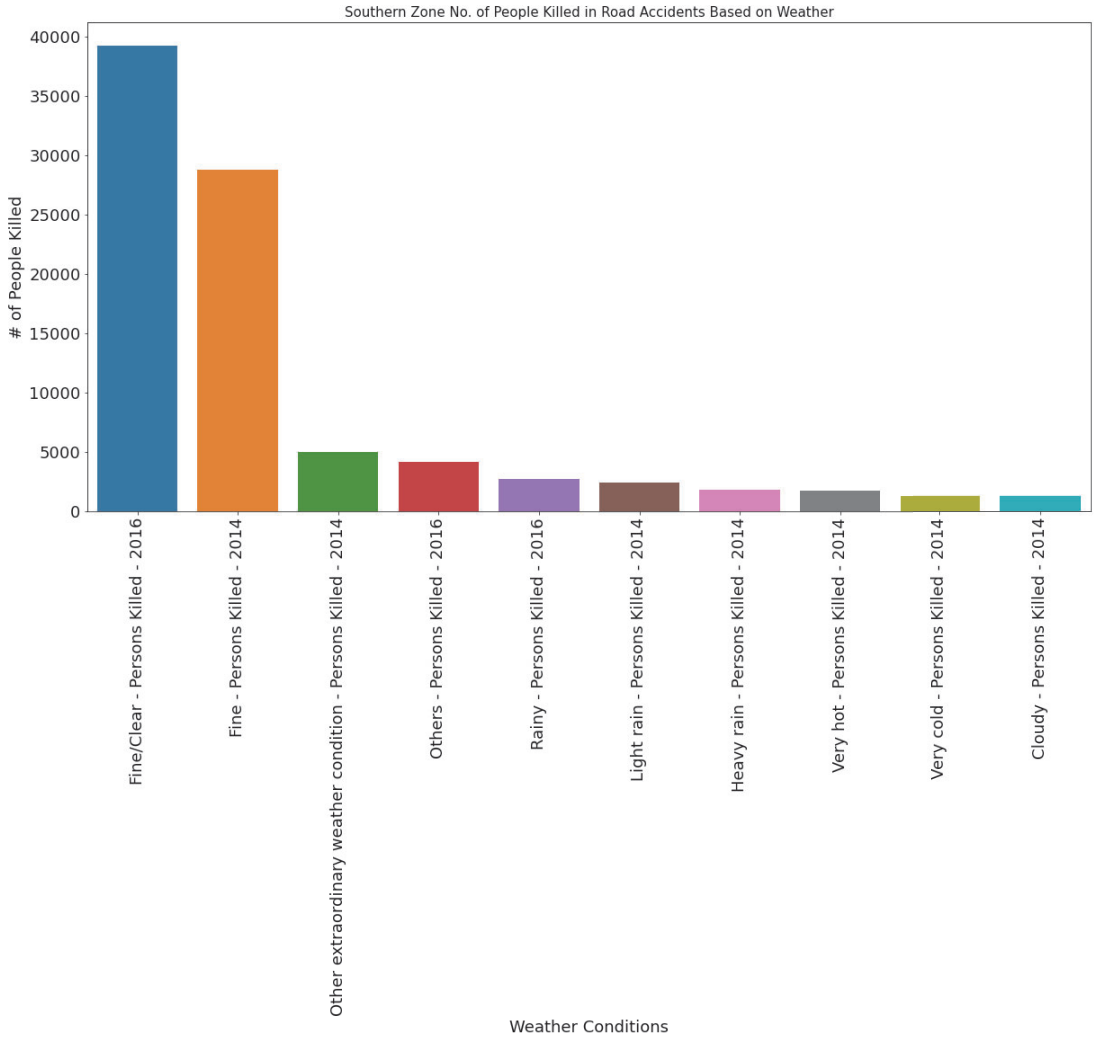
13.0.1 Tamil Nadu reports the highest number of people killed and injured in road accidents, followed by Karnataka. At this point, let us use the other data files we have to see what probable causes.

14 Weather Conditions - No. of People Killed in Road Accidents (South Zone)

[64]:
```
sub_df = weather_df_killed[weather_df_killed['Zones'] == 'South Zone']
df =pd.pivot_table(sub_df, index=['Zones'],aggfunc=np.sum).reset_index()
df = df.T.reset_index()
df = df.rename(columns = {'index': 'Weather Conditions', 0: 'Total'})
df = df.drop(df.index[0])
df = df.sort_values(by = ['Total'], ascending=False).head(10)
df

fig,ax = plt.subplots(1,1, figsize=(20,10))
sns.barplot(x=df['Weather Conditions'],y=df['Total'])
plt.ylabel('# of People Killed')
plt.title('Southern Zone No. of People Killed in Road Accidents Based on␣
 ↪Weather', fontsize=15)
plt.xticks(rotation=90)
```

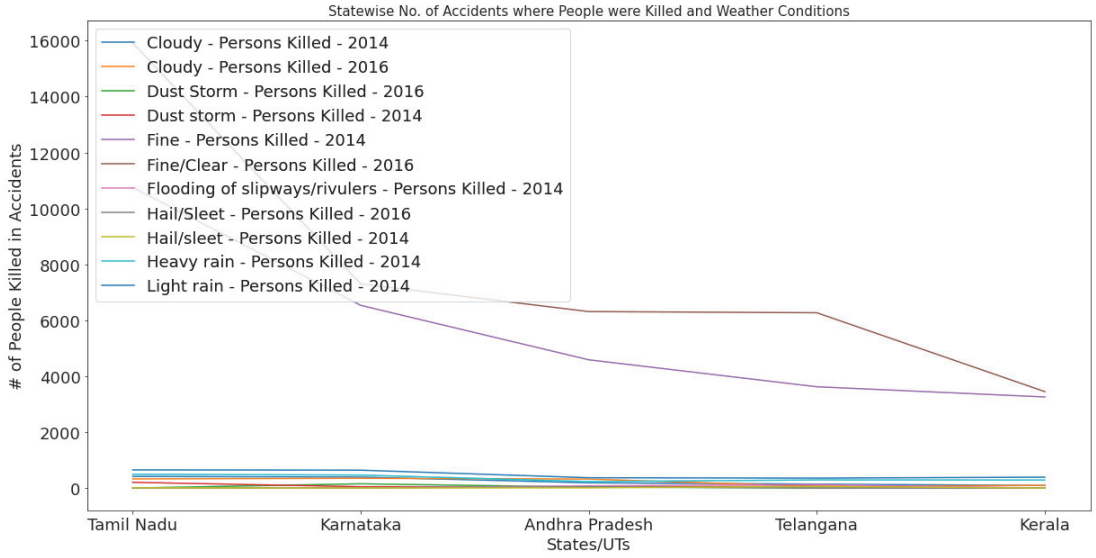
[64]:
```
(array([0, 1, 2, 3, 4, 5, 6, 7, 8, 9]),
 [Text(0, 0, 'Fine/Clear - Persons Killed - 2016'),
  Text(1, 0, 'Fine - Persons Killed - 2014'),
  Text(2, 0, 'Other extraordinary weather condition - Persons Killed - 2014'),
  Text(3, 0, 'Others - Persons Killed - 2016'),
```

```
Text(4, 0, 'Rainy - Persons Killed - 2016'),
Text(5, 0, 'Light rain - Persons Killed - 2014'),
Text(6, 0, 'Heavy rain - Persons Killed - 2014'),
Text(7, 0, 'Very hot - Persons Killed - 2014'),
Text(8, 0, 'Very cold - Persons Killed - 2014'),
Text(9, 0, 'Cloudy - Persons Killed - 2014')])
```



[65]:
```
sub_df = weather_df_killed[weather_df_killed['Zones'] == 'South Zone']
df =pd.pivot_table(sub_df, index=['States/UTs'],aggfunc=np.sum).reset_index()
df = df.reset_index()
cols = list(df.columns[2:13])
df = df.sort_values(by=cols, ascending=False).head(5)

fig,ax = plt.subplots(1,1, figsize=(20,10))
```

```
for i,reason in enumerate(cols):
    sns.lineplot(x=df['States/UTs'],y=df[reason], label=cols[i])
    plt.ylabel('# of People Killed in Accidents')
    plt.title('Statewise No. of Accidents where People were Killed and Weather␣
 ↪Conditions', fontsize=15)
    plt.legend(loc='upper left')
```



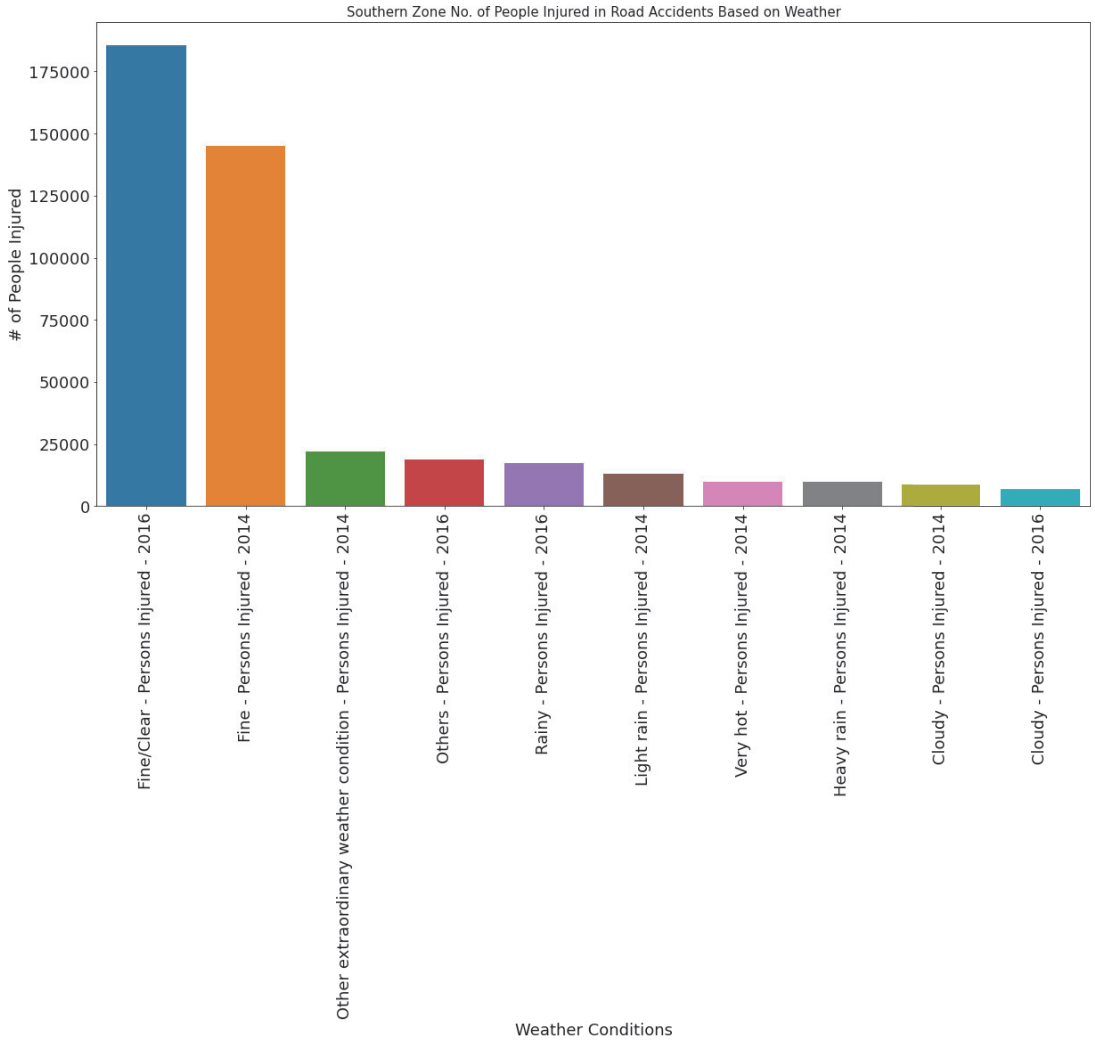
14.0.1 We can clearly infer that most deaths happend when the weather conditions are clear.

15 Weather Conditions - No. of People Injured in Road Accidents (South Zone)

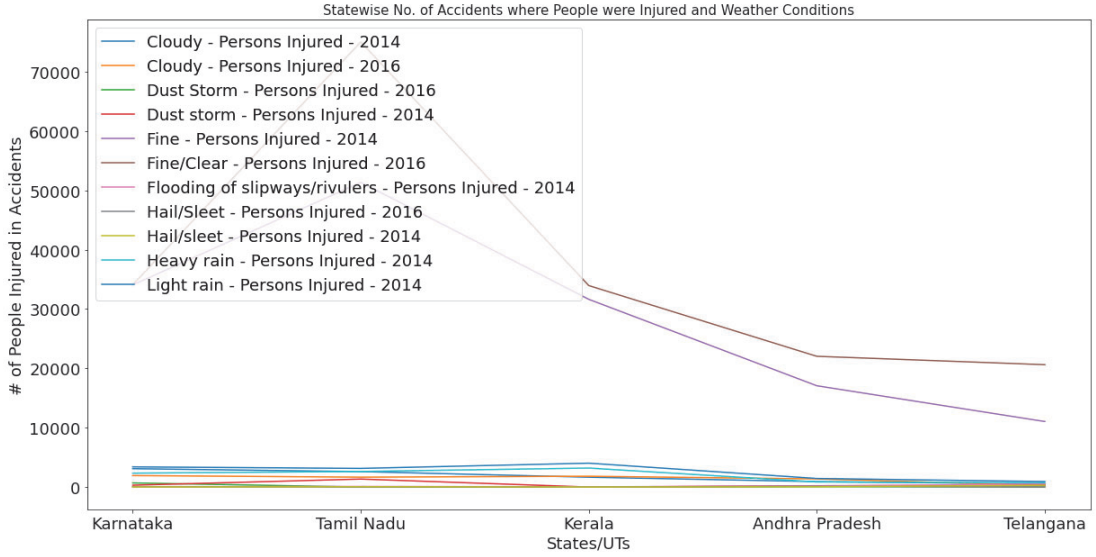
[66]:
```
sub_df = weather_df_injured[weather_df_injured['Zones'] == 'South Zone']
df =pd.pivot_table(sub_df, index=['Zones'],aggfunc=np.sum).reset_index()
df = df.T.reset_index()
df = df.rename(columns = {'index': 'Weather Conditions', 0: 'Total'})
df = df.drop(df.index[0])
df = df.sort_values(by = ['Total'], ascending=False).head(10)
df

fig,ax = plt.subplots(1,1, figsize=(20,10))

sns.barplot(x=df['Weather Conditions'],y=df['Total'])
plt.ylabel('# of People Injured')
plt.title('Southern Zone No. of People Injured in Road Accidents Based on␣
 ↪Weather', fontsize=15)
```

```
plt.xticks(rotation=90)
```

[66]:
```
(array([0, 1, 2, 3, 4, 5, 6, 7, 8, 9]),
 [Text(0, 0, 'Fine/Clear - Persons Injured - 2016'),
  Text(1, 0, 'Fine - Persons Injured - 2014'),
  Text(2, 0, 'Other extraordinary weather condition - Persons Injured - 2014'),
  Text(3, 0, 'Others - Persons Injured - 2016'),
  Text(4, 0, 'Rainy - Persons Injured - 2016'),
  Text(5, 0, 'Light rain - Persons Injured - 2014'),
  Text(6, 0, 'Very hot - Persons Injured - 2014'),
  Text(7, 0, 'Heavy rain - Persons Injured - 2014'),
  Text(8, 0, 'Cloudy - Persons Injured - 2014'),
  Text(9, 0, 'Cloudy - Persons Injured - 2016')])
```



[67]:
```
sub_df = weather_df_injured[weather_df_injured['Zones'] == 'South Zone']
df =pd.pivot_table(sub_df, index=['States/UTs'],aggfunc=np.sum).reset_index()
df = df.reset_index()
cols = list(df.columns[2:13])
df = df.sort_values(by=cols, ascending=False).head(5)

fig,ax = plt.subplots(1,1, figsize=(20,10))
for i,reason in enumerate(cols):
    sns.lineplot(x=df['States/UTs'],y=df[reason], label=cols[i])
    plt.ylabel('# of People Injured in Accidents')
    plt.title('Statewise No. of Accidents where People were Injured and Weather␣
 ↪Conditions', fontsize=15)
    plt.legend(loc='upper left')
```



Most of the accidents where people were killed happened when the weather was fine or had clear skies. 2016 reported around 38k accidents where people were killed while the weather was clear. Since we're looking at southern zone, it is evident that we wont see weather conditions like 'hail or snow' The more prevalent weather conditions are rainy/cloudy which the graph shows. Let's similarly look for Injuries. There wont be much difference in the features or X axis.

As per the https://www.prsindia.org/policy/vital-stats/overview-road-accidents-india Fewer accidents are caused due to neglect of civic bodies (2.8%), defect in motor vehicle (2.3%), and poor weather conditions (1.7%)

Though I cannot confirm for the first two, I see that poor weather condition was seldom the reason for a road accident. Here's another article which confrims

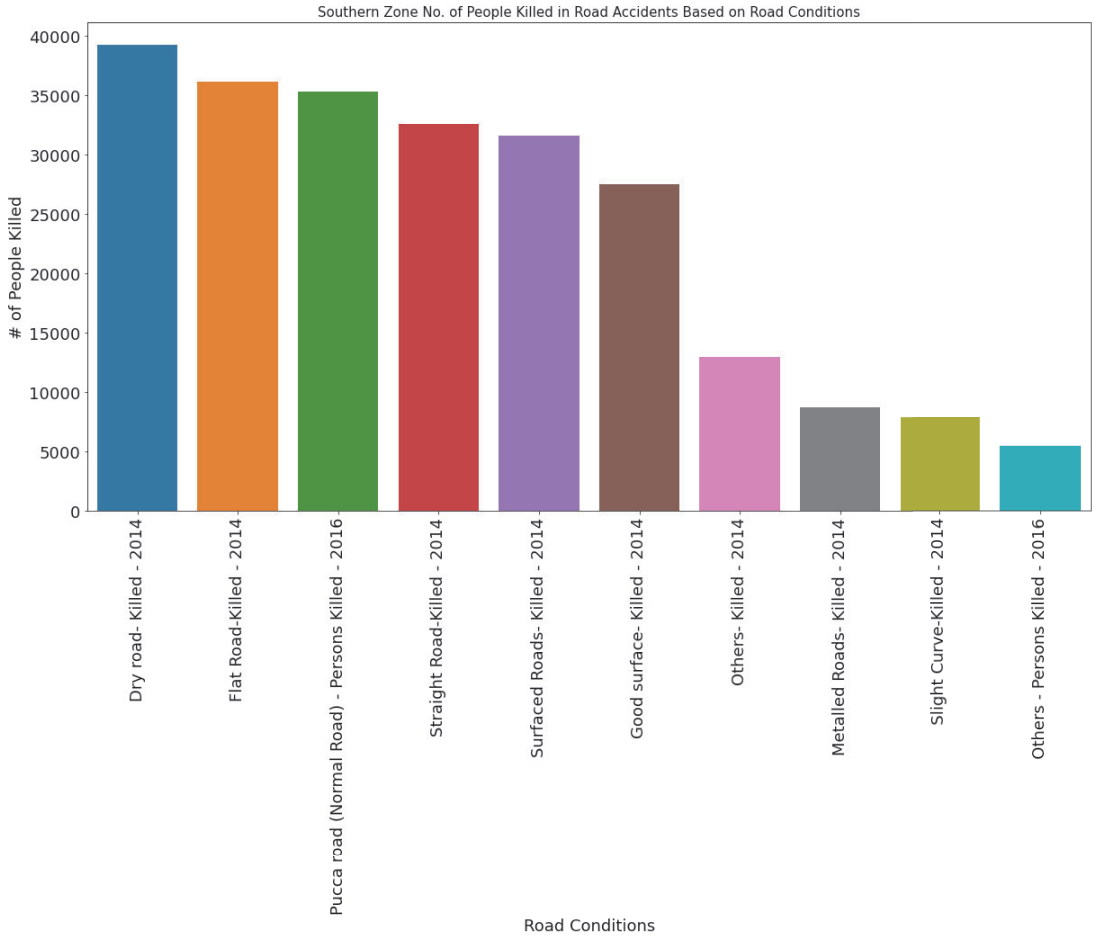
this. https://www.financialexpress.com/india-news/over-70-percent-road-accidents-occurred-on-bright-sunny-days-report/1348638/

Now, that we have the weather details, let look at the road conditions to see how that affects the no. of people killed/injured in road accidents

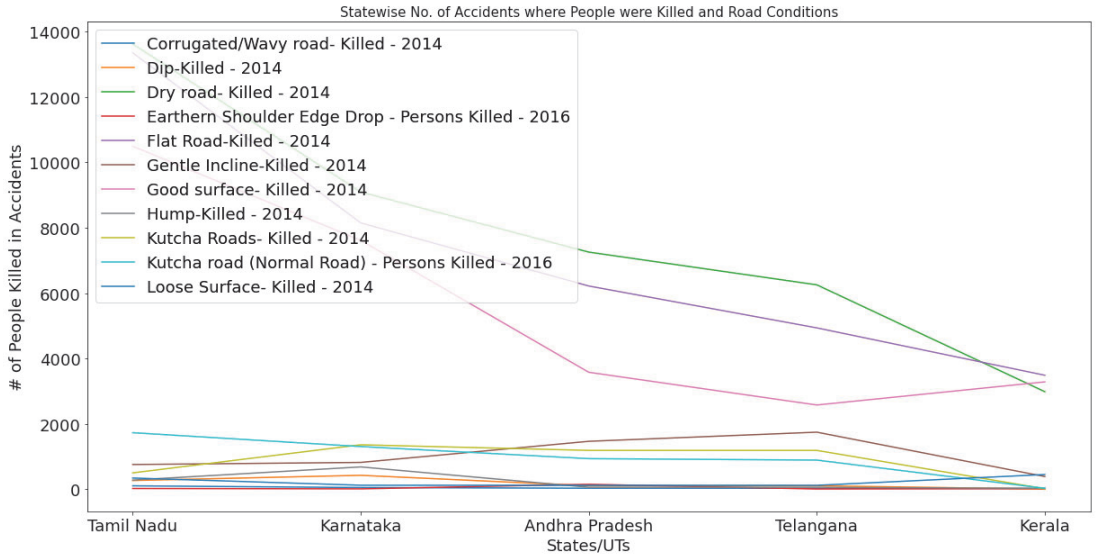
16 Road Conditions - No. of People Killed in Road Accidents (South Zone)

[68]:
```
sub_df = roadcond_df_killed[roadcond_df_killed['Zones'] == 'South Zone']
df =pd.pivot_table(sub_df, index=['Zones'],aggfunc=np.sum).reset_index()
df = df.T.reset_index()
df = df.rename(columns = {'index': 'Road Conditions', 0: 'Total'})
df = df.drop(df.index[0])
df = df.sort_values(by = ['Total'], ascending=False).head(10)
df

fig,ax = plt.subplots(1,1, figsize=(20,10))
sns.barplot(x=df['Road Conditions'],y=df['Total'])
plt.ylabel('# of People Killed')
plt.title('Southern Zone No. of People Killed in Road Accidents Based on Road␣
 ↪Conditions', fontsize=15)
plt.xticks(rotation=90)
```

[68]:
```
(array([0, 1, 2, 3, 4, 5, 6, 7, 8, 9]),
 [Text(0, 0, 'Dry road- Killed - 2014'),
  Text(1, 0, 'Flat Road-Killed - 2014'),
  Text(2, 0, 'Pucca road (Normal Road) - Persons Killed - 2016'),
  Text(3, 0, 'Straight Road-Killed - 2014'),
  Text(4, 0, 'Surfaced Roads- Killed - 2014'),
  Text(5, 0, 'Good surface- Killed - 2014'),
  Text(6, 0, 'Others- Killed - 2014'),
  Text(7, 0, 'Metalled Roads- Killed - 2014'),
  Text(8, 0, 'Slight Curve-Killed - 2014'),
  Text(9, 0, 'Others - Persons Killed - 2016')])
```



[69]:
```
sub_df = roadcond_df_killed[roadcond_df_killed['Zones'] == 'South Zone']
df =pd.pivot_table(sub_df, index=['States/UTs'],aggfunc=np.sum).reset_index()
df = df.reset_index()
cols = list(df.columns[2:13])
df = df.sort_values(by=cols, ascending=False).head(5)

fig,ax = plt.subplots(1,1, figsize=(20,10))
for i,reason in enumerate(cols):
    sns.lineplot(x=df['States/UTs'],y=df[reason], label=cols[i])
    plt.ylabel('# of People Killed in Accidents')
    plt.title('Statewise No. of Accidents where People were Killed and Road␣
 ↪Conditions', fontsize=15)
    plt.legend(loc='upper left')
```

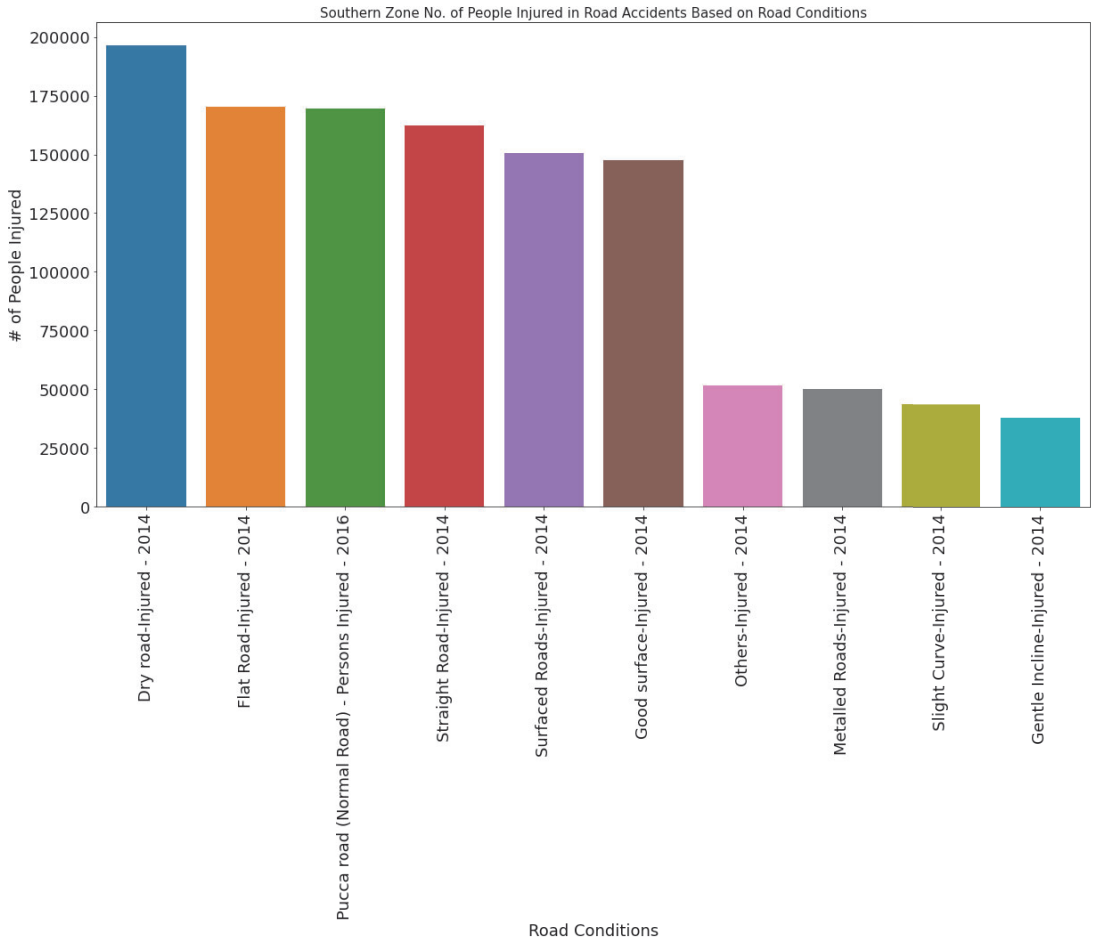


17 Road Conditions - No. of People Injured in Road Accidents (South Zone)

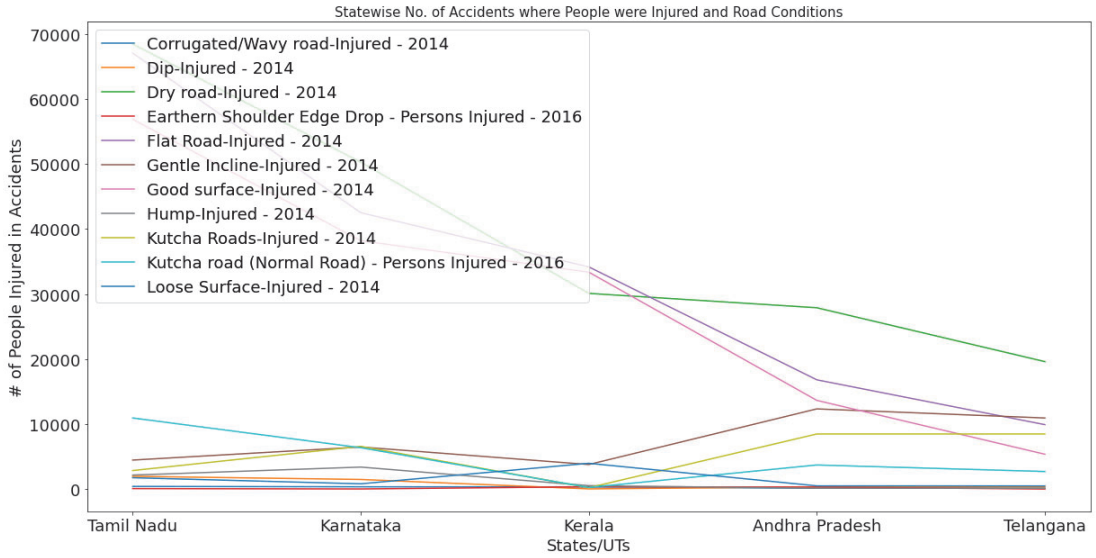
[70]:
```
sub_df = roadcond_df_injured[roadcond_df_injured['Zones'] == 'South Zone']
df =pd.pivot_table(sub_df, index=['Zones'],aggfunc=np.sum).reset_index()
df = df.T.reset_index()
df = df.rename(columns = {'index': 'Road Conditions', 0: 'Total'})
df = df.drop(df.index[0])
df = df.sort_values(by = ['Total'], ascending=False).head(10)
df

fig,ax = plt.subplots(1,1, figsize=(20,10))
sns.barplot(x=df['Road Conditions'],y=df['Total'])
plt.ylabel('# of People Injured')
plt.title('Southern Zone No. of People Injured in Road Accidents Based on Road␣
 ↪Conditions', fontsize=15)
plt.xticks(rotation=90)
```

[70]:
```
(array([0, 1, 2, 3, 4, 5, 6, 7, 8, 9]),
 [Text(0, 0, 'Dry road-Injured - 2014'),
  Text(1, 0, 'Flat Road-Injured - 2014'),
  Text(2, 0, 'Pucca road (Normal Road) - Persons Injured - 2016'),
  Text(3, 0, 'Straight Road-Injured - 2014'),
  Text(4, 0, 'Surfaced Roads-Injured - 2014'),
  Text(5, 0, 'Good surface-Injured - 2014'),
  Text(6, 0, 'Others-Injured - 2014'),
  Text(7, 0, 'Metalled Roads-Injured - 2014'),
```

```
Text(8, 0, 'Slight Curve-Injured - 2014'),
Text(9, 0, 'Gentle Incline-Injured - 2014')])
```



[71]:
```
sub_df = roadcond_df_injured[roadcond_df_injured['Zones'] == 'South Zone']
df =pd.pivot_table(sub_df, index=['States/UTs'],aggfunc=np.sum).reset_index()
df = df.reset_index()
cols = list(df.columns[2:13])
df = df.sort_values(by=cols, ascending=False).head(5)

fig,ax = plt.subplots(1,1, figsize=(20,10))
for i,reason in enumerate(cols):
    sns.lineplot(x=df['States/UTs'],y=df[reason], label=cols[i])
    plt.ylabel('# of People Injured in Accidents')
    plt.title('Statewise No. of Accidents where People were Injured and Road␣
 ↪Conditions', fontsize=15)
    plt.legend(loc='upper left')
```



As per the ET article cited earlier in the notebook, aside from factoring safety considerations into the design of a road, the quality of Indian roads has seen a notable improvement over the past decade or so. For instance, an annual survey of business executives conducted by the World Economic Forum on the quality of roads in around 140 countries reflected that. Between 2008 and 2018, India's rank in road quality rose from 87 to 51. The share of paved roads in our road network has increased from half in March 2008 to nearly two-thirds in March 2016, according to the latest available figures. India has a road network of 5.6 million km, of which national highways contribute just 2% and state highways 3%. Rural roads account for the lion's share, at 70%.

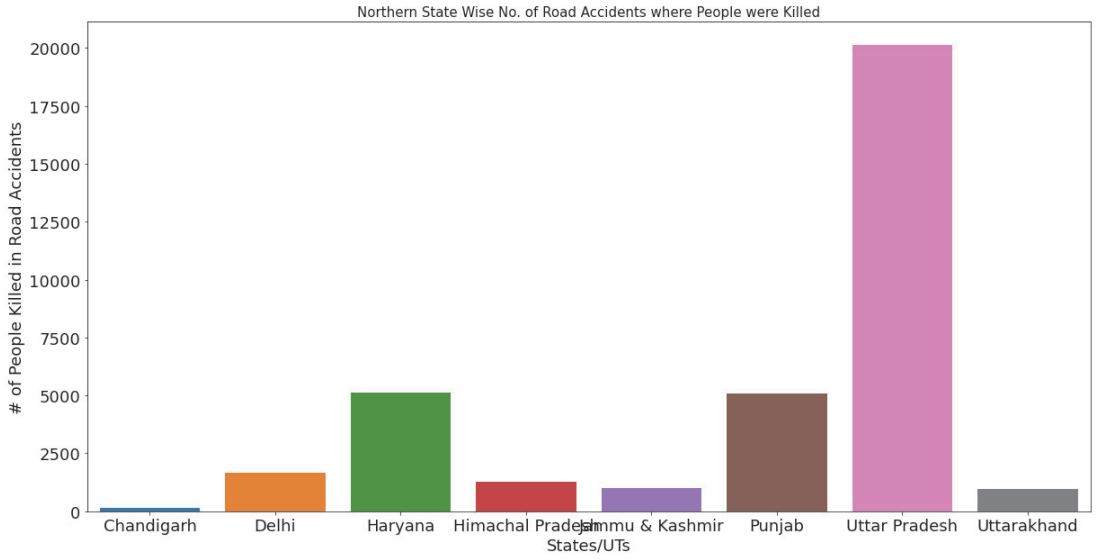
17.1 This zone-wise and state-wise comparison from two data sources eventually conclude on the same point that TN leads in Southern Zone with the most of people killed or injured in road accidents.

17.1.1 Let's perform a similar analysis for Northen Zone to check for UP.

18 Northern-Zone States: No. of Road Accidents where People were Killed

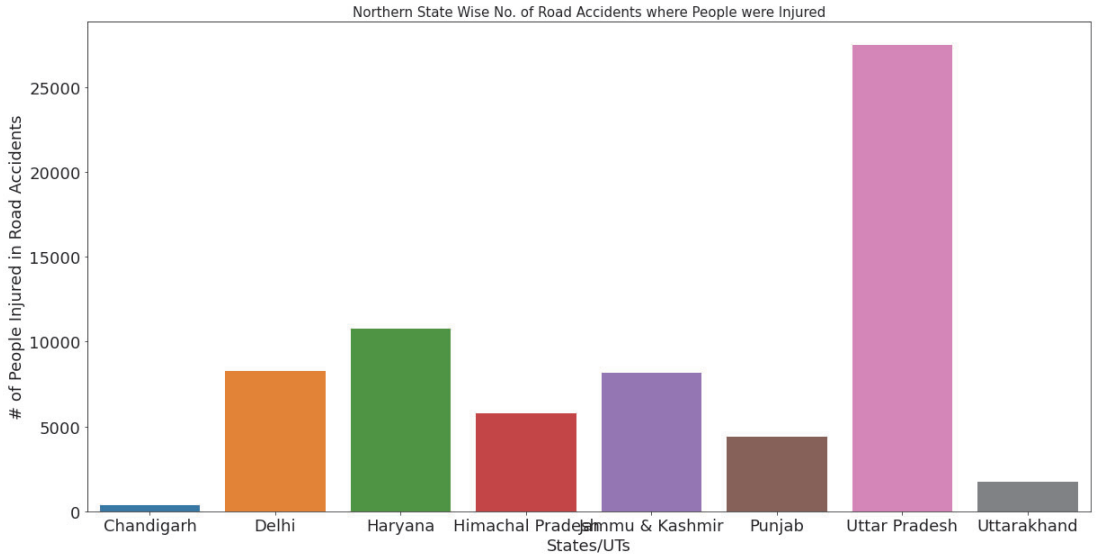
[72]:
```
sub_df = killed_df[killed_df['Zones'] == 'North Zone']
df =pd.pivot_table(sub_df, index=['States/UTs'],values=[2014, 2015, 2016,␣
↪2017],aggfunc=np.sum).reset_index()
df

years = [2014,2015,2016,2017]
fig,ax = plt.subplots(1,1, figsize=(20,10))
for i,year in enumerate(years):
```

```
    sns.barplot(x=df['States/UTs'],y=df[year])
    plt.ylabel('# of People Killed in Road Accidents')
    plt.title('Northern State Wise No. of Road Accidents where People were␣
 ↪Killed', fontsize=15)
```



19 Northern-Zone States: No. of Road Accidents where People were Injured

[73]:
```
sub_df = injured_df[injured_df['Zones'] == 'North Zone']
df =pd.pivot_table(sub_df, index=['States/UTs'],values=[2014, 2015, 2016,␣
 ↪2017],aggfunc=np.sum).reset_index()
df

years = [2014,2015,2016,2017]
fig,ax = plt.subplots(1,1, figsize=(20,10))
for i,year in enumerate(years):
    sns.barplot(x=df['States/UTs'],y=df[year])
    plt.ylabel('# of People Injured in Road Accidents')
    plt.title('Northern State Wise No. of Road Accidents where People were␣
 ↪Injured', fontsize=15)
```



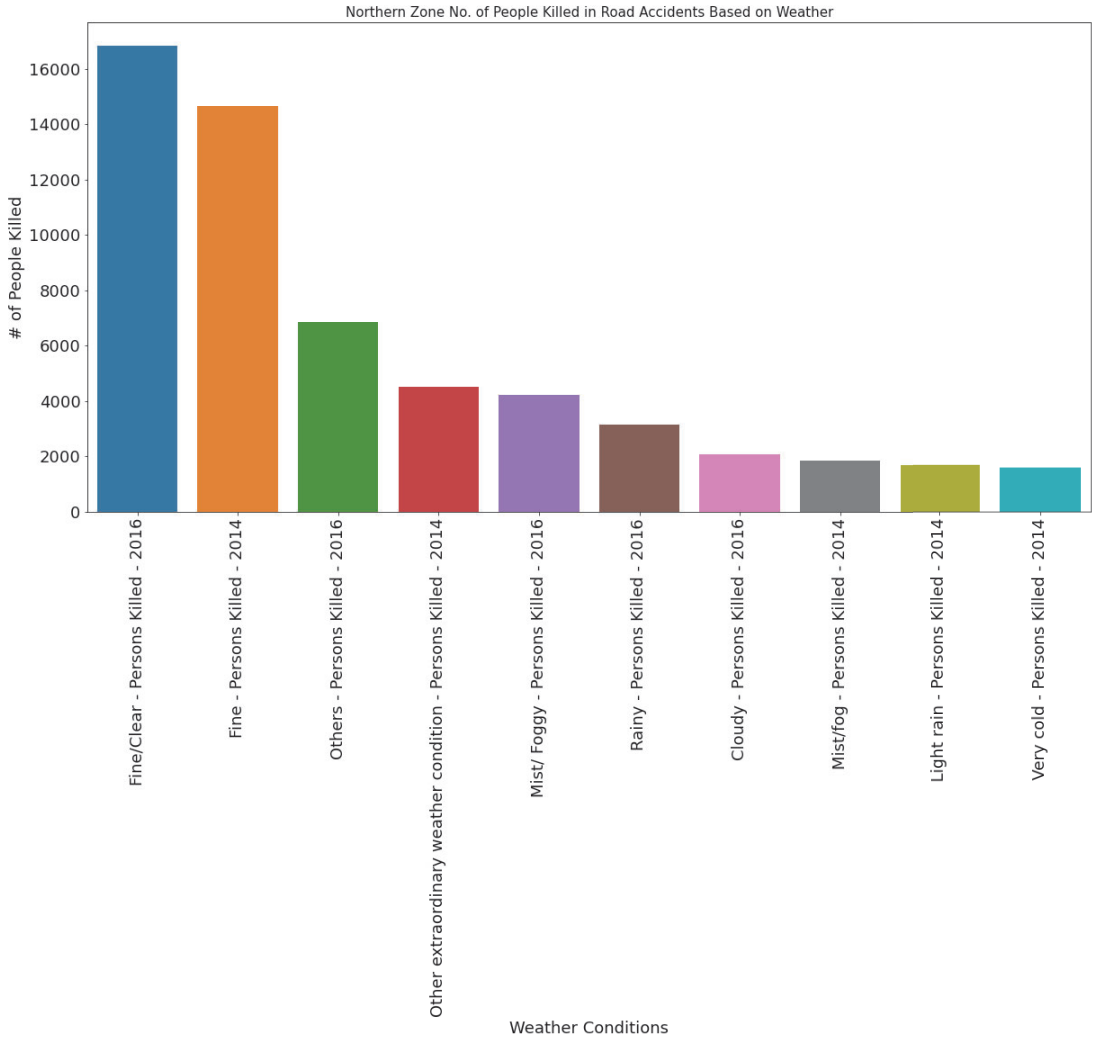
The top three states in the northern zone are Uttar Pradesh, Haryana and Delhi for most number of kills/injuries in road accidents.

20 Weather Conditions - No. of People Killed in Road Accidents (North Zone)

[74]:
```
sub_df = weather_df_killed[weather_df_killed['Zones'] == 'North Zone']
df =pd.pivot_table(sub_df, index=['Zones'],aggfunc=np.sum).reset_index()
df = df.T.reset_index()
df = df.rename(columns = {'index': 'Weather Conditions', 0: 'Total'})
df = df.drop(df.index[0])
df = df.sort_values(by = ['Total'], ascending=False).head(10)
df

fig,ax = plt.subplots(1,1, figsize=(20,10))
sns.barplot(x=df['Weather Conditions'],y=df['Total'])
plt.ylabel('# of People Killed')
plt.title('Northern Zone No. of People Killed in Road Accidents Based on␣
 ↪Weather', fontsize=15)
plt.xticks(rotation=90)
```

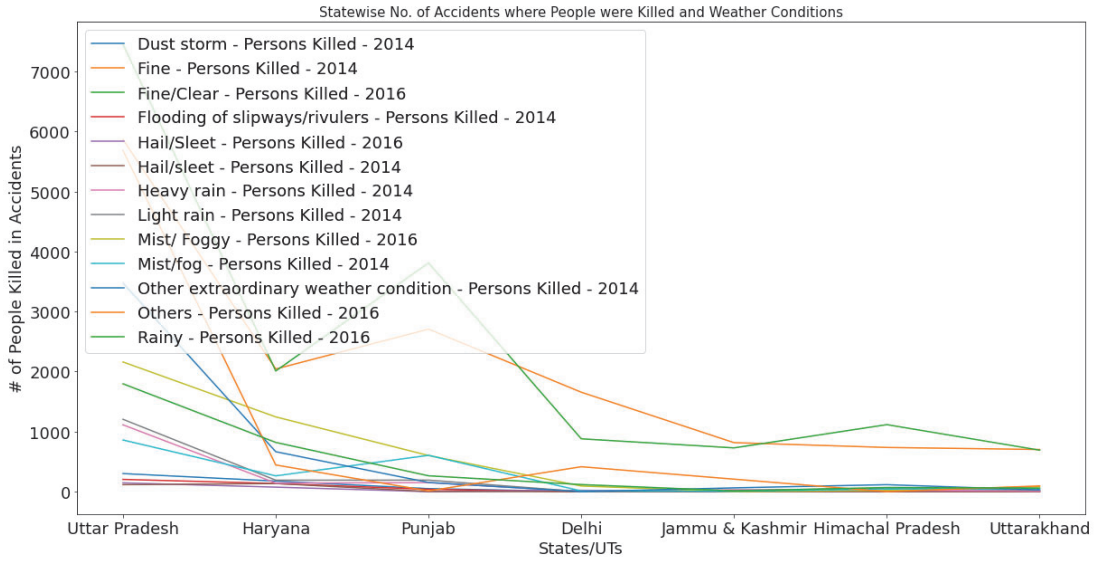
[74]:
```
(array([0, 1, 2, 3, 4, 5, 6, 7, 8, 9]),
 [Text(0, 0, 'Fine/Clear - Persons Killed - 2016'),
  Text(1, 0, 'Fine - Persons Killed - 2014'),
  Text(2, 0, 'Others - Persons Killed - 2016'),
  Text(3, 0, 'Other extraordinary weather condition - Persons Killed - 2014'),
  Text(4, 0, 'Mist/ Foggy - Persons Killed - 2016'),
```

```
Text(5, 0, 'Rainy - Persons Killed - 2016'),
Text(6, 0, 'Cloudy - Persons Killed - 2016'),
Text(7, 0, 'Mist/fog - Persons Killed - 2014'),
Text(8, 0, 'Light rain - Persons Killed - 2014'),
Text(9, 0, 'Very cold - Persons Killed - 2014')])
```



[75]:
```
sub_df = weather_df_killed[weather_df_killed['Zones'] == 'North Zone']
df =pd.pivot_table(sub_df, index=['States/UTs'],aggfunc=np.sum).reset_index()
df = df.reset_index()
cols = list(df.columns[5:18])
df = df.sort_values(by=cols, ascending=False).head(7)

fig,ax = plt.subplots(1,1, figsize=(20,10))
for i,reason in enumerate(cols):
```

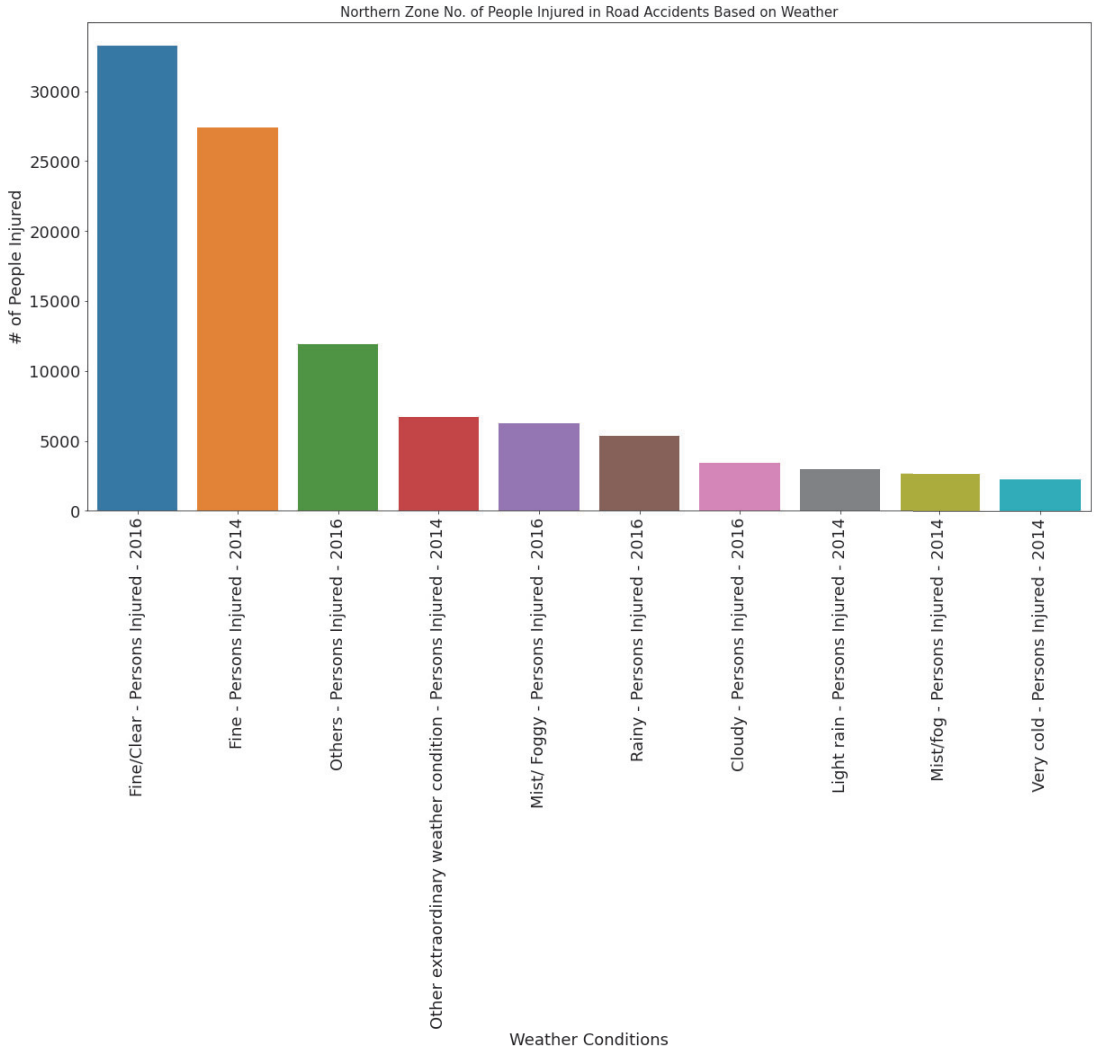
```
    sns.lineplot(x=df['States/UTs'],y=df[reason], label=cols[i])
    plt.ylabel('# of People Killed in Accidents')
    plt.title('Statewise No. of Accidents where People were Killed and Weather␣
 ↪Conditions', fontsize=15)
    plt.legend(loc='upper left')
```



21 Weather Conditions - No. of People Injured in Road Accidents (North Zone)

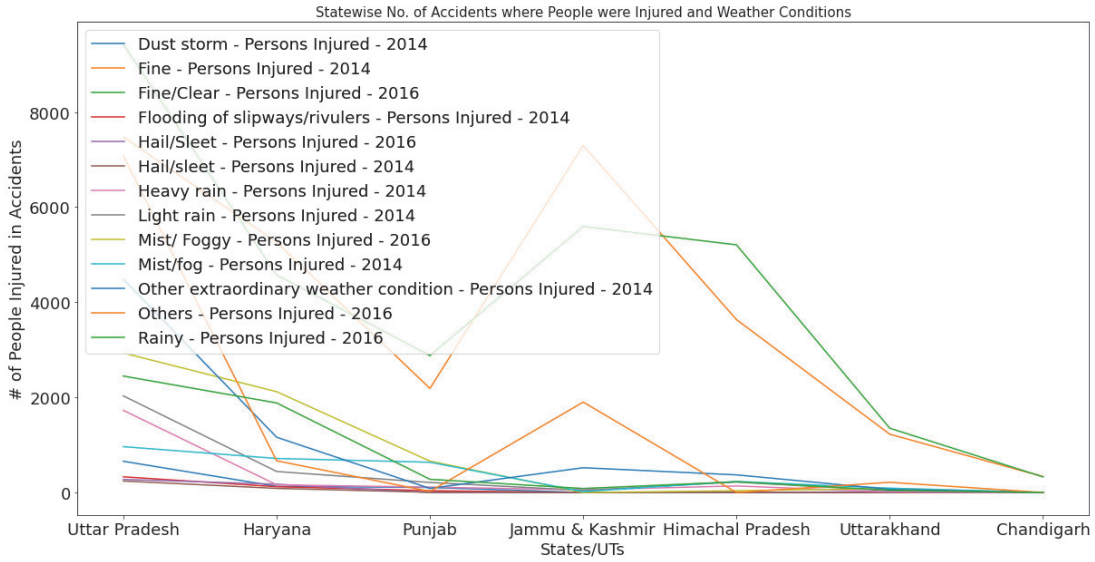
[76]:
```
sub_df = weather_df_injured[weather_df_injured['Zones'] == 'North Zone']
df =pd.pivot_table(sub_df, index=['Zones'],aggfunc=np.sum).reset_index()
df = df.T.reset_index()
df = df.rename(columns = {'index': 'Weather Conditions', 0: 'Total'})
df = df.drop(df.index[0])
df = df.sort_values(by = ['Total'], ascending=False).head(10)
df

fig,ax = plt.subplots(1,1, figsize=(20,10))

sns.barplot(x=df['Weather Conditions'],y=df['Total'])
plt.ylabel('# of People Injured')
plt.title('Northern Zone No. of People Injured in Road Accidents Based on␣
 ↪Weather', fontsize=15)
plt.xticks(rotation=90)
```

[76]:
```
(array([0, 1, 2, 3, 4, 5, 6, 7, 8, 9]),
 [Text(0, 0, 'Fine/Clear - Persons Injured - 2016'),
  Text(1, 0, 'Fine - Persons Injured - 2014'),
  Text(2, 0, 'Others - Persons Injured - 2016'),
  Text(3, 0, 'Other extraordinary weather condition - Persons Injured - 2014'),
  Text(4, 0, 'Mist/ Foggy - Persons Injured - 2016'),
  Text(5, 0, 'Rainy - Persons Injured - 2016'),
  Text(6, 0, 'Cloudy - Persons Injured - 2016'),
  Text(7, 0, 'Light rain - Persons Injured - 2014'),
  Text(8, 0, 'Mist/fog - Persons Injured - 2014'),
  Text(9, 0, 'Very cold - Persons Injured - 2014')])
```



[77]:
```
sub_df = weather_df_injured[weather_df_injured['Zones'] == 'North Zone']
df =pd.pivot_table(sub_df, index=['States/UTs'],aggfunc=np.sum).reset_index()
```

```
df = df.reset_index()
cols = list(df.columns[5:18])
df = df.sort_values(by=cols, ascending=False).head(7)

fig,ax = plt.subplots(1,1, figsize=(20,10))
for i,reason in enumerate(cols):
    sns.lineplot(x=df['States/UTs'],y=df[reason], label=cols[i])
    plt.ylabel('# of People Injured in Accidents')
    plt.title('Statewise No. of Accidents where People were Injured and Weather␣
 ↪Conditions', fontsize=15)
    plt.legend(loc='upper left')
```



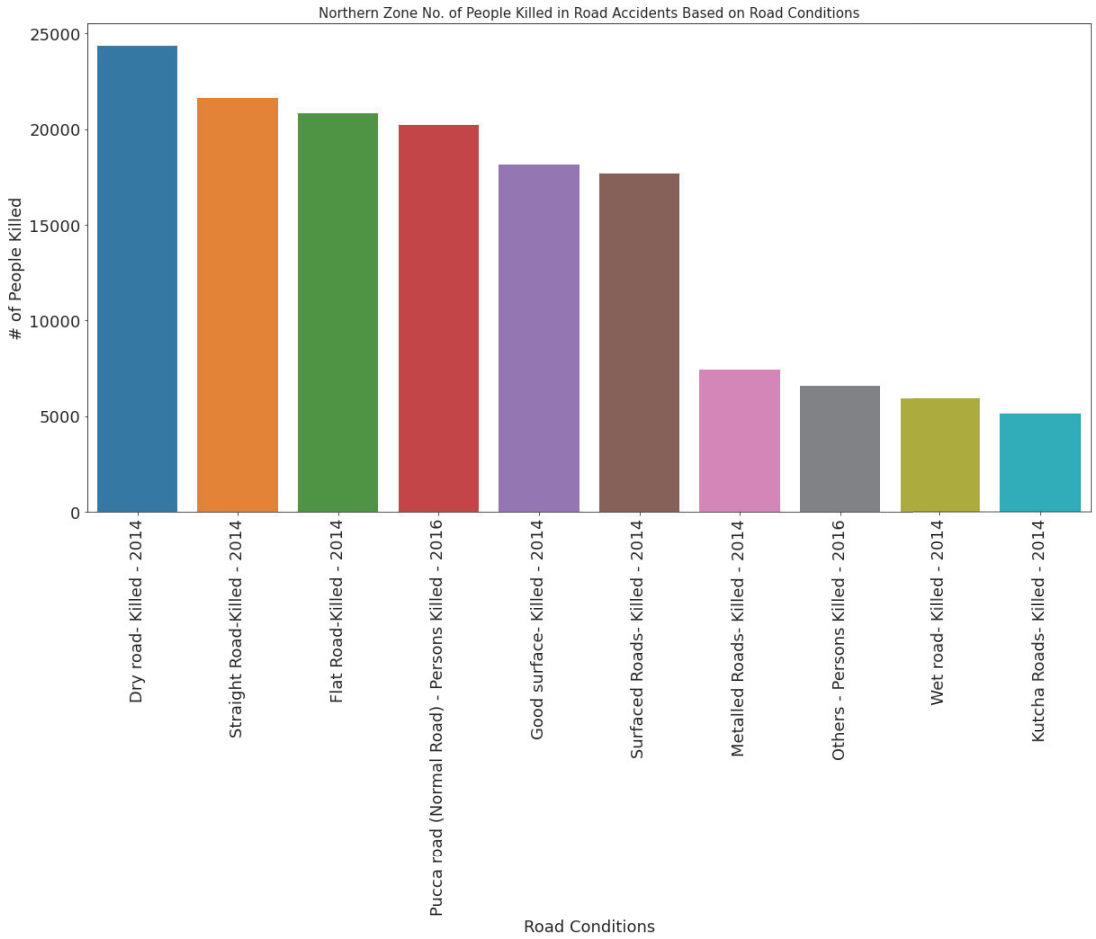
The same pattern for weather continues for the northern states as well. Though here we see features which are appropriate for northern states - mist/foggy, hail etc.

22 Road Conditions - No. of People Killed in Road Accidents (North Zone)

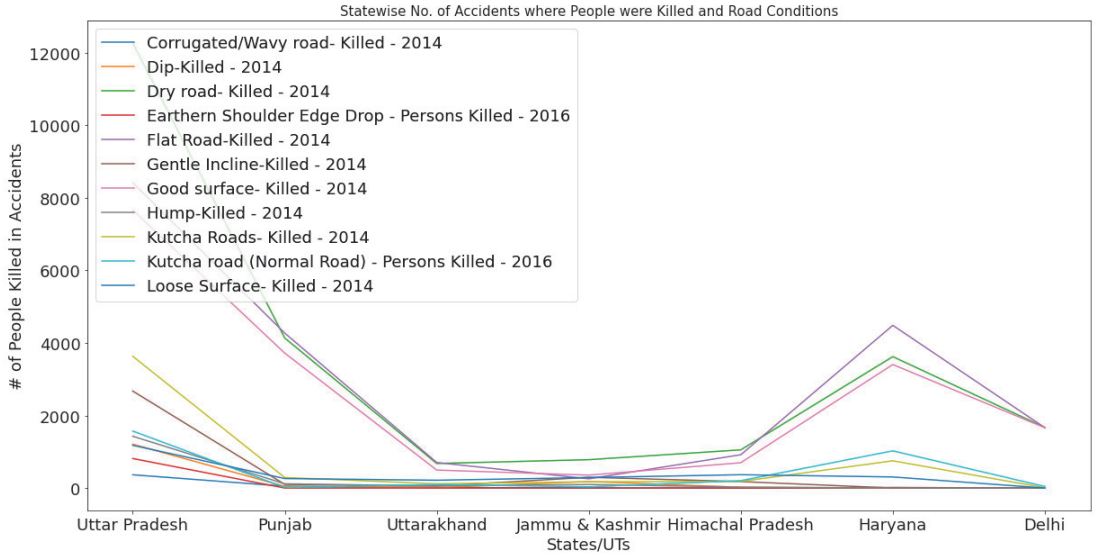
[78]:
```
sub_df = roadcond_df_killed[roadcond_df_killed['Zones'] == 'North Zone']
df =pd.pivot_table(sub_df, index=['Zones'],aggfunc=np.sum).reset_index()
df = df.T.reset_index()
df = df.rename(columns = {'index': 'Road Conditions', 0: 'Total'})
df = df.drop(df.index[0])
df = df.sort_values(by = ['Total'], ascending=False).head(10)
df
```

```
fig,ax = plt.subplots(1,1, figsize=(20,10))
sns.barplot(x=df['Road Conditions'],y=df['Total'])
plt.ylabel('# of People Killed')
plt.title('Northern Zone No. of People Killed in Road Accidents Based on Road␣
 ↪Conditions', fontsize=15)
plt.xticks(rotation=90)
```

[78]:
```
(array([0, 1, 2, 3, 4, 5, 6, 7, 8, 9]),
 [Text(0, 0, 'Dry road- Killed - 2014'),
  Text(1, 0, 'Straight Road-Killed - 2014'),
  Text(2, 0, 'Flat Road-Killed - 2014'),
  Text(3, 0, 'Pucca road (Normal Road) - Persons Killed - 2016'),
  Text(4, 0, 'Good surface- Killed - 2014'),
  Text(5, 0, 'Surfaced Roads- Killed - 2014'),
  Text(6, 0, 'Metalled Roads- Killed - 2014'),
  Text(7, 0, 'Others - Persons Killed - 2016'),
  Text(8, 0, 'Wet road- Killed - 2014'),
  Text(9, 0, 'Kutcha Roads- Killed - 2014')])
```



[79]:
```
sub_df = roadcond_df_killed[roadcond_df_killed['Zones'] == 'North Zone']
df =pd.pivot_table(sub_df, index=['States/UTs'],aggfunc=np.sum).reset_index()
df = df.reset_index()
cols = list(df.columns[2:13])
df = df.sort_values(by=cols, ascending=False).head(7)

fig,ax = plt.subplots(1,1, figsize=(20,10))
for i,reason in enumerate(cols):
    sns.lineplot(x=df['States/UTs'],y=df[reason], label=cols[i])
    plt.ylabel('# of People Killed in Accidents')
    plt.title('Statewise No. of Accidents where People were Killed and Road␣
 ↪Conditions', fontsize=15)
    plt.legend(loc='upper left')
```

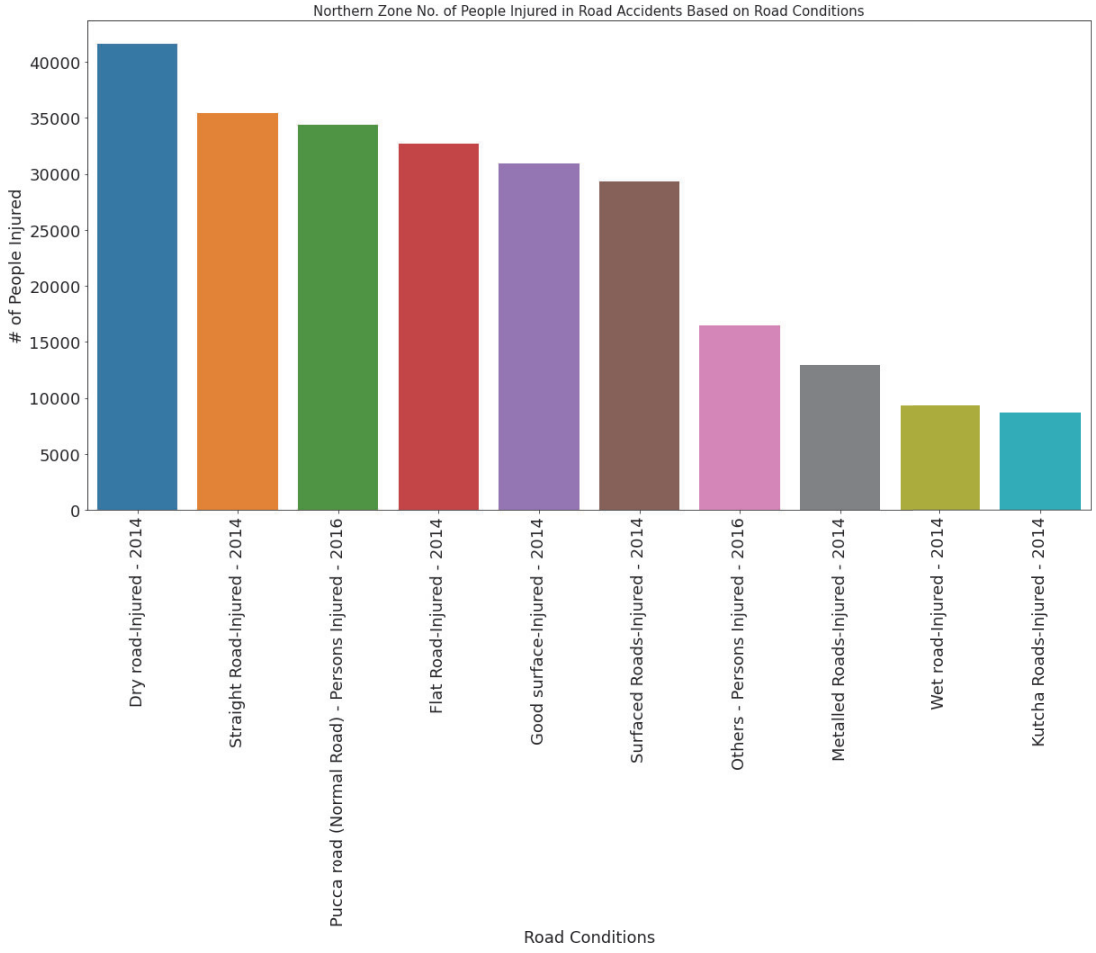


23 Road Conditions - No. of People Injured in Road Accidents (North Zone)

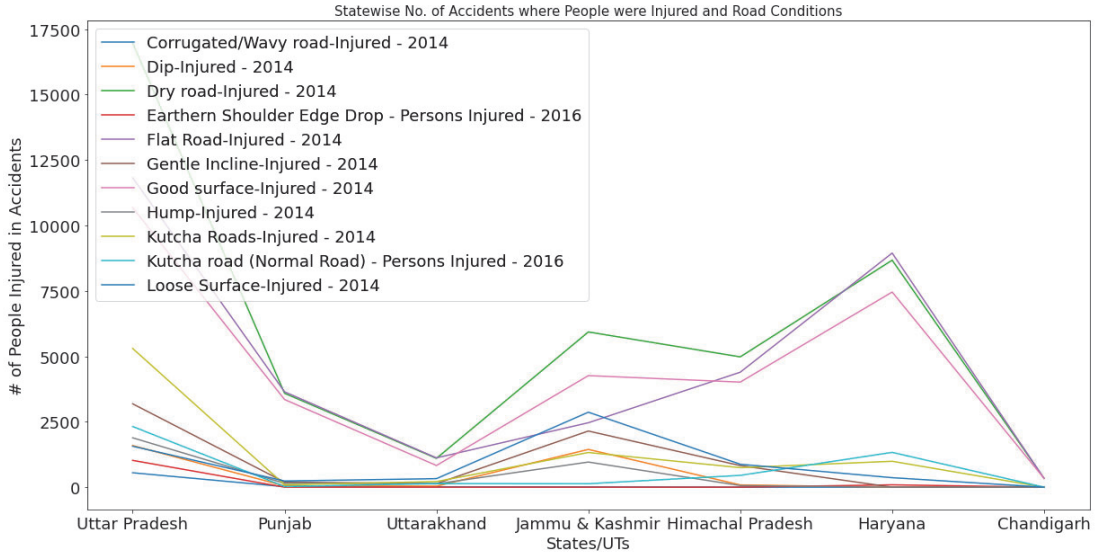
[80]:
```
sub_df = roadcond_df_injured[roadcond_df_injured['Zones'] == 'North Zone']
df =pd.pivot_table(sub_df, index=['Zones'],aggfunc=np.sum).reset_index()
df = df.T.reset_index()
df = df.rename(columns = {'index': 'Road Conditions', 0: 'Total'})
df = df.drop(df.index[0])
df = df.sort_values(by = ['Total'], ascending=False).head(10)
df
```

```
fig,ax = plt.subplots(1,1, figsize=(20,10))
sns.barplot(x=df['Road Conditions'],y=df['Total'])
plt.ylabel('# of People Injured')
plt.title('Northern Zone No. of People Injured in Road Accidents Based on Road␣
 ↪Conditions', fontsize=15)
plt.xticks(rotation=90)
```

[80]:
```
(array([0, 1, 2, 3, 4, 5, 6, 7, 8, 9]),
 [Text(0, 0, 'Dry road-Injured - 2014'),
  Text(1, 0, 'Straight Road-Injured - 2014'),
  Text(2, 0, 'Pucca road (Normal Road) - Persons Injured - 2016'),
  Text(3, 0, 'Flat Road-Injured - 2014'),
  Text(4, 0, 'Good surface-Injured - 2014'),
  Text(5, 0, 'Surfaced Roads-Injured - 2014'),
  Text(6, 0, 'Others - Persons Injured - 2016'),
  Text(7, 0, 'Metalled Roads-Injured - 2014'),
  Text(8, 0, 'Wet road-Injured - 2014'),
  Text(9, 0, 'Kutcha Roads-Injured - 2014')])
```



[81]:
```
sub_df = roadcond_df_injured[roadcond_df_injured['Zones'] == 'North Zone']
df =pd.pivot_table(sub_df, index=['States/UTs'],aggfunc=np.sum).reset_index()
df = df.reset_index()
cols = list(df.columns[2:13])
df = df.sort_values(by=cols, ascending=False).head(7)

fig,ax = plt.subplots(1,1, figsize=(20,10))
for i,reason in enumerate(cols):
    sns.lineplot(x=df['States/UTs'],y=df[reason], label=cols[i])
    plt.ylabel('# of People Injured in Accidents')
    plt.title('Statewise No. of Accidents where People were Injured and Road␣
 ↪Conditions', fontsize=15)
    plt.legend(loc='upper left')
```



[ ]:

24 Linear Regression

[82]:
```
# read the train and test dataset
k_data= killed_df
i_data= injured_df

k_data= k_data.drop(columns = ['States/UTs','Zones'])
i_data= i_data.drop(columns = ['States/UTs','Zones'])
```

```
print(k_data.head())

# shape of the dataset
print('\nShape of killed data :',k_data.shape)
print('\nShape of injured data :',i_data.shape)

# Now, we need to predict the missing target variable in the test data
# target variable - Item_Outlet_Sales

# seperate the independent and target variable on killed data
train_x = k_data.drop(columns=[2017],axis=1)
train_y = k_data[[2017]]

# seperate the independent and target variable on killed data
test_x = i_data.drop(columns=[2017],axis=1)
test_y = i_data[[2017]]

model = LinearRegression()

# fit the model with the killed data
model.fit(train_x,train_y)

# coefficeints of the killed model
print('\nCoefficient of model :', model.coef_)

# intercept of the model
print('\nIntercept of model',model.intercept_)

# predict the target on the injured dataset
predict_train = model.predict(train_x)
print('\n2017 killed on killed data',predict_train)

# Root Mean Squared Error on killed dataset
rmse_train = mean_squared_error(train_y,predict_train)**(0.5)
print('\nRMSE on killed dataset : ', rmse_train)

# predict the target on the injured dataset
predict_test = model.predict(test_x)
print('\n2017 injured on injured data',predict_test)

# Root Mean Squared Error on injured dataset
rmse_test = mean_squared_error(test_y,predict_test)**(0.5)
print('\nRMSE on injured dataset : ', rmse_test)
```

```
   2014  2015  2016  2017
0  7908  8297  8541  8060
1   119   127   149   110
2  2522  2397  2572  2783
```

```
3  4913  5421  4901  5554
4  4022  4082  3908  4136

Shape of killed data : (36, 4)

Shape of injured data : (36, 4)

Coefficient of model : [[-1.667533    1.70188226  0.86866233]]

Intercept of model [31.57199162]

2017 killed on killed data [[ 8.38448311e+03]
 [ 1.78705299e+02]
 [ 2.13966506e+03]
 [ 5.32220018e+03]
 [ 3.66657004e+03]
 [ 3.69143348e+02]
 [ 7.65136577e+03]
 [ 5.22366465e+03]
 [ 1.00153270e+03]
 [ 7.70183801e+02]
 [ 3.20228152e+03]
 [ 1.07489686e+04]
 [ 4.14478425e+03]
 [ 9.97292993e+03]
 [ 1.24035627e+04]
 [ 5.83497306e+01]
 [ 2.38193642e+02]
 [ 4.31579785e+01]
 [-1.24832465e+01]
 [ 4.67653912e+03]
 [ 5.06341055e+03]
 [ 9.85165881e+03]
 [ 1.26155601e+02]
 [ 1.62792161e+04]
 [ 6.88684533e+03]
 [ 1.37251768e+02]
 [ 9.56949684e+02]
 [ 1.97204703e+04]
 [ 6.52887592e+03]
 [ 4.71292842e+01]
 [ 1.63835992e+02]
 [ 4.46250667e+01]
 [ 1.11047220e+02]
 [ 1.38761914e+03]
 [ 3.24406540e+01]
 [ 3.91670449e+02]]
```

```
RMSE on killed dataset :  276.09276705936844

2017 injured on injured data [[ 2.63265253e+04]
 [ 4.68594531e+02]
 [ 6.54547295e+03]
 [ 5.50032896e+03]
 [ 1.21948320e+04]
 [ 2.15555541e+03]
 [ 1.63546678e+04]
 [ 1.26351570e+04]
 [ 4.43359224e+03]
 [ 7.15808009e+03]
 [ 2.93483503e+03]
 [ 4.96126805e+04]
 [ 4.42494146e+04]
 [ 5.30212869e+04]
 [ 3.11473524e+04]
 [ 7.45649877e+02]
 [ 3.63376135e+02]
 [-1.24267819e+02]
 [-1.21781832e+02]
 [ 1.14946996e+04]
 [ 4.44132140e+03]
 [ 1.96994835e+04]
 [ 2.46592890e+02]
 [ 7.75127754e+04]
 [ 2.40440178e+04]
 [ 4.79347998e+02]
 [ 1.80572702e+03]
 [ 2.40760151e+04]
 [ 1.03646264e+04]
 [ 4.03561114e+02]
 [ 3.22061372e+02]
 [ 1.49497506e+02]
 [ 1.47386897e+02]
 [ 6.48795018e+03]
 [ 3.50101054e+01]
 [ 2.01817802e+03]]

RMSE on injured dataset :  1407.2231630658664
```

[83]:
```
ac_lr=model.score(test_x, test_y)
print(f"Model R^2: {ac_lr}")
print('Slope:' ,model.coef_)
print('Intercept:',model.intercept_)
```

```
Model R^2: 0.9940402101620864
Slope: [[-1.667533    1.70188226  0.86866233]]
```

```
Intercept: [31.57199162]
```

[84]:
```
Y_Pred = model.predict(test_x)
print('Linear Regression R squared: %.2f' % model.score(test_x, test_y))
```

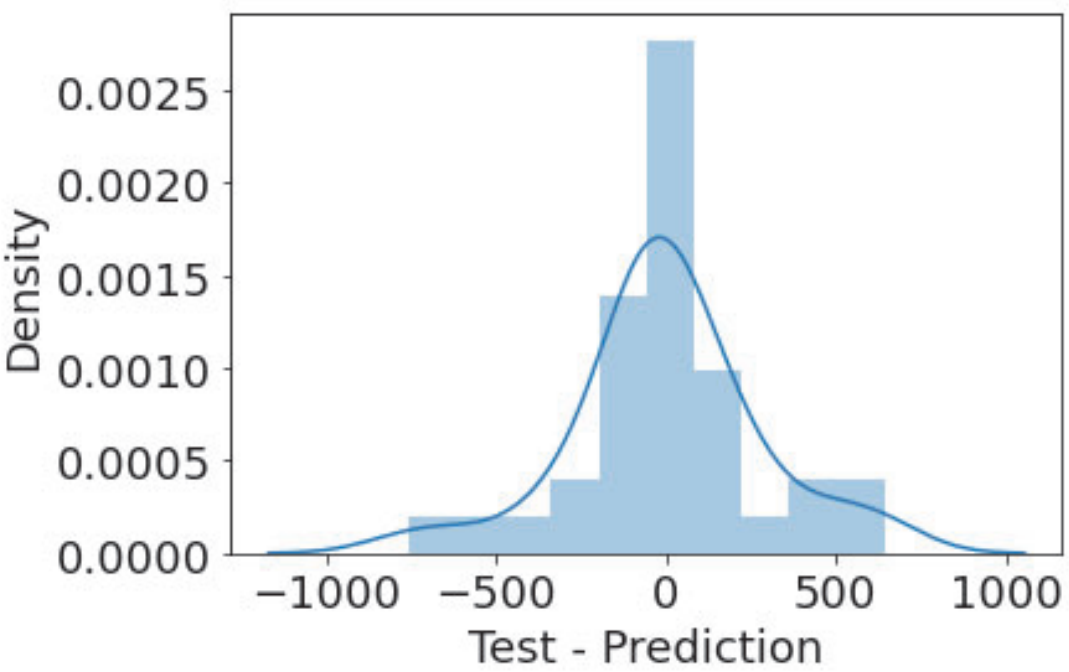
```
Linear Regression R squared: 0.99
```

[85]:
```
mse = mean_squared_error(Y_Pred, test_y)
rmse = np.sqrt(mse)
print('Linear Regression RMSE: %.2f' % rmse)
```

```
Linear Regression RMSE: 1407.22
```

[86]:
```
model = LinearRegression()
model.fit(train_x,train_y)
predictions = model.predict(train_x)
sns.distplot(train_y - predictions, axlabel="Test - Prediction")
plt.show()
```

```
C:\Users\sjaru\anaconda3\lib\site-packages\seaborn\distributions.py:2557:
FutureWarning:

`distplot` is a deprecated function and will be removed in a future version.
Please adapt your code to use either `displot` (a figure-level function with
similar flexibility) or `histplot` (an axes-level function for histograms).
```



[ ]:

25 K__Means

[87]:
```
#k_data= killed_df
#i_data= injured_df
# read the train and test dataset
train_data = k_data
test_data = i_data

# shape of the dataset
print('Shape of killed data :',train_data.shape)
print('Shape of injured data :',test_data.shape)

# Now, we need to divide the training data into differernt clusters
# and predict in which cluster a particular data point belongs.

model = KMeans()

# fit the model with the training data
model.fit(train_data)

# Number of Clusters
print('\nDefault number of Clusters : ',model.n_clusters)

# predict the clusters on the train dataset
predict_train = model.predict(train_data)
print('\nCLusters on killed data',predict_train)

# predict the target on the test dataset
predict_test = model.predict(test_data)
print('Clusters on injured data',predict_test)

# Now, we will train a model with n_cluster = 3
model_n3 = KMeans(n_clusters=3)

# fit the model with the training data
model_n3.fit(train_data)

# Number of Clusters
print('\nNumber of Clusters : ',model_n3.n_clusters)

# predict the clusters on the train dataset
predict_train_3 = model_n3.predict(train_data)
print('\nCLusters on killed data',predict_train_3)
```

```
# predict the target on the test dataset
predict_test_3 = model_n3.predict(test_data)
print('Clusters on injured data',predict_test_3)
```

```
Shape of killed data : (36, 4)
Shape of injured data : (36, 4)

Default number of Clusters :  8

CLusters on killed data [3 2 6 0 0 2 3 0 2 2 6 5 0 5 1 2 2 2 2 0 0 5 2 7 3 2 2 4
3 2 2 2 2 6 2 2]
Clusters on injured data [4 2 3 3 1 6 4 5 0 3 0 4 4 4 4 2 2 2 2 5 0 4 2 4 4 2 6
4 1 2 2 2 2 3 2 6]

Number of Clusters :  3

CLusters on killed data [0 1 1 0 0 1 0 0 1 1 1 2 0 0 2 1 1 1 1 0 0 2 1 2 0 1 1 2
0 1 1 1 1 1 1 1]
Clusters on injured data [2 1 0 0 2 1 2 2 0 0 0 2 2 2 2 1 1 1 1 2 0 2 1 2 2 1 1
2 2 1 1 1 1 0 1 1]
```

[88]:
```
km = KMeans(n_clusters=2)
km.fit(k_data)
print(km.fit(k_data))
print('\n')

def convertToCluster(cluster):
    if cluster=='Yes':
        return 1
    else:
        return 0
k_data['Cluster'] = k_data[2014].apply(convertToCluster)
print(confusion_matrix(k_data['Cluster'],km.labels_))
print(classification_report(k_data['Cluster'],km.labels_))
```

```
KMeans(n_clusters=2)


[[27  9]
 [ 0  0]]
              precision    recall  f1-score   support

           0       1.00      0.75      0.86        36
           1       0.00      0.00      0.00         0

    accuracy                           0.75        36
   macro avg       0.50      0.38      0.43        36
weighted avg       1.00      0.75      0.86        36
```

```
C:\Users\sjaru\anaconda3\lib\site-
packages\sklearn\metrics\_classification.py:1245: UndefinedMetricWarning:

Recall and F-score are ill-defined and being set to 0.0 in labels with no true
samples. Use `zero_division` parameter to control this behavior.

C:\Users\sjaru\anaconda3\lib\site-
packages\sklearn\metrics\_classification.py:1245: UndefinedMetricWarning:

Recall and F-score are ill-defined and being set to 0.0 in labels with no true
samples. Use `zero_division` parameter to control this behavior.

C:\Users\sjaru\anaconda3\lib\site-
packages\sklearn\metrics\_classification.py:1245: UndefinedMetricWarning:

Recall and F-score are ill-defined and being set to 0.0 in labels with no true
samples. Use `zero_division` parameter to control this behavior.
```

[ ]:

26 Random Forest

[89]:
```
from sklearn.ensemble import RandomForestRegressor
clf = RandomForestRegressor(n_estimators=10)
clf.fit(train_x,train_y)
y_pred = clf.predict(test_x)
ac_rf=clf.score(test_x, test_y)
ac_rf
```

```
<ipython-input-89-89fe581333d1>:3: DataConversionWarning:

A column-vector y was passed when a 1d array was expected. Please change the
shape of y to (n_samples,), for example using ravel().
```

[89]:
```
0.3652928878587758
```

[90]:
```
from sklearn import metrics
print('Mean Absolute Error:', metrics.mean_absolute_error(test_y, y_pred))
print('Mean Squared Error:', metrics.mean_squared_error(test_y, y_pred))
print('Root Mean Squared Error:', np.sqrt(metrics.mean_squared_error(test_y,␣
 ↪y_pred)))
```

```
Mean Absolute Error: 6088.5666666666675
Mean Squared Error: 210896013.04055557
Root Mean Squared Error: 14522.259226461823
```

27 Decision Tree

[91]:
```
from sklearn.tree import DecisionTreeRegressor
regressor = DecisionTreeRegressor()
regressor.fit(train_x,train_y)
y_pred = regressor.predict(test_x)
```

[92]:
```
y_pred = regressor.predict(test_x)
ac_dt=regressor.score(test_x, test_y)
print(f"Model R^2: {ac_dt}")
print('Mean Absolute Error:', metrics.mean_absolute_error(test_y, y_pred))
print('Mean Squared Error:', metrics.mean_squared_error(test_y, y_pred))
print('Root Mean Squared Error:', np.sqrt(metrics.mean_squared_error(test_y,␣
 ↪y_pred)))
```

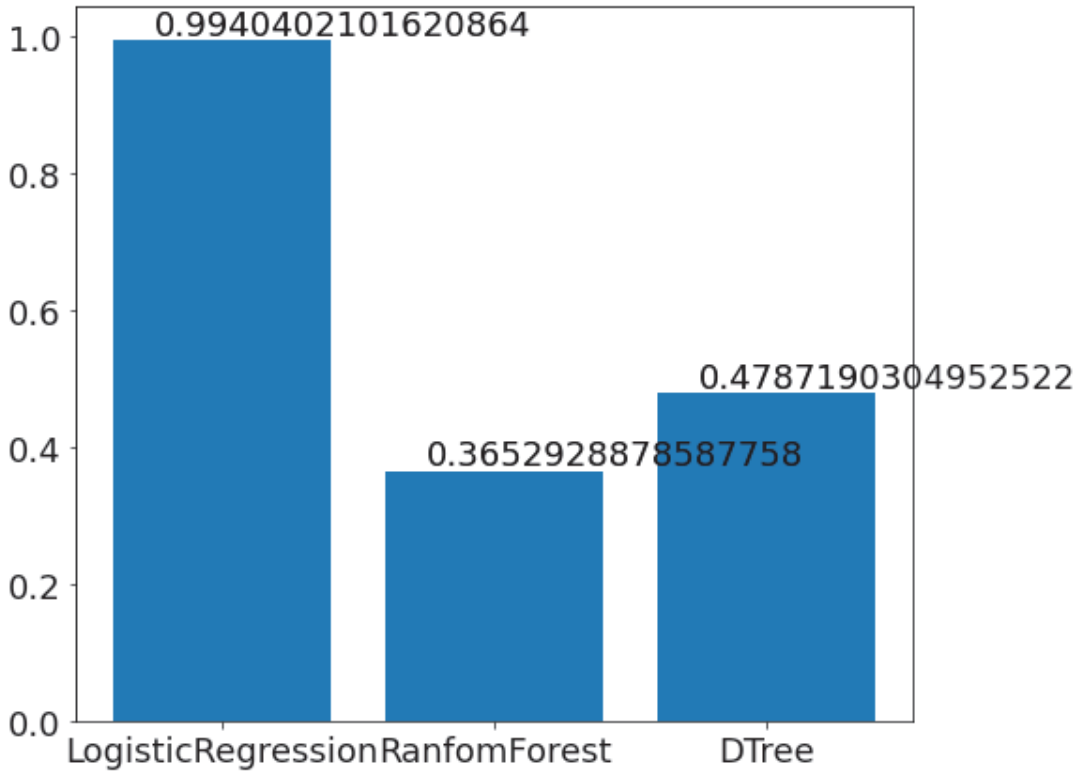
```
Model R^2: 0.4787190304952522
Mean Absolute Error: 5245.277777777777
Mean Squared Error: 173207572.5
Root Mean Squared Error: 13160.834794951268
```

[ ]:

[ ]:

28 Accuracy score graph

[93]:
```
plt.figure(figsize= (8,7))
ac = [ac_lr,ac_rf,ac_dt]
name = ['LogisticRegression','RanfomForest','DTree']
plt.bar(name,ac)

xlocs, xlabs = plt.xticks()
for i, v in enumerate(ac):
    plt.text(xlocs[i] - 0.25, v + 0.01, str(v))
```


0.9940402101620864
0.4787190304952522
0.3652928878587758
1.0
0.8
0.6
0.4
0.2
0.0
LogisticRegression
RanfomForest
DTree

CONCLUSION

The main aim of this paper was to analyze the road accidents in India at National, State and Zonal city level. Although, road accidents are relatively higher in extreme weather and during working hours, it was concluded from this study that it is not the condition of the roads or weather that causes most of the accidents but the unhealthy and unaware behaviour of people while driving that causes most of the accidents.

Thus we have also cleared the common misconception we had about weather and unruly/winding roads contributing to most number of accidents. Without increased efforts and new initiatives, the total number of road traffic deaths in India is likely to cross the mark of 250,000 by the year 2025. There is thus an urgent need to recognize the worsening situation in road deaths and injuries and to take appropriate action like the existing and new safety rules and measures should be effectively passed and implemented.

We have also concluded from this study that the number of road accidents in Kutcha roads and good roads is almost similar also the no of road accidents that is recorded during the weather condition of clear sky is very high, this conclusion strengths our objective that it is not road conditions and weather conditions that is determining the rate of accidents.

FUTURE WORK

In future the project can be further extended to analyse the same situations globally to have a possible conclusion by collecting global datasets and using new methodologies as follows

- Clustering
- Supervised learning
- Neural Network

It can also be utilised for globally comparing the datasets and take precautive measures from a board perspective

REFERENCES

1. https://www.analyticsvidhya.com/blog/2017/09/common-machine-learning-algorithms/?#
2. Road Accidents in India (2014-2017) | Kaggle
3. Over 1.51 lakh died in road accidents last year; UP tops among states | India News - Times of India (indiatimes.com)
4. https://data.gov.in/
5. katreparitosh/Multi-Dimentional-Data-Analytics-of-Road-Accidents-in-India: Multi-dimensional Analytics Project on Road Accidents of India. (github.com)
6. https://economictimes.indiatimes.com/industry/transportation/roadways/indian-roads-fatalities-in-mishaps-high-despite-better-construction-and-use-of-tech/articleshow/68443824.cms?from=mdr
7. https://reader.elsevier.com/reader/sd/pii/S2352146517307913?token=4ADB32DF3CEE92D8574B34132B92AB053E9FF54072488412F0EDF74F8C2D58C26454142A35C0019B572A0E79AB021BC6
8. https://www.business-standard.com/article/current-affairs/roads-accidents-led-to-three-deaths-every-minute-in-india-in-2017-report-118083000440_1.html
9. https://www.prsindia.org/policy/vital-stats/overview-road-accidents-india
10. https://www.financialexpress.com/india-news/over-70-percent-road-accidents-occurred-on-bright-sunny-days-report/1348638/
11. http://jhtransport.gov.in/causes-of-road-accidents.html