1.) Title of the project

Car Crash Analysis

2.) Group members with roll numbers

Aditya Satam MT2020058

Mohit Choudhary MT2020134

Shubham Ahlawat MT2020163

3.) Background of the problem

The accident datasets collected using two APIs that provide streaming traffic incident data.  
These APIs broadcast traffic data captured by a variety of entities, such as state departments of transportation, law enforcement agencies, traffic cameras and traffic sensors within the road-networks.  
Accidents can be used for various applications such as real-time car accident prediction, studying car accidents hotspot locations, casualty analysis and extracting cause and effect rules to predict car accidents.

4.) Problem statement

Challenges about future accident predictions and understanding accident patterns in formulating policies to guide or mitigate it.

5.) Objective(s)

a) To identify top States and top Cities with most accidents.  
b) To analyze accident trends over time series.  
c) To find Temporal data for most of the accidents.  
d) To analyze accidents in streets over its severity level.  
e) To Spatial analyze accidents on the map.  
f) To Predict total no. of accidents in different states for future dates   
 using different supervised learning techniques.

6.) Research questions

Is there a change in accident severity levels over the years?  
Do Weather Conditions have effect on Accidents?

7.) Data to be used

US\_Accidents.csv file. This file contains details of 1 million traffic accidents that took place in the US, from February 2016 to Dec 2019.

8.) Possible methods to be adopted

Exploratory Data Analysis, Understanding the Data, Cleaning the Data, Relationship Analysis, Modeling the Data via Decision Tree Regression Model & Random Forest Regression Model.

9.) Expected results

a) State: California (CA) has most no. of accidents & State: North Carolina (NC) has least no. of accidents.  
b) City: Houston has most no. of accidents & City: Nashville has least no. of accidents.  
c) Year: 2015 has more accidents as compared to year: 2017.  
d) There are more accidents in month of June & least accidents in month of Feb & May.  
e) Thursday witness more accidents as compared to weekend days: Saturday & Sunday.  
f) Severity 2 has more accidents as compared to Severity 1 & 4.  
g) Cloudy has most accidents and Rainy Situation has least accidents.  
h) Street: "FDR Dr N" has most accidents but with least severity and "George Washington Brg" has least accidents with highest severity.  
i) Prediction of total no. of accidents for future dates with score 70% using Decision Tree method.  
j) Prediction of total no. of accidents for future dates with score 98% using Random Forest method.

10.) Keywords

Exploratory Data Analysis, Decision Tree Regression Model, Random Forest Regression Model.  
<https://www.kaggle.com/sobhanmoosavi/us-accidents>