Introduction / Problem Description :

In the United States and throughout much of the world, car accidents are a leading cause of serious injury and death. In fact, in the U.S. alone, at least 38,800 people were killed in motor vehicle collisions in 2019. One of the worst affected city from these motor vehicle accidents in USA is Seattle.

Seattle is s a seaport city on the West Coast of the United States. It is the largest city in the state of Washington, as well as the largest in the Pacific Northwest. As of the latest census, there were 713,700 people living in Seattle. Seattle residents get around by car, trolley, streetcar, public bus, bicycle, on foot, and by rail. With such bustling streets, it's no surprise that Seattle sees car accidents every day.

The results of these accidents are mostly severe. People loose their live or suffer from serious injuries which remain there for very long, sometimes. These people who survive traffic accidents can face hefty medical bills, thousands of dollars in lost wages and property damage, pain and suffering, and lost quality of life. A collision could cause permanent disability, stripping the victim of the life he or she might have had.

Car accidents happen for a host of reasons, including behavioral, environmental, and situational. The behavioral reasons inlcude the carelessness from the driver's side like - Drug and Alcohol impairment, Over Speeding, Driver's Distraction, etc. There can be some environmental reasons as well like- bad lightning conditions, rainy and windy weather. Vehicle sudden malfunctioning can be another reason for accidents.

A small number of car accidents are inevitable and can't be prevented. Most of them, however, could at least be prevented, and many result from poor decisions by someone who should have done better.

It would be great if real-time road conditions and alerts can be provided to estimate the trip safeness. In this way, it can be decided beforehand if the driver will take the risk, based on reliable information. For example, if the driver is going towards a route where there is windy and rainy weather, and low lightning condition, and he gets a prior notification about it, he can try to be more alert whle driving on that route, or even try to change the route, if feasible.

Through this project, I will be trying to assess the severity and different reasons for these accidents and will try to figure out the ways in which these figures can be brought down significantly.

The target audience of the project, who will be most benefitted by the outcome of this project are - Seattle people, local Seattle government, police, rescue groups, and the insurance companies as well.

Data Description :

Data, if enquired propely, can give us a lot of insights. The dataset that we will be working to analyze the accident data is provided by Seattle Police Department(SPD) and recorded by Traffic Records. The dataset contains all types of collisions between 2004 to Present . This dataset is updated weekly and contains 194673 rows and 38 columns.

Out of these 38 columns, first column is 'SEVERITYCODE', which is our dependent variable. Rest of the 37 columns are the attributes. All the attibutes are not important for our model, and so from these 37 attributes, we will choose some of the important attributes for our model. They consist of :

- **ADDRTYPE** : Collision address type: Alley, Block, or Intersection
- **LOCATION** : Description of the general location of the collision.
- **PERSONCOUNT** : The total number of people involved in the collision.
- **PEDCOUNT** : The number of pedestrians involved in the collision.
- **PEDCYLCOUNT** : The number of bicycles involved in the collision.
- **VEHCOUNT** : The number of vehicles involved in the collision.
- **INJURIES :** The number of total injuries in the collision.
- **INCDTTM** : The date and time of the incident.
- **FATALITIES :** The number of fatalities in the collision.
- **JUNCTIONTYPE** : Category of junction at which collision took place.
- **UNDERINFL** : Whether or not a driver involved was under the influence of drugs or alcohol.
- **WEATHER** : A description of the weather conditions during the time of the collision.
- **ROADCOND** : The condition of the road during the collision.
- **LIGHTCOND** : The light conditions during the collision.
- **SPEEDING** : Whether or not speeding was a factor in the collision (Y/N).
- **SEGLANEKEY** : A key for the lane segment in which the collision occurred.
- **CROSSWALKKEY** : A key for the crosswalk at which the collision occurred.
- **HITPARKEDCAR** : Whether or not the collision involved hitting a parked car.

We will further process the dataset in order to make it apt for our use.

The dataset we will be using can be downloaded from the link provided by Coursera.