Introduction:

The purpose of this project is to warn user/people given the weather and the road conditions about the possibility of user/people getting into a car accident and how severe it would be, so that user/people would drive more carefully or even change travel if possible. It is an effort to reduce the frequency of car collisions which can save lives and damages happen to the common people due to car accidents.

Data understanding and preparation:

For this project we will be using a dataset available under below location.

https://s3.us.cloud-object-storage.appdomain.cloud/cf-courses-data/CognitiveClass/DP0701EN/version-2/Data-Collisions.csv

We will be using 'SEVERITYCODE' as our predictor or target variable because it is used measure the severity of an accident from 0 to 5 within the dataset. The main attributes used to weigh the severity of an accident from the given dataset are 'WEATHER', 'ROADCOND' and 'LIGHTCOND' as these attributes can be contributing factors to accidents. As these attributes type is “object” need to convert or create new attributes out of them with type as “int”, we will us label encoding to convert attributes in the desired type.

```
WEATHER          object
ROADCOND         object
LIGHTCOND        object
SEVERITYCODE      int64
WEATHER_le        int32
ROADCOND_le       int32
LIGHTCOND_le      int32
dtype: object
```

Balancing the Dataset :

Also the given dataset in not balanced, to balance it we can either down-sample majority class or up-sample minority class in the dataset, we will down-sample the dataset based on the SEVERITYCODE target column.

```
2    58188
1    58188
Name: SEVERITYCODE, dtype: int64
```

Below is the head of the converted dataset.

	WEATHER	ROADCOND	LIGHTCOND	SEVERITYCODE	WEATHER_le	ROADCOND_le	LIGHTCOND_le
0	Raining	Wet	Dark - Street Lights On	1	6	8	2
1	Clear	Dry	Daylight	1	1	0	5
2	Unknown	Unknown	Unknown	1	10	7	8
3	Clear	Dry	Daylight	1	1	0	5
4	Clear	Dry	Daylight	1	1	0	5