Prediction of Road Accident Severity in Rainy Weather



Image source: https://unsplash.com/photos/p3Ip8U0eNNM

# Introduction

The goal of this project is to predict the severity of road accidents in rainy weather due to various conditions.

Why is rainy weather selected for this analysis?

Road accidents are a common occurrence, and wet roads along with poor visibility increase the risk further during the rainyseason. Lack of awareness is a major cause of accidents.

Benefits of this project:

* This will help the community to stay safe and avoid damage or loss due to accidents in rainy weather.
* This will also help the Road Safety team to take necessary precautions.

# Data Understanding

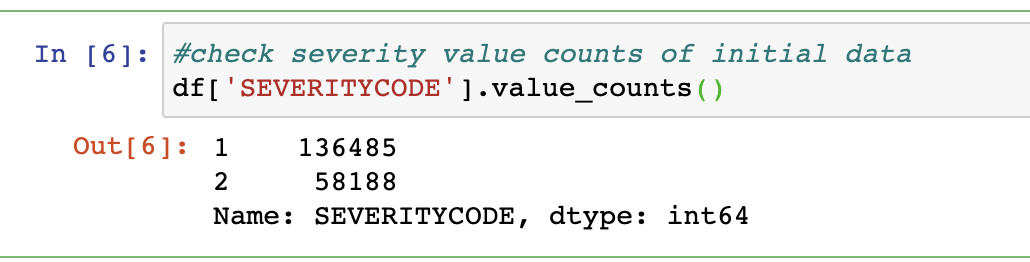
Data used in this project is provided by Coursera-IBM Data Science course. This is raw data which requires analysis and transformation to apply machine learning models.

By analysing the raw data, the following conclusion can be done.

1. Severity code is the target parameter or predictor variable which as it shows the severity of the accidents.
2. Data clean up is required as few columns are not required for analysis.
3. Since the goal is to get predictions for rainy weather, only rows corresponding to ‘Rainy’ Weather can be used.
4. ROADCOND and LIGHTCOND are different categories that can be derived from the Weather column.
5. Convert raw unbalanced data to balanced dataset.

## Severity code

For Rainy weather, the SEVERITY CODE is either 1 or 2, where 1 indicates it is Safe to travel and 2 indicates damage to life or property. This can be used as a Target variable to derive a solution.



## Data Clean Up

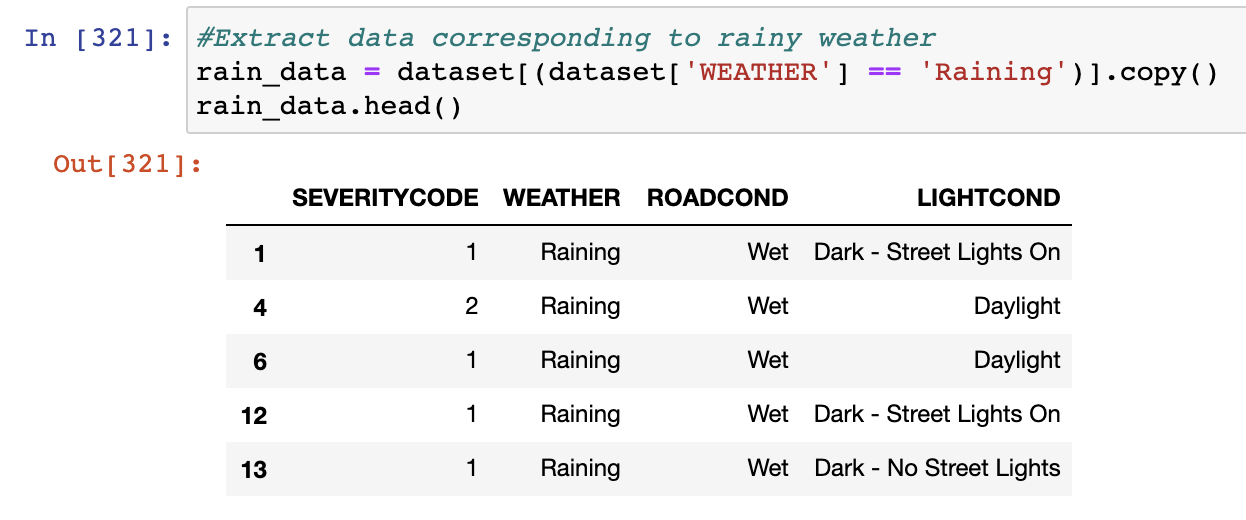
Post extracting csv data to the data frame, a clean up is required to remove unwanted data.

Columns excluding SEVERITY CODE, WEATHER, ROADCOND, LIGHTCOND can be removed. This creates a clean data set with only required columns.



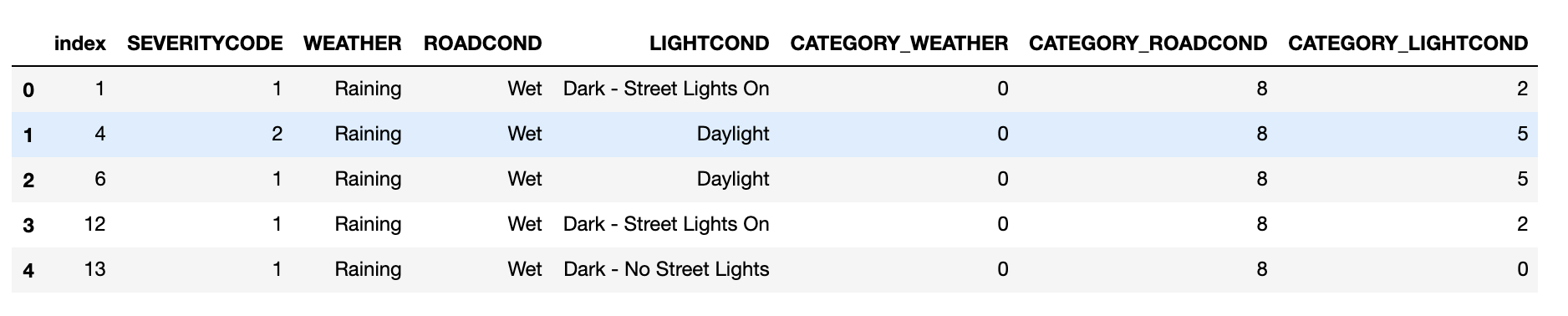
## Extract Required Data

Since analysis is based on Rainy weather, rows including other weather conditions can be removed. This creates a clean data set with only required rows.



## Categorise data

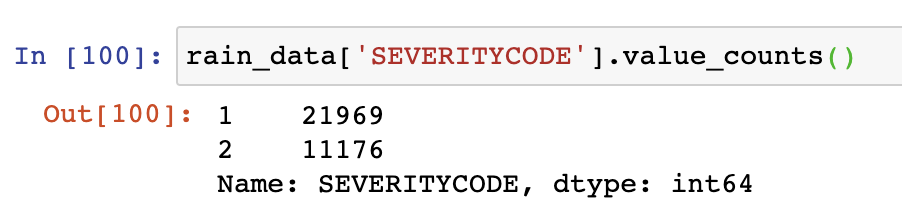
ROADCOND and LIGHTCOND are two columns which impact target variables along with Weather data. One of the major reasons why we convert categorical variables into factors i.e number because to make Analysis easy and effective.

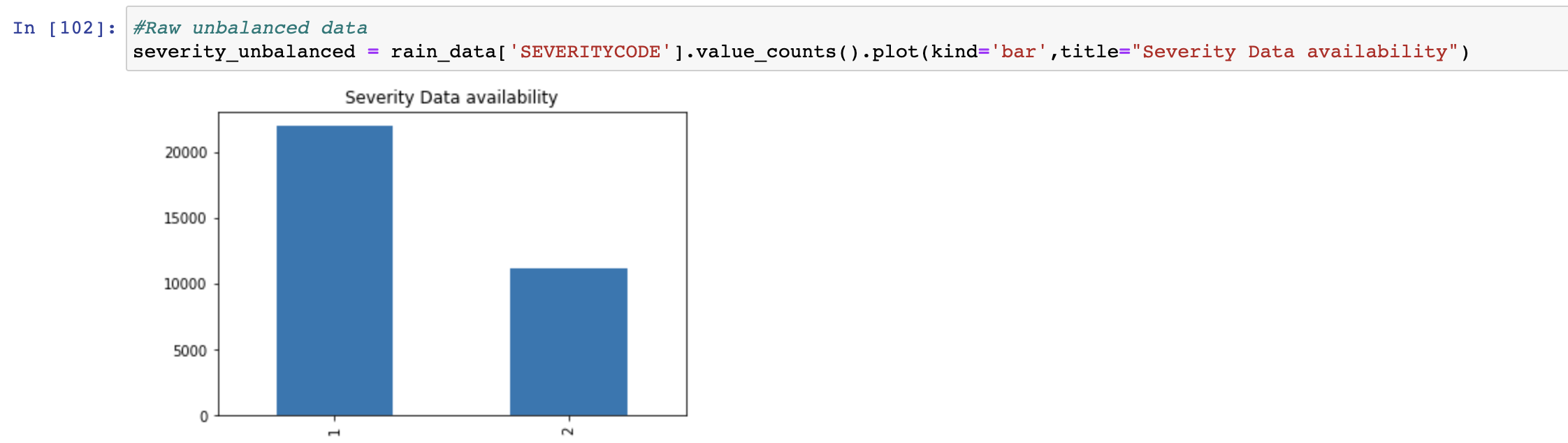


## Convert Unbalanced data to balanced dataset

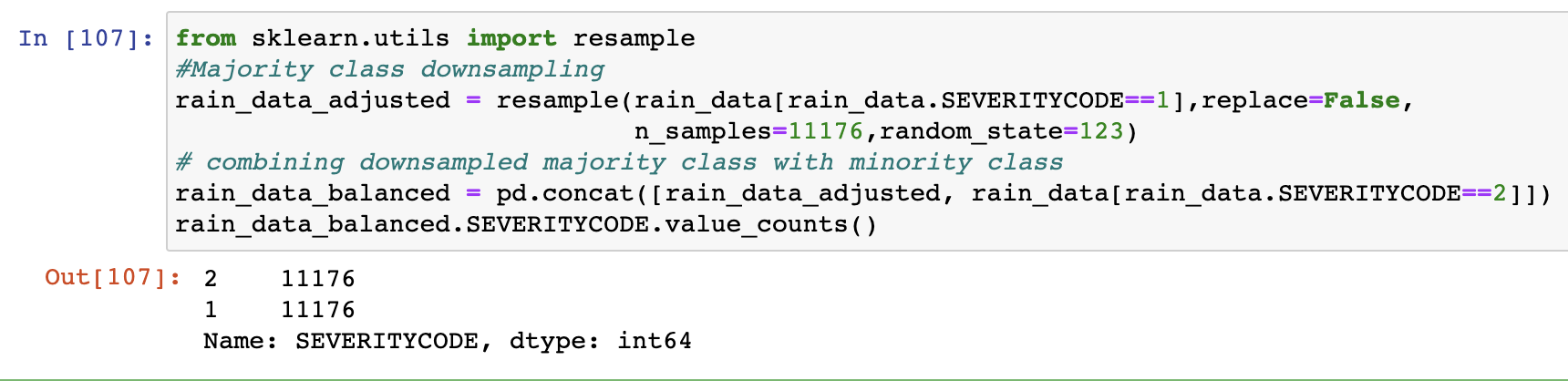
Unbalanced data refers to classification problems where we have unequal instances for different classes. Having unbalanced data is actually very common in general. Downsampling method is followed to achieve this.The main goal of downsampling (and upsampling) is to increase the discriminative power between the two classes.

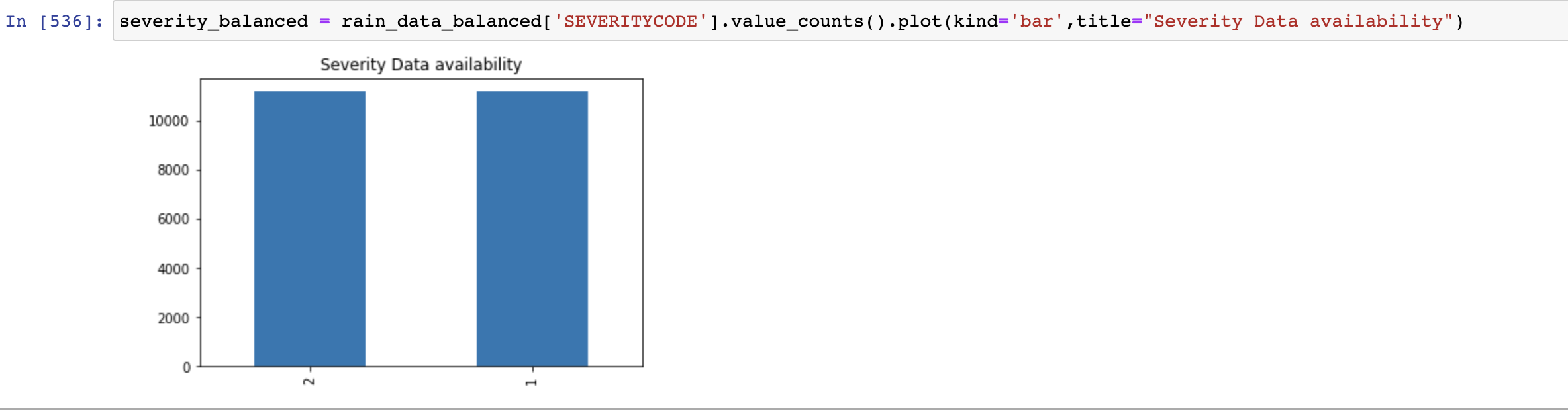
Here is a plot of Unbalanced data:





## Downsampling step





Following machine learning models are used for further analysis:

1. K Nearest Neighbor (KNN)
2. Decision Tree
3. Logical Regression