**Libraries used:**

from google.colab import drive

import pandas as pd

import numpy as np

import matplotlib.pyplot as plt

import seaborn as sns

from sklearn.preprocessing import StandardScaler

from sklearn.model\_selection import train\_test\_split

from sklearn.linear\_model import LogisticRegression

from sklearn.ensemble import RandomForestClassifier

from sklearn.metrics import accuracy\_score, recall\_score, classification\_report

from datetime import \*

from sklearn.preprocessing import LabelEncoder, OneHotEncoder

from sklearn import preprocessing

from sklearn.metrics import classification\_report

I have used Google Colab to solve this data.

Data contain 1000 rows \* 39 columns.

38 features and 1 target variable.

Missing value are represented by “?” and police\_report, collision\_type & property\_damage have missing values.

My approach includes visualisation of data after visualisation pre-processing and then training model.

Made few new features such as vehicle\_age from the date it was registered.

Also made age group from age of insured and month from date.

Missing values in collision type are only 20% of total so I decided to delete missing value rows.

Missing values in police\_report and property\_damage are more than 1/3rd imputing values using mode seems not suitable. I decided to treat it as new category because of time permit.

Another way could be imputing value after predicting the missing value using available data.

After dealing with missing data I did lablel encoding of categorical and also I did apply standardization to get values in range.