**Name- Abhinav Kulshrestha**

**Student ID- 300337257**

### Predicting The Insurance for a Car

Using Machine learning algorithms

# Introduction and Discovery

* In this project we are using a Car automobile dataset which has various information about the car as follows:

This data set consists of three types of entities: (a) the specification of an auto in terms of various characteristics, (b) its assigned insurance risk rating, (c) its normalized losses in use as compared to other cars. The second rating corresponds to the degree to which the auto is riskier than its price indicates. Cars are initially assigned a risk factor symbol associated with its price. Then, if it is riskier (or less), this symbol is adjusted by moving it up (or down) the scale. Actuaries call this process "symboling". A value of +3 indicates that the auto is risky, -3 that it is probably safe.

## Data Preparation

The dataset of dry beans was acquired from the UCI Machine Learning Repository  
Link: <https://archive.ics.uci.edu/ml/datasets/automobile>

The dataset initially didn’t have any column names specifically given but were allotted afterwards

Like symbolling, wheelbase, width, height etc.

All the null valued rows were dropped like engine size mileage etc. which aren’t risk factors for the insurance.

Graphical user interface

Description automatically generated

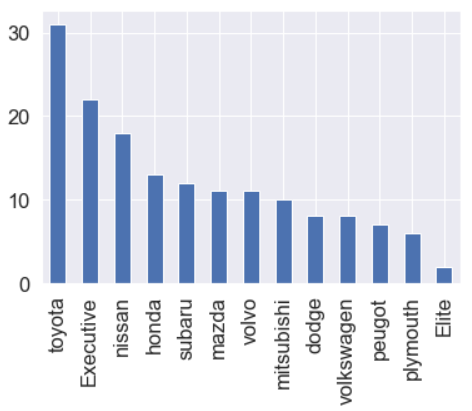
We Also changed the datatypes of various columns that were numerical

Some interesting plots:

Chart, histogram

Description automatically generated

Most cars are lying under the width bracket of 63 to 67 inches



Most cars present are of Toyota make followed by executive cars.

After Going through the correlation of all the features we also observed the following heatmap of correlation

A picture containing chart

Description automatically generated

Chart, scatter chart

Description automatically generated

The weight of the Car is steadily increasing with the price with certain outreached data points

Chart, bar chart

Description automatically generated

The car’s engine mostly of the type OHC. Pushrod engines are required of high maintenance which may effect the insurance.

## Feature Selections

From the given heatmap and applying different feature selection procedures we observed that there were 4 features that were important to predict the car’s Price that were with engine-size, width, length and curb weight of the car.

Chart, line chart

Description automatically generated

Chart, line chart

Description automatically generated

* 1. **Model Planning**

In the data frame, we had to create Dummy values first for the categorical columns.

3 different feature models were created using following selection techniques:

**Manual Selection**: In this selection, we selected those 5 features which are having highest correlation with the target.

**Variance Threshold Selection**: Using this selection method, we selected those features which were having variance of more than 0.6.

**Select K Best**: Using this selection method, we selected top 10 best features out of all.

After performing linear regression with all those features we figured out the best model that fitted for the lowest r2 Value and highest RMSE value :

A picture containing shape

Description automatically generated

## Conclusion:

From the above regression models and outcomes, we concur the following

1. The insurance of the Car highly depends on its weight, Size and engine. So they can be the major factors for determining the insurance
2. symbolling of the car as mentioned above is not a guarantee for a reliable car.
3. The larger the size of the car will result in increase of its insurance price
4. Major risk factors that should be included while calculating the car insurance should be its Engine, size.