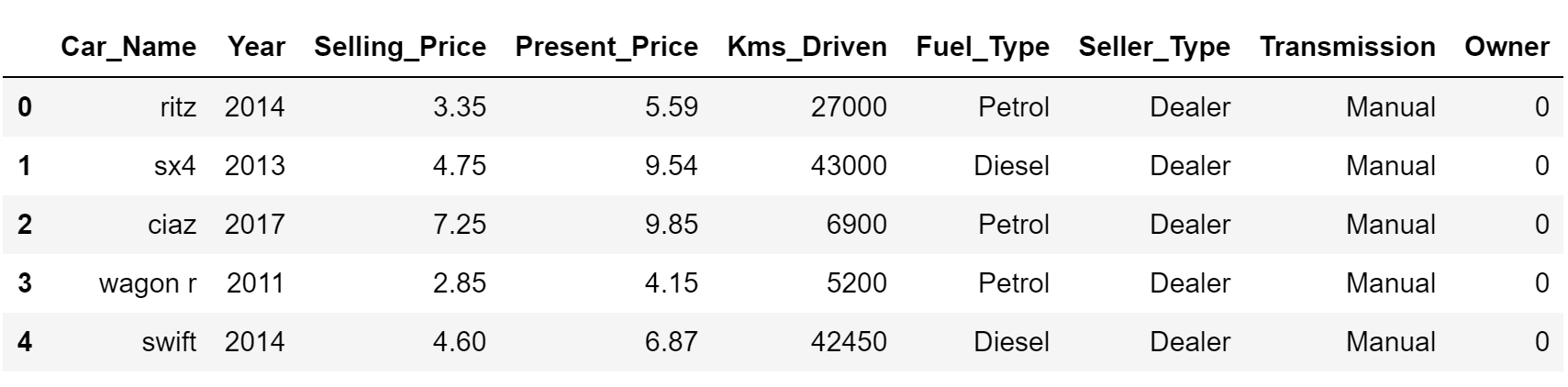
**Predicting Used Cars Prices**

**Introduction**

Predicting the price of a used car helps determine if a car is worth the price quoted by a dealer or an individual. Several factors, including mileage, make, model, year, etc. can influence the actual worth of a car. With the help of existing data our goal is to design a Machine learning model to predict the price of used cars. This application will thereby help sellers know at what price range the car can be sold, also buyers save money and buy cars for reasonable prices.

**Dataset and Pre- Processing**

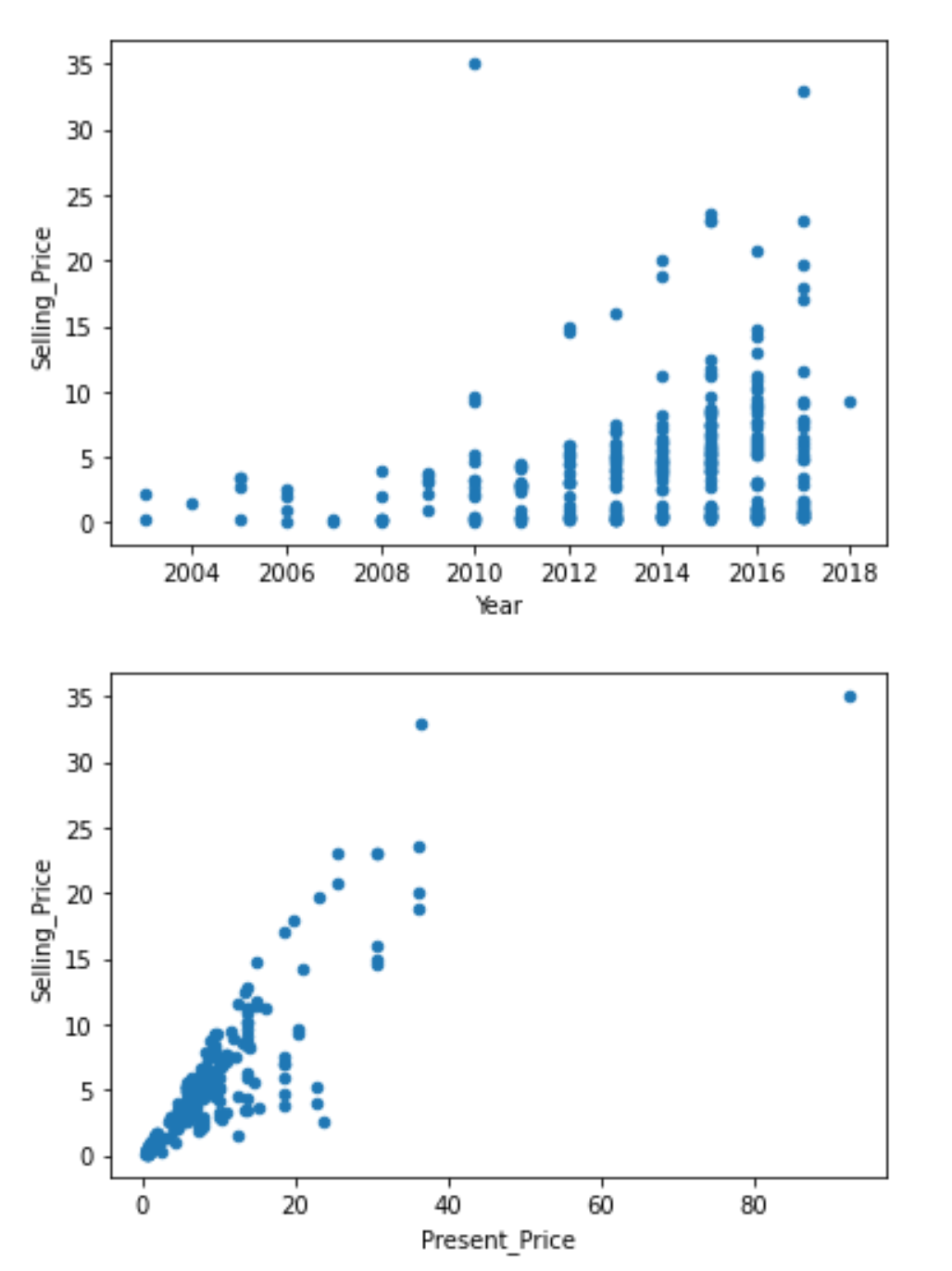
For this project, we are using the dataset on used car sales listed on [www.cardekho.com](http://www.cardekho.com/), available on Kaggle[1]. The reason behind the choice of dataset is because we were crazy about cars and dream to buy a muscle car in future. The features available in this dataset are Car\_Name, Year, Selling\_Price, Present\_Price, Kms\_Driven, Fuel\_Type, Seller\_Type, Transmission, Owner.



* Convert Categorical Columns to Numeric Features – The feature columns Fuel\_Type, Seller\_Type, Transmission have categorial values which where replaced with numerical values.
* Percentage change of the feature Kms\_Driven by dividing the column values by hundred. This helped us avoid NaN values as parameter output.

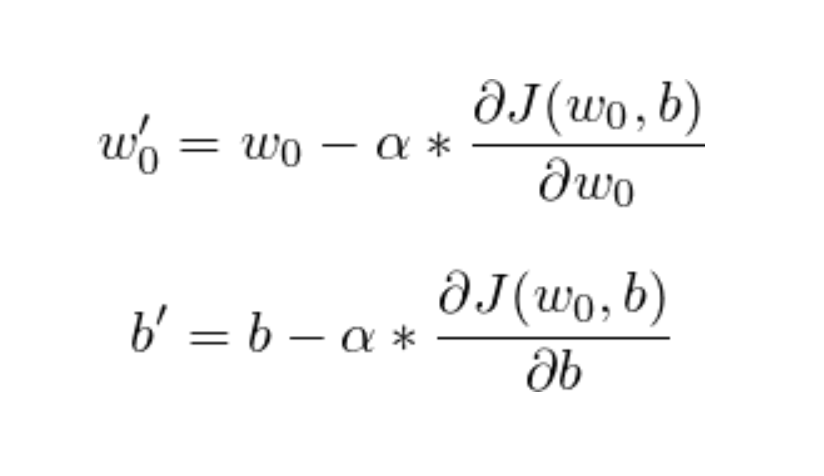
**Method**

In this project, [scatter plot](https://en.wikipedia.org/wiki/Scatter_plot) is used to display the relationship between target and training features. Here, a **linear dependency between features can be observed** — with the increase of values on axis x, values on the y-axis are linearly increasing or decreasing accordingly. It’s great because if that was not the case (e.g. relationship would be exponential), then it would be hard to fit a line through all the points and different algorithm should be considered.



Gradient descent is an optimization algorithm used to minimize some function by iteratively moving in the direction of steepest descent as defined by the negative of the gradient. In this project, we use gradient descent to update the parameters of our model.

To update model parameters, so the convergence is achieved, the following math has to be iteratively applied:



where:

w’ — new weight value,

b’ — new bias value,

w — current weight value,

b — current bias value,

α — learning rate,

dJ(w\_0, b)/dw\_0 — derivative of J with respect to w\_0,

dJ(w\_0, b)/db— derivative of J with respect to b.

**Difficulties faced during the project**

There were many difficulties faced during this project, but the hardest part was to find an appropriate dataset for the project. Choosing the Training rate for the model was one other difficulty faced by us which required us to do lot of trial and error before arriving at the highest possible accuracy.

The feature column Kms\_Driven had large value which lead to NAN values while calculating the parameters. To overcome this, we had to divide the feature value by hundred.

Result

|  |  |  |
| --- | --- | --- |
| Alpha | Iterations | Accuracy |
| .00000001 | 10000000 | 94.0120 |
| .0000001 | 10000000 | 99.7751 |
| .0000001 | 1000000 | 94.0120 |
| .0000001 | 100000 | 85.2915 |

**Future Work and References**

With the same dataset using Logistic regression we can also classify the cars as ‘good/bad’ or also rate the car on a scale of 1 to 10 based on the feature set.

1. <https://www.kaggle.com/nehalbirla/vehicle-dataset-from-cardekho>
2. <https://towardsdatascience.com/machine-learning-models-explained-to-a-five-year-old-f2f540d9dcea>