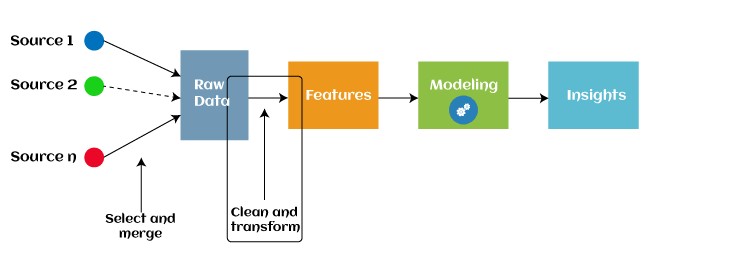
**FUTURE SALES PREDICTION – DEVOLPMENT PART 2**

**Feature Engineering :**

**→** Feature engineering refers to manipulation — addition, deletion, combination, mutation — of your data set to improve machine learning model training, leading to better performance and greater accuracy. Effective feature engineering is based on sound knowledge of the business problem and the available data sources.

**Feature Engineering Techniques :**

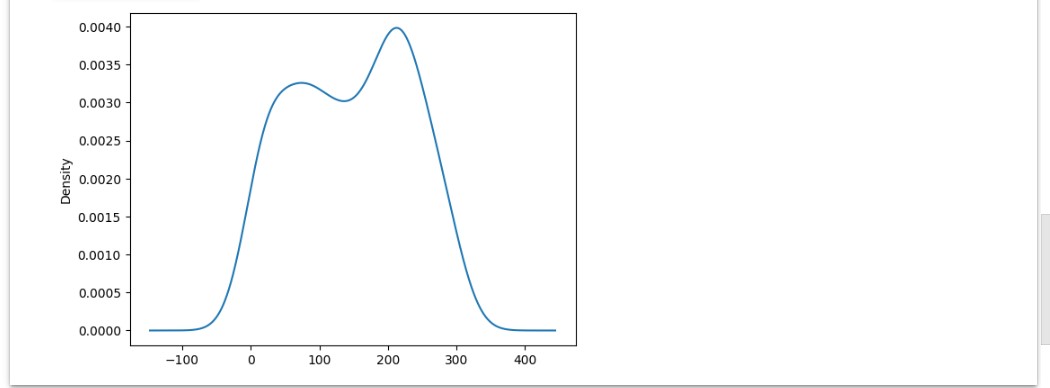
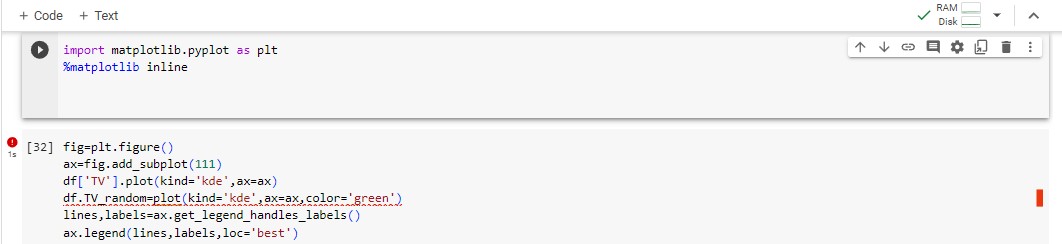
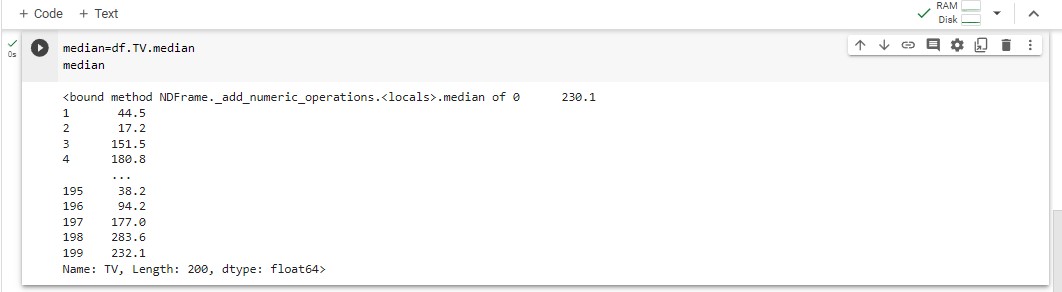
**→**Imputation **→**Handling Outliers

**→**Log transform

**→**Binning

**Imputation :**

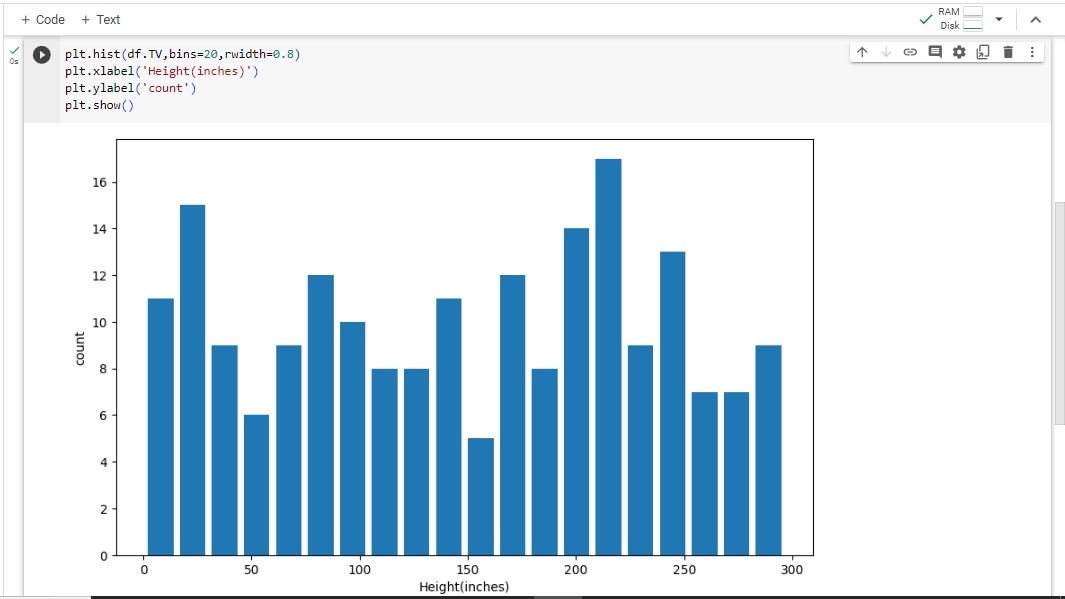
* Feature engineering deals with inappropriate data, missing values, human interruption, general errors, insufficient data sources, etc. Missing values within the dataset highly affect the performance of the algorithm, and to deal with them "Imputation" technique is used. Imputation is responsible for handling irregularities within the dataset.



**Handling Outliers :**

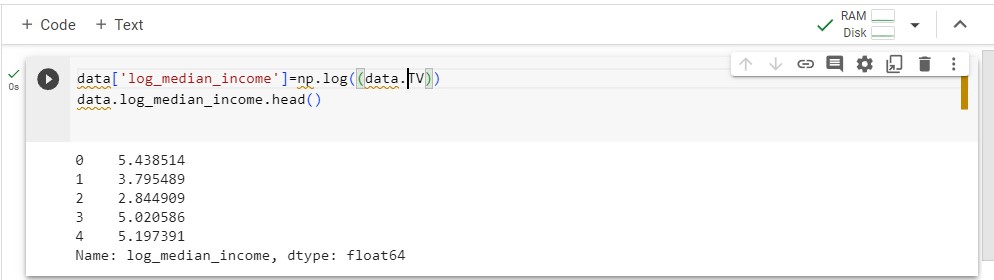
* Standard deviation can be used to identify the outliers. For example, each value within a space has a definite to an average distance, but if a value is greater distant than a certain value, it can be considered as an outlier. Zscore can also be used to detect outliers.



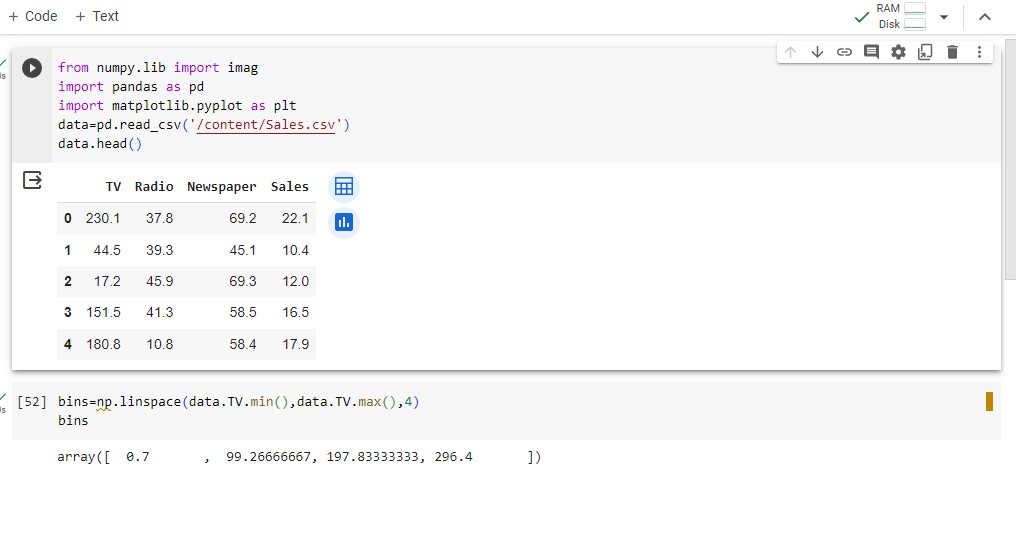


**Log transform :**

* Logarithm transformation or log transform is one of the commonly used mathematical techniques in machine learning. Log transform helps in handling the skewed data, and it makes the distribution more approximate to normal transformation. It also reduces the effects of outliers on the data, as because of the normalization of magnitude differences, a model becomes much robust.



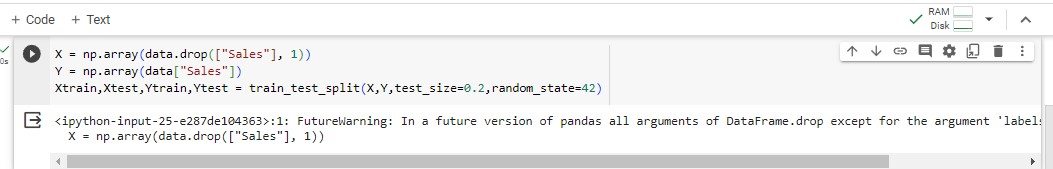
**Binning :**



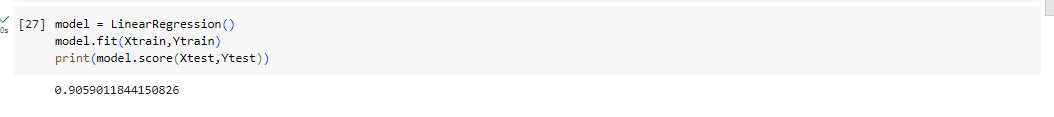
* In machine learning, overfitting is one of the main issues that degrade the performance of the model and which occurs due to a greater number of parameters and noisy data. However, one of the popular techniques of feature engineering, "binning", can be used to normalize the noisy data. This process involves segmenting different features into bins.

**Model Training :**

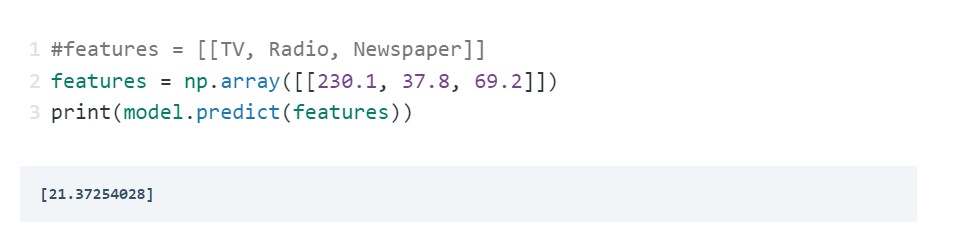
* Now in this section, I will train a machine learning model to predict the future sales of a product. But before I train the model, let’s split the data into training and test sets:



* Now let’s train the model to predict future sales:



* Now let’s input values into the model according to the features we have used to train it and predict how many units of the product can be sold based on the amount spent on its advertising on various platforms:

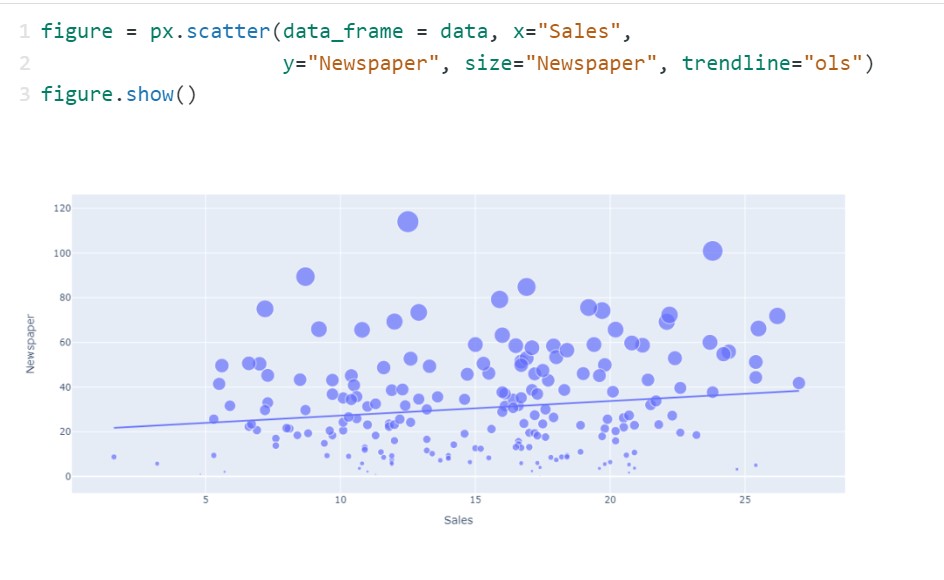


**Evaluation :**

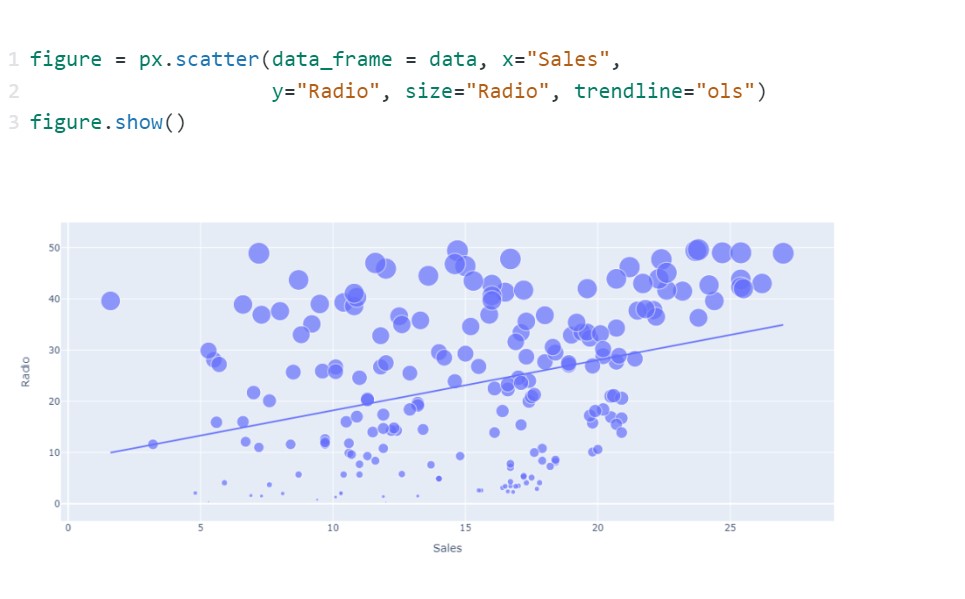
* + Now let’s visualize the relationship between the amount spent on advertising on TV and units sold:



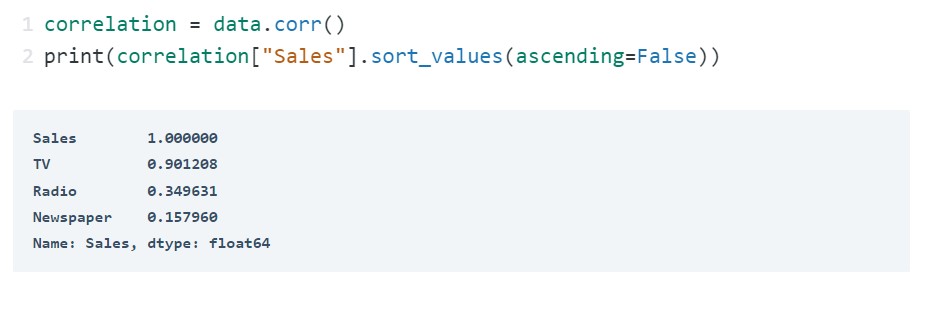
* + Now let’s visualize the relationship between the amount spent on advertising on newspapers and units sold:



* + Now let’s visualize the relationship between the amount spent on advertising on radio and units sold:



* + Out of all the amount spent on advertising on various platforms, I can see that the amount spent on advertising the product on TV results in more sales of the product. Now let’s have a look at the correlation of all the columns with the sales column:



**→** So this is how we can evaluate a machine learning model and perform the feature engineering for future sales prediction