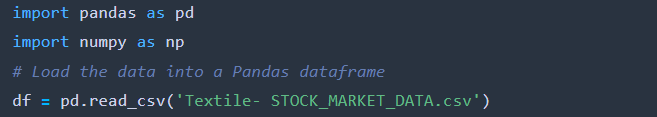
Name: RISHU VERMA Name: SIDDHARTH DOGRA  
Department: MBA in BABD Department: MBA IN BABD   
University: Thapar University University: Thapar University

**TEXTILE STOCK PRICE PREDICTION MODEL**

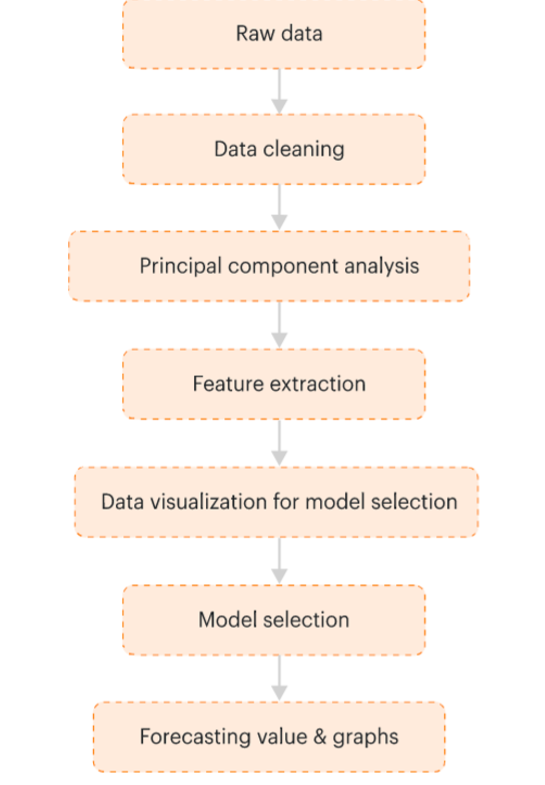
**APRIL 29,2023**

# Overview

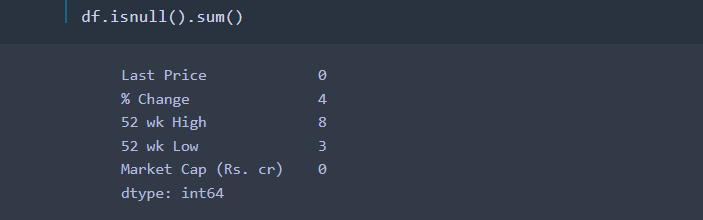
The dataset we are using here to train a textile stock price prediction model was downloaded from yahoo finance. It contains data about all the main features that contribute to the price of a stock . So let’s start this task by importing the necessary Python libraries and the dataset:



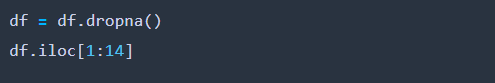


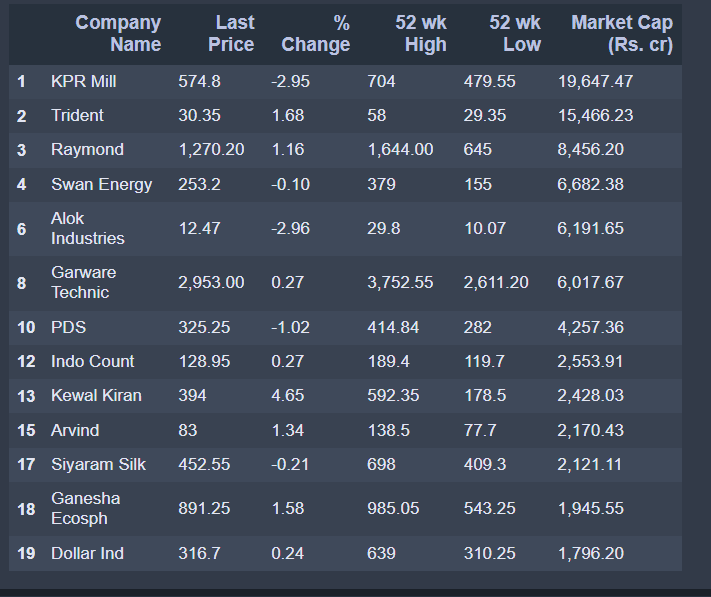


There are 6 columns and 99 rows in this dataset, so it is very important to check whether or not this dataset contains null values before going any further:

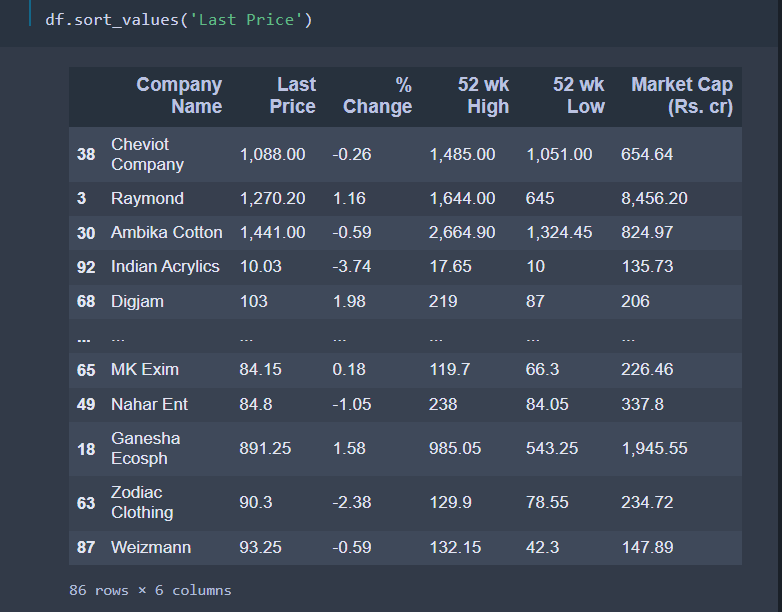


So this dataset does have a few null values, lets drop rows containing missing values and handle it:

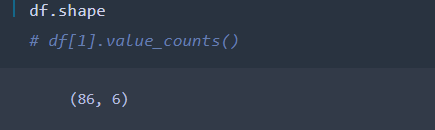
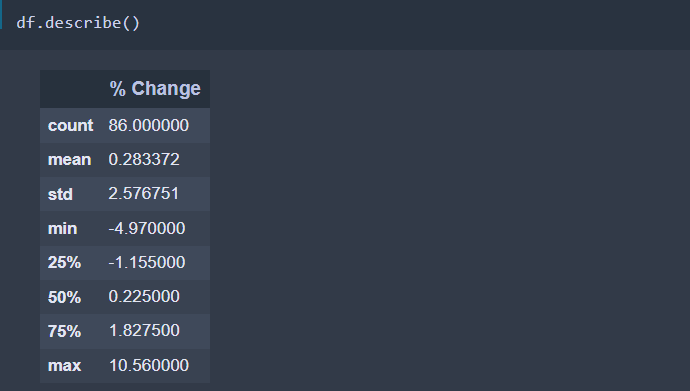




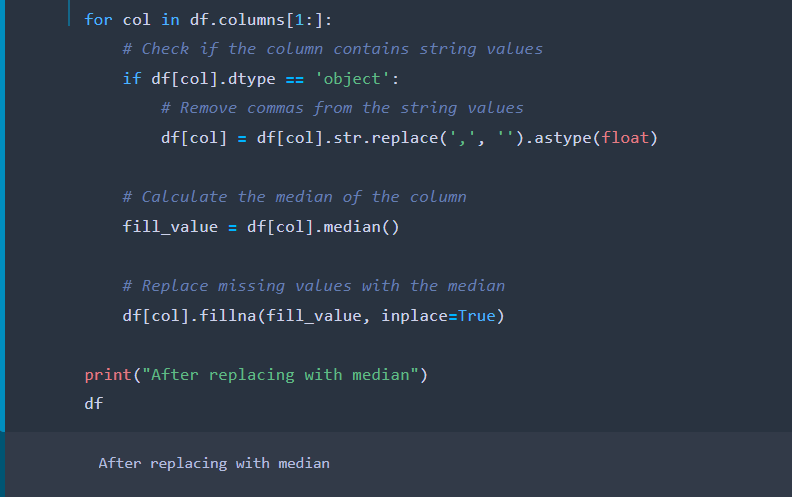
**SORTING ON BASIS OF LAST PRICE**



now let’s look at some of the other important insights to get an idea of what kind of data we’re dealing with:

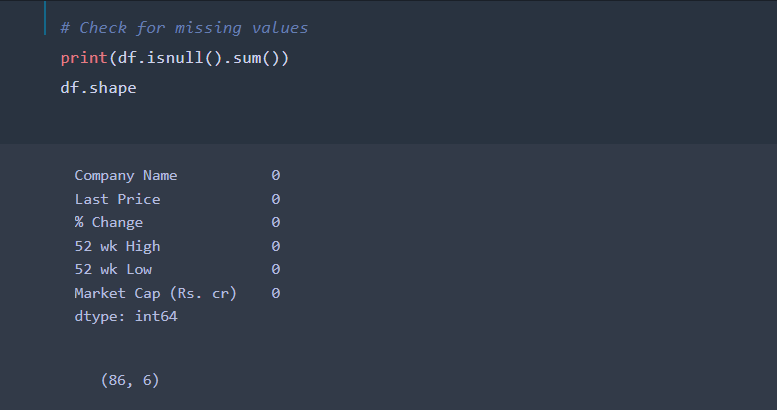
* 
* 

Now lets try to Replace Missing Values With Median:





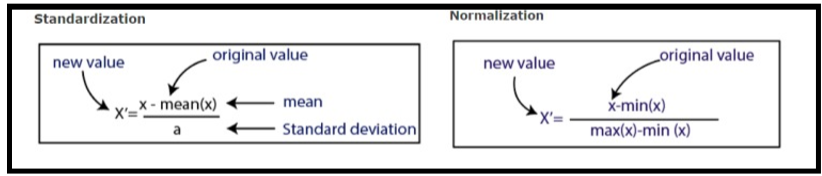
Now to see if there is any missing value left:

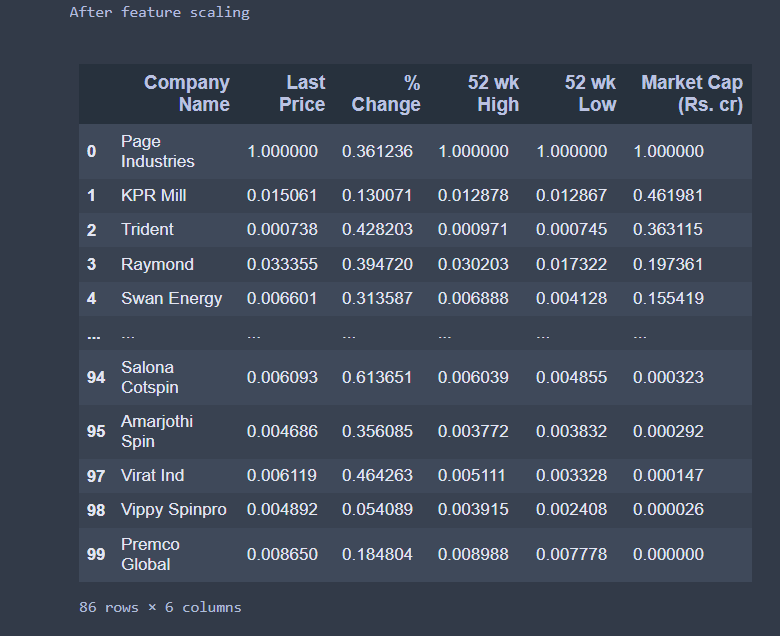
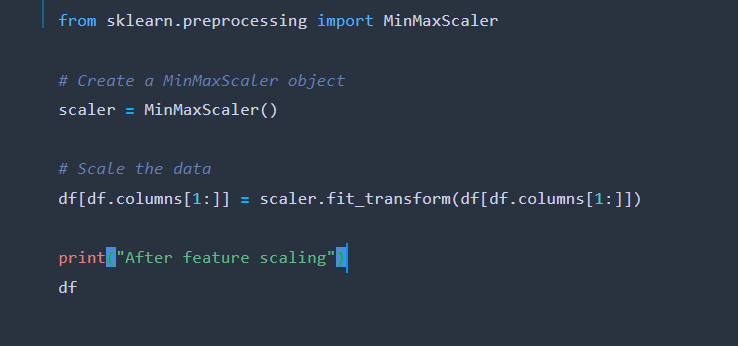


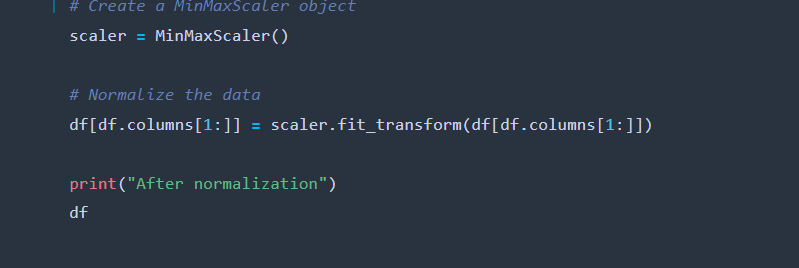
We can see no missing values left,all have been taken care off.

***Standard Scale or Normalization***: The variables in the data set maybe in different scale or units, so it becomes difficult for the machine algorithm to do regression and give optimized and accurate results. In order to overcome this

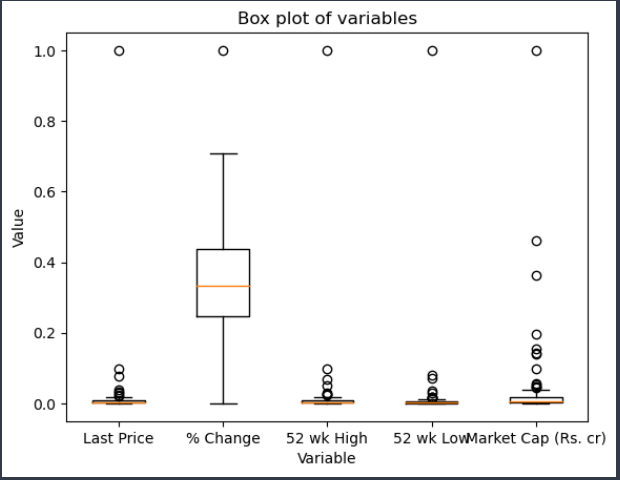
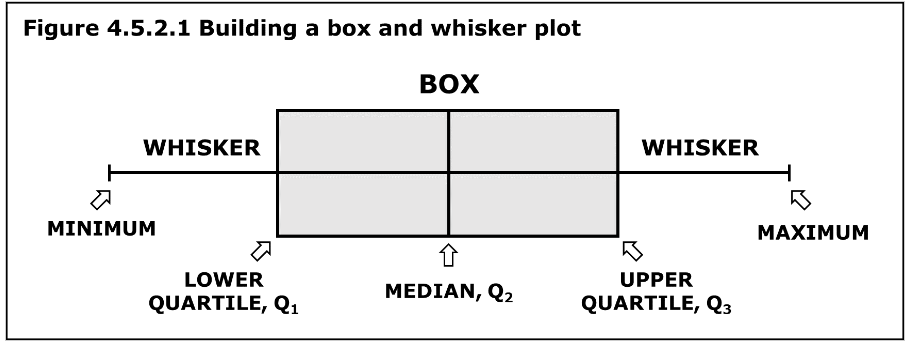
problem, a need arises to scale such data within a specific range. Hence, we either use standard scalar or normalization technique to bring such data within a specific range. This helps to ensure uniformity within a data.



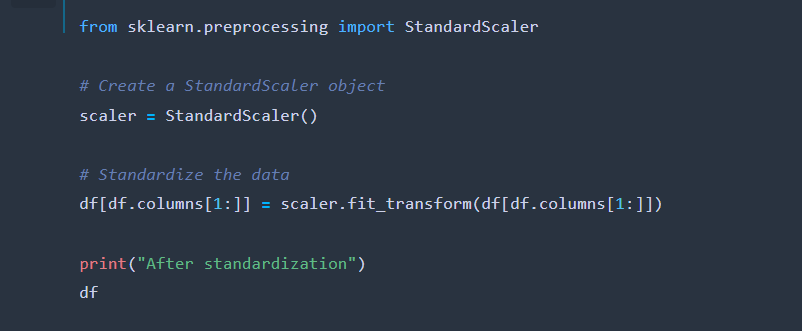




**DATA VISUALIZATION**

For visualization we had used matplotlib library.In this we take help of box plot as it is useful for indicating whether a distribution is skewed and whether there are potential unusual observations (outliers) in the dataset.

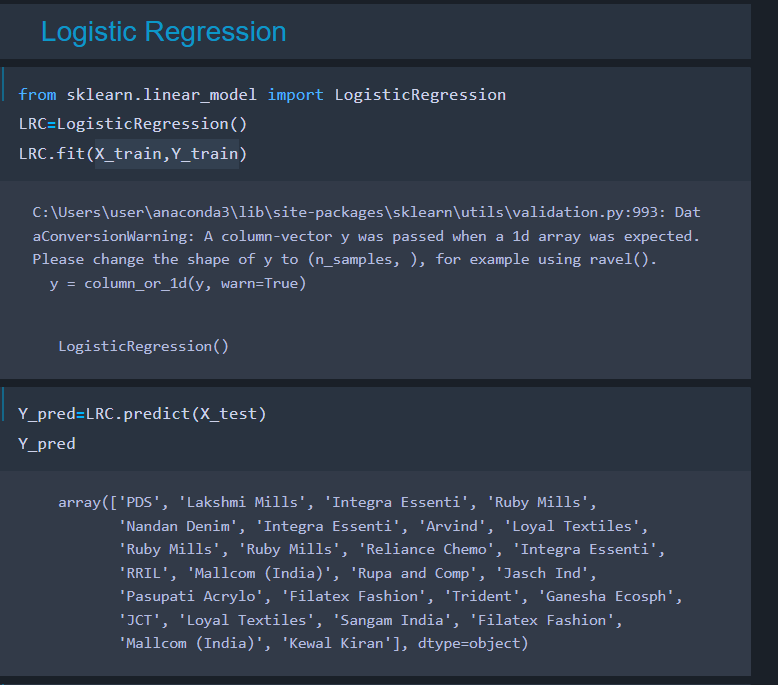
# **Standardization**

The result of **standardization** is that the features will be rescaled to ensure the mean and the standard deviation to be 0 and 1, respectively.The concept of standardization comes into the picture when continuous independent variables are measured at different scales.

The various machine learning algorithms used for the prediction stock price are Logestic regression, Multiple linear regression, KNN Regression, Decision tree regression, SVR (Support vector regression), Random Forest regression. Out of these models, the model which gives the most accurate prediction of the stock price is selected.

1. Logestic Regression: Logistic regression is another technique borrowed by machine learning from the field of statistics.

It is the go-to method for binary classification problems (problems with two class values). In this post you will discover the logistic regression algorithm for machine learning.

2. Decision Tree Regression: A decision tree is used to develop regression models and has a tree-like structure. It gradually divides a dataset into smaller and smaller subgroups depending on the information gain value for each unique feature while also developing an associated decision tree. Finally, we have a tree containing decision nodes and leaf nodeAfter doing this let’s make a heatmap using seaborn library.

A heatmap (or heat map) is a graphical representation of data where values are depicted by color.The variation in color may be by hue or intensity, giving obvious visual cues to the reader about how the phenomenon is clustered or varies over space. There are two fundamentally different categories of heat maps: the cluster heat map and the spatial heat map.

