**Standardization vs Normalization**

**Is feature scaling mandatory?** **when to use standardization?** **when to use normalization? what will happen to the distribution of the data? what is the effect on outliers? Will the accuracy of the model increase?**

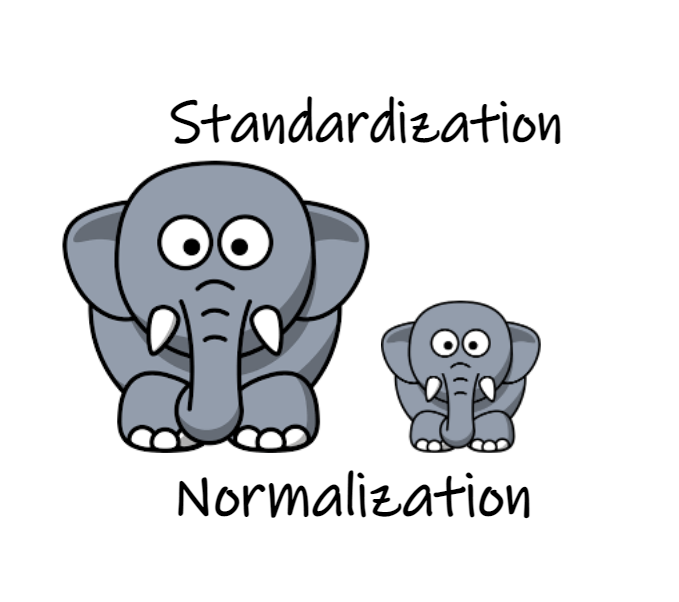


Image by author

**Table of Contents:**· [**What is Standardization?**](https://medium.com/@gowthamsr37/which-feature-scaling-technique-to-use-standardization-vs-normalization-9dcf8eafdf8c#d0af)∘ [Effect on the distribution of data:](https://medium.com/@gowthamsr37/which-feature-scaling-technique-to-use-standardization-vs-normalization-9dcf8eafdf8c#d752)  
 ∘ [Effect of Standardization on different Machine Learning algorithms:](https://medium.com/@gowthamsr37/which-feature-scaling-technique-to-use-standardization-vs-normalization-9dcf8eafdf8c#1800)  
 ∘ [Effect on Outliers:](https://medium.com/@gowthamsr37/which-feature-scaling-technique-to-use-standardization-vs-normalization-9dcf8eafdf8c#f17a) ·[**What is Normalization?**](https://medium.com/@gowthamsr37/which-feature-scaling-technique-to-use-standardization-vs-normalization-9dcf8eafdf8c#4bd9)  
 ∘ [Effect on Distribution of the data:](https://medium.com/@gowthamsr37/which-feature-scaling-technique-to-use-standardization-vs-normalization-9dcf8eafdf8c#4046)  
 ∘ [Effect of Normalization on different Machine Learning algorithms:](https://medium.com/@gowthamsr37/which-feature-scaling-technique-to-use-standardization-vs-normalization-9dcf8eafdf8c#ad5d)  
 ∘ [Effect on outliers:](https://medium.com/@gowthamsr37/which-feature-scaling-technique-to-use-standardization-vs-normalization-9dcf8eafdf8c#9729)**·** [**Observations:**](https://medium.com/@gowthamsr37/which-feature-scaling-technique-to-use-standardization-vs-normalization-9dcf8eafdf8c#34e7)

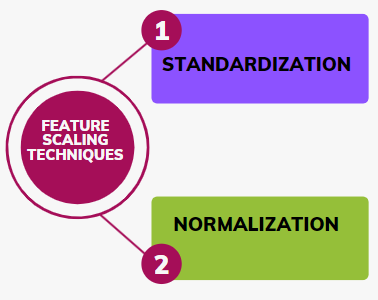
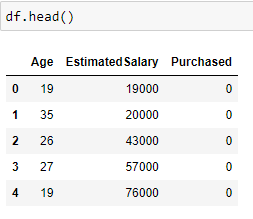


Image by author

*The above questions are frequently asked in interviews too, I will try to answer the above questions in this blog by providing suitable examples. We will use sklearn’s StandardScaler and MinMaxScaler.*

Let us consider a dataset in which **Age** and **Estimated Salary** are the *input features* and we have to predict if the product is **Purchased**(*output label*) or not purchased.

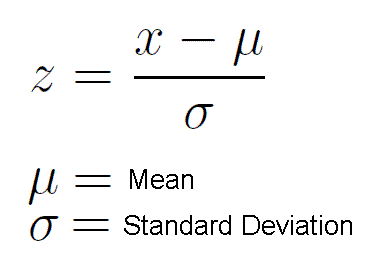
Take a look at the first 5 rows of our data.



The first 5 rows of the dataset

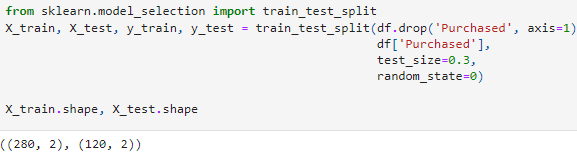
**What is Standardization?**

Standardizationor Z-Score Normalization is one of the feature scaling techniques, here the transformation of features is done by subtracting from the mean and dividing by standard deviation. This is often called Z-score normalization. The resulting data will have the **mean as 0 and the standard deviation as 1.**



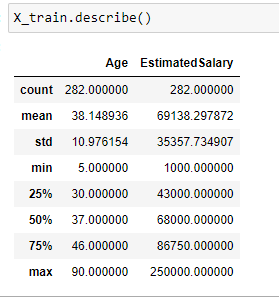
Formula to calculate Z score

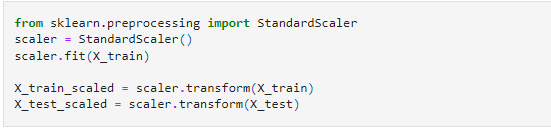
So now we have seen the formula of standard scaling, Now we shall look at how it can be applied to our dataset.

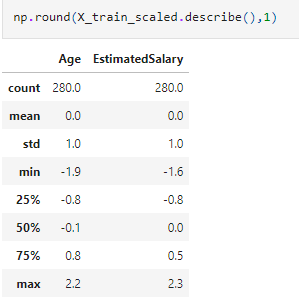


First, we shall divide our data into train and test sets and apply a standard scaler.

**Description of the dataset:**

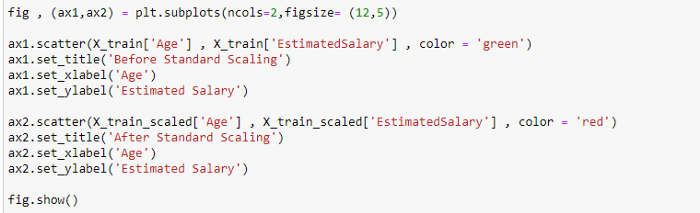


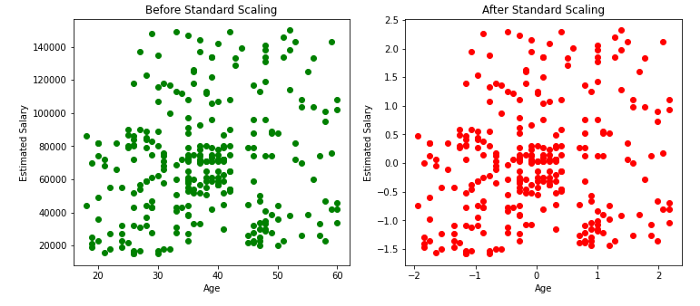




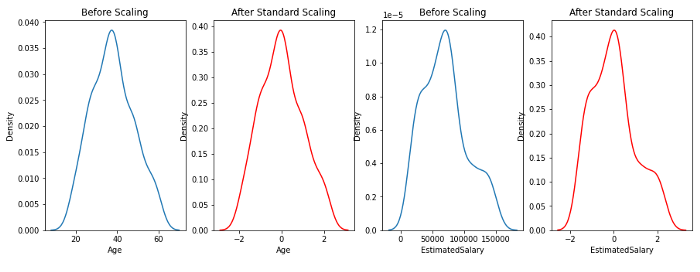
Note that the described method applied to X\_train\_scaled data shows that **the mean is 0 and the standard deviation is 1** after applying the standard scaler.

**Effect on the distribution of data:**





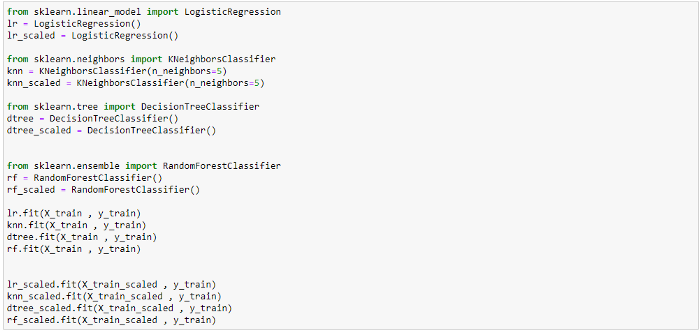
Effect of standard scaling on the distribution of data.



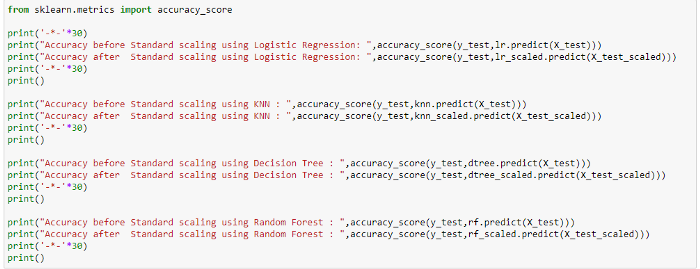
Effect of standard scaling on the distribution of data.

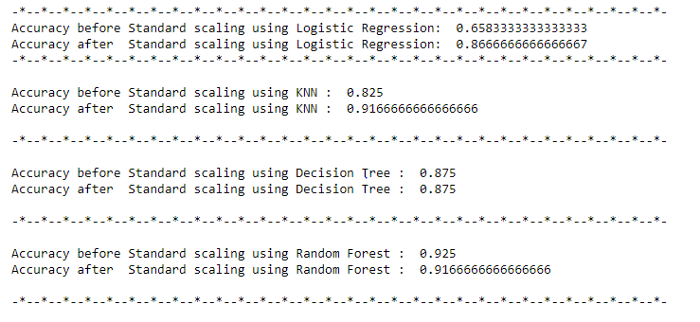
From the above scatter plots and KDE plots we can note that the distribution of the data remains the same even after applying standard Scaler, only the scale changes.

**Effect of Standardization on different Machine Learning algorithms:**



training various models





the behavior of different machine learning algorithms before and after scaling.

In the above examples, the accuracy of Logistic regression and KNN increased significantly after scaling. But there was no effect on accuracy when the decision tree or random forest was used.

**Effect on Outliers:**

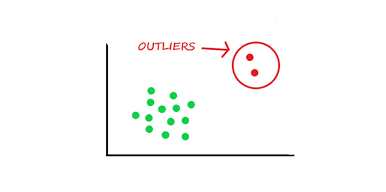
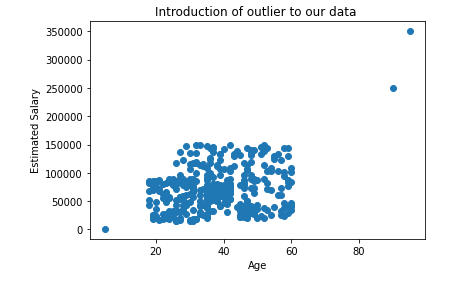
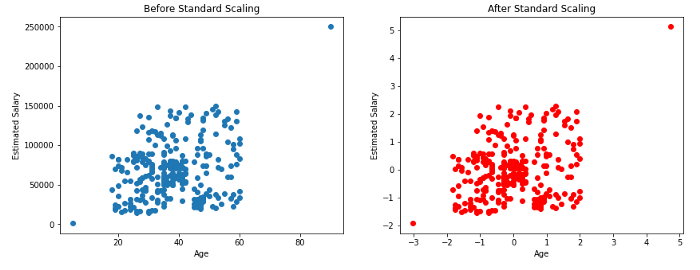


Image by author

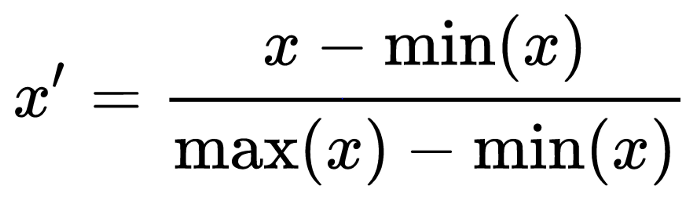




The above plots show that the outliers in our data will be still the outliers even after applying the standard scaling. So, as data scientists, it is our responsibility to handle the outliers.

**What is Normalization?**

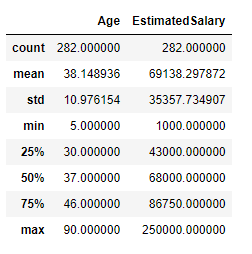
MinMaxScaling(a commonly used normalization technique) is one of the feature scaling techniques, it transforms features by subtracting from the minimum value of the data and dividing by (maximum minus minimum).



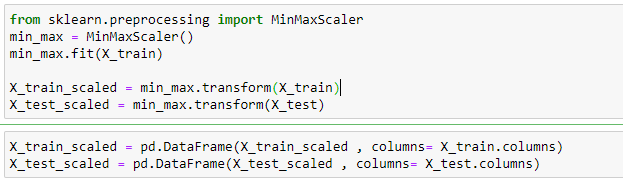
Formula for MinMaxScaling

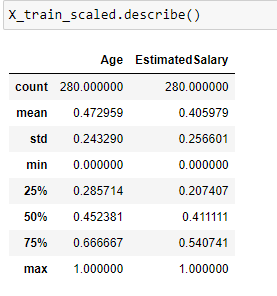
So now we have seen the formula min-max scaling, Now we shall look at how it can be applied to our dataset.

**Description of the dataset:**



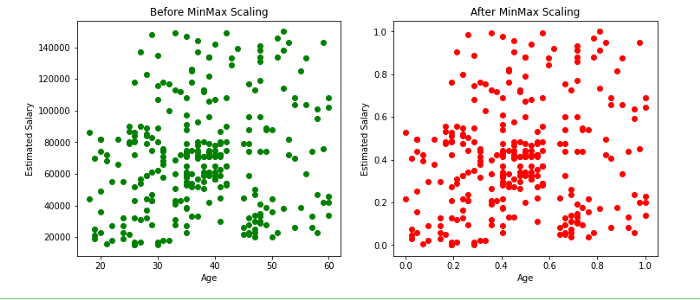
describe the data before min-max scaling





Note that the **minimum** value of both the input features Age and Estimated Salary has become **0** and the **maximum** value has become **1** after applying MinMax scaling.

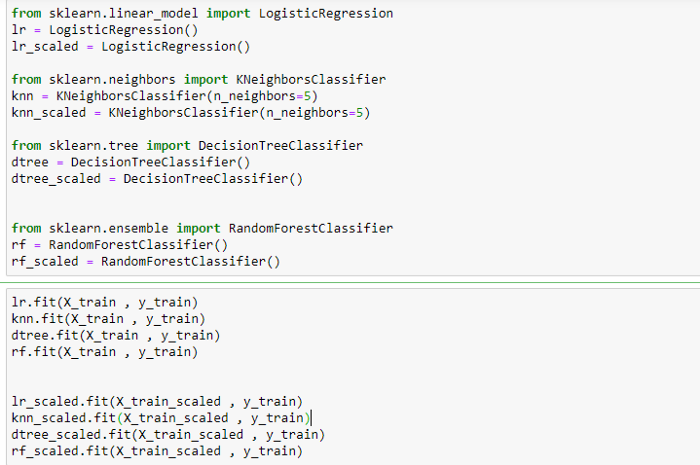
**Effect on Distribution of the data:**

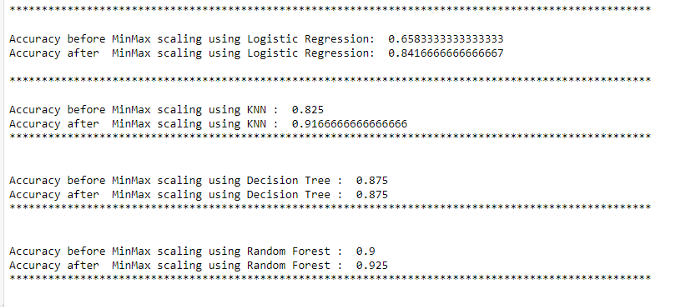


Effect of scatter plot before and after Min-Max scaling

From the above scatter plots and KDE plots we can note that the distribution of the data remains the same even after applying the min-max scaler, only the scale changes.

**Effect of Normalization on different Machine Learning algorithms:**

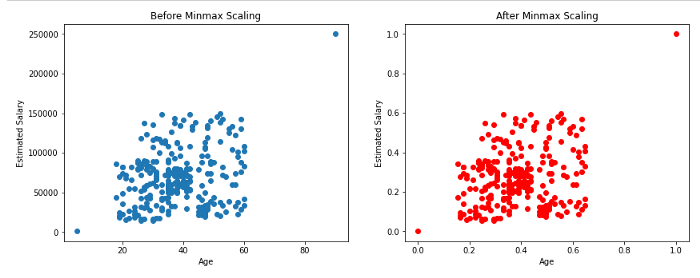




Effect on accuracy

In the above examples, the accuracy of Logistic regression and KNN increased significantly after scaling. But there was no effect on accuracy when the decision tree or random forest was used.

**Effect on outliers:**



Effect of normalization on outliers.

As shown above, there will not be any effect on outliers even after applying min-max scaling.

**Observations:**

* The resulting data **after standardization** will have a **mean of 0** and a **standard deviation of 1**, whereas the resulting data after **min-max scaling** will have a **minimum value as0** and a **maximum value of 1** (Here the mean and standard deviation can be anything).
* The scatter plots and KDE plots above show that there will be **no change in the distribution of data** before and after applying the standard scaler or min-max scaler, only the scale changes.
* The feature scaling step has to be performed while applying algorithms where **distance gets calculated** (Eg: KNN, KMEANS) and involves **gradient descent** (Eg: Linear and Logistic regressions, neural networks).
* There will not be any effect of scaling when we use tree-based algorithms like decision trees or random forests.
* In the above examples, the accuracy of **Logistic regression** and **KNN** **increased significantly** after scaling. But there was **no effect** on accuracy when the **decision tree** or **random forest** was used.
* **Outliers in the dataset will still remain an outlier** even after applying the feature scaling methods, as data scientists, it is our responsibility to handle the outliers.
* There is no hard rule to tell which technique to use, but we need to check both standardization and normalization and decide based on the result which one to use.

Please visit the mentioned links to get the full code. [Normalization](https://github.com/gowthamsr37/How-to-use-normalization/blob/main/Normalization.ipynb) [Standardization](https://github.com/gowthamsr37/Why-standardization-is-required/blob/main/Effect%20of%20Standardization.ipynb)