

BLACK FRIDAY

- Importing important libraries for reading the dataset.
- PANDAS And NUMPY
- Reading dataset and assigning data to a variable.
- After that checking for the null values.
- Some Nans are shown in two columns of dataset.
- A column in dataset is like as having a nominal data.
- Let's delete this column.
- Importing encoder (Label Encoder, Ordinal Encoder) for encoding the required column.
- After encoding the required column again, check for dataset.
- Because two columns are having null values, are such data type that we cannot use mean or mode method.
- I tried many times to filling Nans by still it shows Nans.
- So, we filled these Nans by 0.
- Checking again null values for dataset.
- Changing required column from float datatype to int datatype.
- Importing libraries for visualization for dataset.
- For checking the trend of dataset, will plot the data.
- For checking, data is normally distributed or not.
- First Distribution plot.
- Data is shown that it is not normally distributed.

- Second boxplot for checking the presence of outliers.
- We will also check outliers by using Quantiles and IQR.
- Some of outliers are shown in only our target variable. We will not do any type of changes.
- Plotting count plot individually for each column.
- Finding Quantiles and IQR for checking outliers individually.
- After checking for outliers individually, removed outliers where present.
- Again, plotting distribution plot and boxplot for checking the shape of data and for seeing outliers.
- Now, plotting heatmap.
- By plotting heatmap, it will have cleared that there is no any type of correlation among columns.
- No any type of correlation is found by plotting heatmap.
- Split dataset into two variables for further process.
- Scatter plot for 'x' variable data to 'y' variable data.
- Importing Standard Scaler and train test split.
- Standardize the data.
- Train test split.
- Also importing metrics for checking the error of each model.
- First model Linear Regression
- Second model Decision Tree
- Third model Random Forest
- Fourth model AdaBoost

- Fifth model XGB
- Among all these five models, Random Forest is giving good accuracy on train data and test data.
- Ada Boost also is giving some good accuracy.
- We will do hyperparameter tuning for these two models.
- After doing hyperparameter tuning, Random Forest is giving good score.
- After doing hyperparameter tuning, AdaBoost is giving low score.
- Considering Random Forest for saving and further processing.
- By using Pickle, we will save Random Forest.