# **Project-Black Friday Sales Predection**

* Project Introduction:

Black Friday is an informal name for the Friday following Thanksgiving Day in the United States, which is celebrated on the fourth Thursday of November. The day after Thanksgiving has been regarded as the beginning of the United States Christmas shopping season since 1952.

* Important libraries:

1. **Pandas** - Helps us to create Datastructure and manipulate Datastructure.

- Also used to read the Dataset.

1. **Numpy** - Numerical Python

- used for working with arrays.

- also has functions for working in domain of linear algebra, fourier transform, and matrices.

1. **Seaborn** - data visualization library built on top of Matplotlib.
2. **LabelEncoder class**- From library sklearn.preprocessing is used to transform categorical (non-numeric) data into numerical labels.
3. **StandardScaler**- specifically used to standardize features in a dataset.
4. **train\_test\_split** - to split a dataset into training and testing sets.

* Mchine Learning Models:

1. **LinearRegression**- finding the best-fit line
2. **DecisionTreeRegressor**
3. **RandomForestRegressor**

* Load and View the Dataset
* Data Understanding:

1. shape()- Total no.of Rows and Columns.(550068,12)
2. describe()-Descriptive Statistics and shape of our datset.
3. Info()-Help us to detect null values and also gives datatypes of columns.
4. Isnull.sum()- Gives clear view of null values (if present)

* Data Visualization:

1. count plot- visualize the frequency distribution of a categorical variable.
2. Bar plots-The height of each bar represents the value of a data point, and the width of each bar represents the category of the data.
3. Box plot- to check outliers

* Data Preprocessing:

Data mining to get useful data

* Visualization using Destribution plot
* Log Transformation: used for transfoem the data and change the data into normal Distribution.
* Train-Test-Split:80% - Training and 20% Testing
* Model Building:

1. Linear Regression- Evaluation Metrics

* r-2 Score
* Root Mean Square Error(RMSE)

1. Decision Tree-Root Mean Square value is less(it perform very well;better than previous model)
2. Random Forest Regressor-better than previos model(RMSE is low and r-2 score is high)

* Conclusion:

Implanted multiple supervised models such as Linear Regressor,Decision Tree Regressor, Random Forest Regressor and XGBOOST Regressor. Out of these supervised models, based on the RMSE scores XGBRegressor/XGBOOST Regressor was the best performer with a score of 2879.