# EDA(Exploratory Data Analysis)

# Exploratory Data Analysis is a process of examining or understanding the data and extracting insights or main characteristics of the data. EDA is generally classified into two methods, i.e. graphical analysis and non-graphical analysis

# EDA is very essential because it is a good practice to first understand the problem statement and the various relationships between the data features before getting your hands dirty.

# ML PipeLine

# Step1: Data Gathering or Data Ingestion

# Step2: EDA

# Step3: Preprocessing

# Step4: Model Building using Machine Learning algorithm

# Supervised: linear Regression, Logistic Regression, SVM, Random Forest, XGBoost, Gradient Boost, Decision Tree, Naïve Bayes.

# Unsupervised: K-means, DB Scan, Hierarchical

# Step5: Evaluation or validation

# EDA

# Profile of data

# Statistical Analysis

# Graph Based Analysis

# Feature Engineering

# Missing Value-> Null value

# Outlier->Extreme Value

# Encoding-> Categorical value to Numerical Value

# Scaling->Compress our data in a certain Range.

# Imbalanced Data-> Up sampling and down sampling

# Transformation

# Feature Selection->selection some certain data for train and test data.

# Dimension Reduction->Remove the column which is not need anymore. (Like. PCA, LDA)

# Reference : [https://www.analyticsvidhya.com](https://www.analyticsvidhya.com/blog/2021/02/introduction-to-exploratory-data-analysis-eda/)

# Note: To Start Machine Learning Practical We have to Follow the both EDA and Feature Engineering.