**APPROACH FOR SALES PREDICTION**

* **UPLOADING LIBRARIES & IMPORTING DATASET**
* Initially uploaded the libraries like pandas, Numpy , Matplotlib and Seaborn.
* **DATA CONCATENATION**
* The test data doesn’t consist of order and sales columns.
* Add the sales columns to test data by keeping constant variable as test.
* Combined both train and test data row wise by using concat statement .
* **DATA CLEANING**
* Made a check on head, tail, shape and data types of the combined dataset.
* Dropped ID, Store id, Order, Date columns dataset.
* Checked null values by using **isnull( ).sum( )** statement.
* **DATA PREPROCESSING**
* Divided combined dataset into numerical and object data.
* Divided the categorical data from numerical data and dropped the categorical columns from numerical data.
* Transferred sales from object data to numerical data.
* Dropped sales column from object data.
* Made dummies for both object and categorical data.
* Concatenated numerical , object dummies data, categorical dummies data columnwise.
* Divided clean data into train and test data.
* Dropped sales columns from the test data.
* Plotted the heat map for both train and test data.
* Plotted the graphical representations like distplot, piechart, histograms, boxplot,line.
* **MACHINE LEARNING MODELS**
* Divided the new train data into **independent variable(X)** and **dependent variable (y)**.
* The **dependent variable (y)** consists only **sales column**.
* The **independent variable (X)** consists **all the columns** **except sales column**.
* Worked on **independent variable (X)** to machine learning models like **Linear Regression, Decision Tree, Random Forest, Gradient Boosting, XGBoost** for model building and found R-Square
* Predicted the values for test data and converted to csv file.
* Comparing all the MSE #XGBoosting is the best fit model for the given dataset.