**Design Document**

I used two approaches two solve the given problem.

1. Machine Learning Approach
2. Timeseries Approach

**Machine Learning Approach:**

Random Forest and XGBoost models are used to train the model. Random Forest is giving better results between two as it is not overfitting. XGboost is highly overfitting.

Process:

1. Read the dataset
2. Used Featurization Techniques to create more features out of ‘DATE’ column
3. Performed EDA and found out that ‘SALES’ are higher in the end of year
4. Used one hot encoding to vectorize categorical features
5. Split the dataset in the proportion of 80:20 for train and test
6. Performed Cross Validation using RandomSearchCV to get the optimum parameters
7. Trained both the models using optimum parameters
8. Predictions were made on test and future data using trained model

**Time Series Approach:**

ARIMA and SARIMA models are used to train the model. The dataset is having seasonality so SARIMA model is better suited for this dataset.

Process:

1. Read the dataset and make ‘DATE’ column as index
2. Visualized data to see check whether it is timeseries or not
3. Used Dickey – Fuller to check whether time series is stationary or not
4. Used AutoCorrelation and Partially Autocorrelation to calculate values of P and Q respectively.
5. Used ARIMA and SARIMA models to Train on the data
6. Prediction using SARIMA model show that forecast is following original data.