Approach for the Jobathon:

The problem is about predicting the sales for a given dataset, the model I choosed is XGBoost regressor as I have used it before and got my hands dirty in using this regressor.

Preprocessing is the initial step which focuses on improving the input data by removing unwanted impurities and redundancy. The pre-processing step I took is checking of missing values, data integrity and data consistency. I have implemented one-hot encoding for the categorical variables as it will be easier for the model to perform proving a higher accuracy. I have extracted the day from the given date which is like feature engineering or feature generation. And created an one hot encoding for the day columns from Monday to Sunday, My main approach was to make the dataset containing values only with zeros and ones so that I have used one hot encoding using python dummies.