20201700994 يوسف احمد بسيم علي غيط
20201700503 علي سامح سعد عبدالرحمن
20201700463 عبدالله ماجد محمد عوض
20201700467 عبدالمجيد حاتم عبدالمجيد عبدربه
20201701014 يوسف ضياء الدين محمد حسين
20201700500 علي ابراهيم علي ابراهيم

1) Drop date column and add (day ,month ,year) columns in DateFormat Function ,three columns better than one column (date).
2) Drop outliers greater than 75% and lower than 25% with outlier Method
3) Check the how many countries in country column ,if the same drop the column
4) Choose the values of the CategoryTree to (MainCategory ,Subcategory) , then drop the original column
5) Encoding all columns with string values
6) Applying DateFormat function on ShipDate and OrderDate
7) Dropping 5 columns with lowest absolute correlation (Shipmonth ,Ordermonth ,Orderday ,Rowid ,Productname) with the Profit Column.
8) Scaling all Data using MinMax Method
9) Splitting Data 20% Test - 80% Training

10) Feature Selection using Anova model with the highest 10 for the polynomial model
11) Generating Polynomial Model with Degree 3 to avoid overfitting.
12) Generating Ridge Model with the worst 8 features (ship_mode ,ship_day ,ship_year ,order_year ,customer_id ,customer_name ,order_id ,postal_code) from the Correlation table.
13) Saving each model using Pickle library
14) Polynomial model is better than ridge because of the feature ,then better accuracy.