INDUSTRIAL COPPER MODELING

PRICE PREDICTION:

Overview:

The Industrial Copper Modeling project focuses on predicting the selling price and status (won or lost) in the industrial copper market using machine learning regression and classification algorithms. By exploring the dataset, performing data cleaning and preprocessing, and applying various machine learning techniques, we aim to develop models that can accurately predict the selling price and status in the copper market.

About Dataset:

The dataset used for this analysis contains information about industrial copper transactions, including variables such as selling price, quantities, and status (won or lost). It provides a comprehensive view of the copper market and factors that influence the outcomes of transactions.

Approach:

- **Removing Null values**
- **Changing dtypes**
- **Handling outliers**
- **Handling skewness**
- **EDA – Analysis the data & find the relationship.**
- **Model selection & Encoding.**
- **Hyper-parameter tuning.**
- **Model Evaluation.**

Null values:

From whole dataset, matrial ref column had more null values. So, replaced all those null values using unknown. Further, checking remaining null values are below 1%, they removed from dataset.

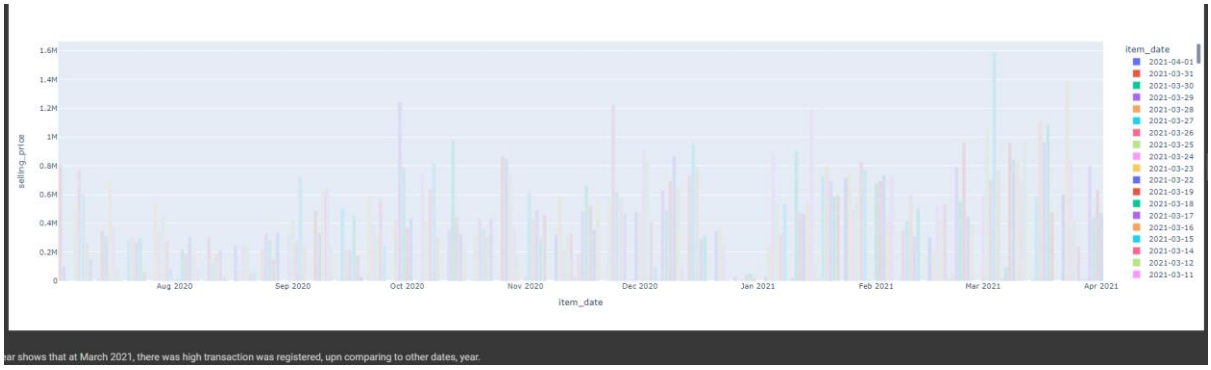
Handling Outliers:

From whole dataset, several columns like application, quantity tons were had outliers. Used, IQR method and removed the outliers from dataset.

EDA:

It clearly shows that all the columns had huge relationship with selling price. (Target).

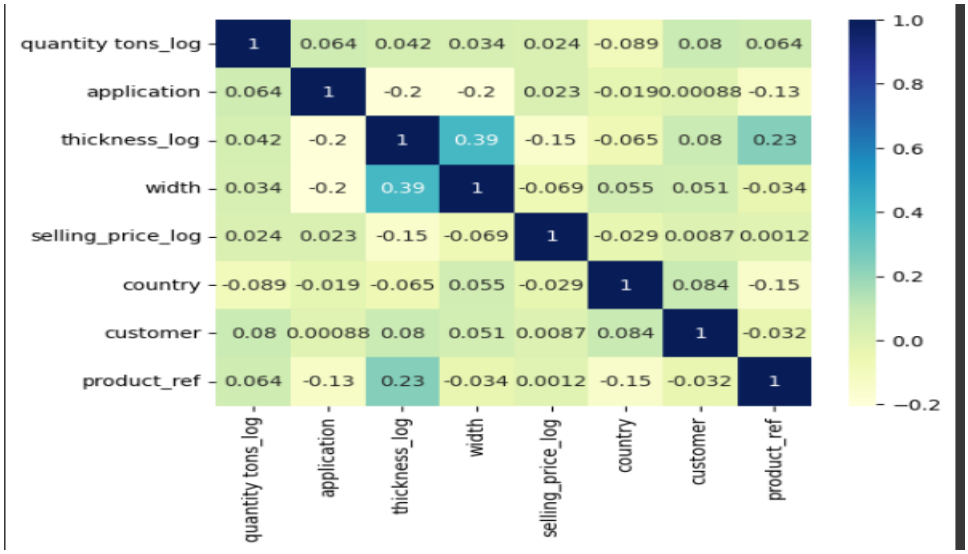
Further, please see below the relationship between transaction date and selling price.



It clearly shows, March 2021 was highest transaction done.

Correlation:

Yes, I previously mentioned, all the columns had serious relationship with target column.



Model selection:

- **Regressor**
- **Classification**

Regressor:

For regression model, I have used DecissionTree, RandomForest, Adaboost, GradientBoosting.

```
from sklearn.preprocessing import OrdinalEncoder
from sklearn.model_selection import train_test_split
from sklearn.metrics import mean_absolute_error, mean_squared_error, r2_score
from sklearn.tree import DecisionTreeRegressor
from sklearn.ensemble import ExtraTreesRegressor
from sklearn.ensemble import RandomForestRegressor
from sklearn.ensemble import AdaBoostRegressor
from sklearn.ensemble import GradientBoostingRegressor
from xgboost import XGBRegressor
from sklearn.preprocessing import StandardScaler, OneHotEncoder
```

Hyper-Parameter Tuning:

Used, Grid search Hyper-paramenter tuning to generalize the model.

Streamlit application:

Where users can select or enter inputs on the columns like Status, item_type, Country, Application.

Once they entered all details, they can find the selling Price of Industrial copper.

PFB: the sample mode of streamlit screenshot for reference!!!!