Task :-

My task is to predict the cost required to ship these sculptures to customers based on the information provided in the dataset.

Basic Steps :-

1. I have imported all the necessary libraries like pandas, numpy, matplotlib, seaborn etc.
2. I imported the training and testing datasets.

Exploratory Data Analysis :-

1. I checked the missing Values
2. I created a list of numerical and categorical features present in the dataset
3. I have plotted the distributions of numerical features
4. I used box plot for detecting outliers
5. I used histograms for plotting categorical features

Data Preparation :-

1. I have replaced missing values of numerical features with mean.
2. I have replaced missing values of categorical features with the high frequency category present in those features.
3. I have dropped unwanted columns like CustomerID, Artist Name , Customer Location etc.
4. I have derived a new feature from the scheduled date and Delivery date.
5. I have dropped both scheduled and delivery dates.
6. I have encoded categorical features with OnehotEncoding.
7. I have transformed the cost feature by applying np.log1p() and abs() to the cost feature.

Model Building :-

1. I have used some basic models like logistic regression, K nearest neighbors (KNN), NaiveBayes, Decision trees but I got less accuracy.
2. I have used ensemble techniques like RandomForest, Ada Boost ,Gradient Boosting . Among these three algorithms Gradient Boosting performed well with a score of 93.2 after Hyperparameter tuning.
3. I have trained Artificial Neural Networks (ANN) as well but ensemble techniques have given a better score compared to ANN's.
4. I have used other ensemble algorithms like XGBoost, CatBoost

Among all these algorithms CatBoost is performed well with a score of 94.83 after Hyper parameter tuning.

After building the model , I have used expm1() to transform the predicted values of the test data.

References :-

1. Krish Naik Youtube Channel :- https://www.youtube.com/channel/UCNU_lfiiWBdtULKOw6X0Dig
2. Sklearn documentation and Catboost documentation.