Abhishek Banait

REPORT

- We first checked for the datatypes of every column in dataset.
- Using df.info we got the overall summary of the data including datatypes and null values.
- As there were no null values, we didn't need any cleaning.
- We plotted boxplots to check outliers which weren't detected.
- We then plotted histplot of Location
- We then plotted the heatmap of correlation of dataset.
- We performed One Hot Encoding on all the categorical data in the form of object dtype.
- Then, we concatenated the original dataframe with the encoded columns
- Doing so, we deleted CustomerID, Name, Gender and Location columns as they weren't needed anymore.
- We again plotted the heatmap of correlation of data with new encoded columns
- After, we separated the predictive features and target feature.
- Then we used the ExtraTreeRegressor to check important features.
- We then split the training and test data
- We used the Logistic Regressor model for prediction
- We trained the model with training data and performed prediction with test data
- We then evaluated the model score and performed cross validation at the end.