Assignment: Predicting Post Engagement

Objective: Build a predictive model to estimate the **Total Interactions** (sum of comments, likes, and shares) a post will receive, based on various features of the post.

Instructions:

Data Preprocessing:

1. **Load the Data:** Import the dataset and inspect it for missing values or anomalies.
2. **Feature Engineering:** Create relevant features from the existing data. For example, you might combine like, comment, and share into Total Interactions if it's not already calculated.
3. **Categorical Variables:** Convert categorical variables like Type, Category, and Paid into numerical representations using one-hot encoding or label encoding.

Exploratory Data Analysis (EDA):

1. **Visualization:** Create visualizations to understand the distribution of features and their relationships with Total Interactions. Examples include histograms, scatter plots, and box plots.
2. **Correlation Analysis:** Examine correlations between features and Total Interactions to identify which variables have the most impact.

Model Building:

1. **Feature Selection:** Choose relevant features that you believe will have the most influence on Total Interactions.
2. **Model Choice:** Implement and train a predictive model. You can use linear regression for a simple approach or try more complex models like decision trees or random forests.
3. **Evaluation:** Assess the model's performance using metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), or R-squared. Split the data into training and testing sets to validate your model's performance.

Interpretation and Reporting:

1. **Model Interpretation:** Explain the significance of the features in your model and how they contribute to predicting Total Interactions.
2. **Reporting:** Provide a summary of your findings, including model performance, feature importance, and any insights gained from the analysis.

Deliverables:

Code: Complete code for data preprocessing, model training, and evaluation.

Report: A report summarizing the EDA (Exploratory Data Analysis), model performance, feature importance, and any recommendations or insights.

Estimated Time:

Data Preprocessing & EDA: 2-3 hours

Model Building & Evaluation: 2-3 hours

Interpretation & Reporting: 1-2 hours

Code Submission:

1. Create a public git repository
2. Push your code in that repository
3. Share link of repository