Predicting Post Engagement - Summary Report

Objective:

The task was to 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.

Data Preprocessing:

The dataset contains several features related to Facebook post performance for a cosmetic brand.
Steps taken:

1. Handling Missing Values: Missing values in the 'like' and 'share' columns were filled with 0. The 'Paid' column's missing value was filled with the mode.

2. Feature Engineering: The Total Interactions was already present in the dataset, so no additional calculation was required.

3. Categorical Variables: Categorical columns (Type, Category, Paid) were converted into numerical formats using one-hot encoding.

Exploratory Data Analysis (EDA):

Key observations from the dataset:

- Likes were the most frequent interaction, followed by shares and comments.
- Visualizations were created for likes, comments, shares, and total interactions to understand the distribution.
- Correlation analysis revealed that the number of likes and shares had the highest correlation with total interactions.

Model Building:

Three models were evaluated:

1. Linear Regression: The model performed perfectly but was considered overly simplistic.

2. Decision Tree: Performed well but had slightly higher errors compared to Random Forest.

3. Random Forest: Selected as the final model due to a balance between accuracy and complexity.

Evaluation Metrics (Random Forest Model):

- Mean Absolute Error (MAE): 10.11
- Mean Squared Error (MSE): 1237.47
- R-Squared (R^2): 0.977

Feature Importance:

The most important features impacting post engagement were:

1. Likes
2. Shares
3. Reach by People who Liked the Page
4. Comments

Recommendations:

1. Focus on Increasing Likes and Shares.
2. Enhance Reach to Page Likers.
3. Optimize Post Timing.

Conclusion:

The Random Forest model provided a reliable way to predict post engagement with high accuracy. By focusing on increasing likes and shares, and optimizing post reach, the brand can enhance the effectiveness of their posts.