■ Project Report: Unlocking YouTube Channel Performance Secrets

1. Project Overview

This project analyzes YouTube channel performance data and builds a machine learning model that predicts video revenue based on engagement and viewership metrics. It uses EDA, feature engineering, and predictive modeling to provide actionable insights for creators and marketers.

2. Dataset Description

Dataset contains 70 columns and ~364 records. Key data includes Video Details, Revenue Metrics, Engagement Metrics, Audience Data, and Monetization. Target variable: Estimated Revenue (USD).

3. Data Cleaning & Preprocessing

- Removed missing values - Converted publish time to datetime - Converted video duration to seconds - Verified dataset integrity

4. Exploratory Data Analysis (EDA)

- Revenue Distribution: skewed (most videos earn little, few earn high) - Correlation Heatmap: strong link between Views & Revenue - Top Performers: identified top 10 revenue videos

5. Feature Engineering

Created: - Revenue per View = Revenue / Views - Engagement Rate = (Likes + Shares + New Comments) / Views * 100

6. Predictive Modeling

Model: Random Forest Regressor Features: Views, Subscribers, Likes, Shares, New Comments, Engagement Rate Target: Estimated Revenue (USD) Performance: MSE ~ moderate, R^2 ~0.65

7. Evaluation – Predicted vs Actual Revenue

Example comparison: Video 0 → Actual: 0.00, Predicted: 9.84 Video 3 → Actual: 14.95, Predicted: 10.05 Video 9 → Actual: 24.60, Predicted: 7.83

8. Feature Importance

Top predictors: 1. Views 2. Engagement Rate 3. Subscribers Secondary: Likes, Shares, New Comments

9. Insights & Recommendations

- More views = higher revenue - Engagement boosts efficiency - Subscriber growth vital for long-term - CTR & Watch Time indirectly drive revenue

10. Deployment

Model saved as youtube_revenue_predictor.pkl for reuse with joblib.

11. Future Improvements

- Hyperparameter tuning - Try Boosting models (XGBoost, LightGBM) - Log-transform skewed revenue - Add geographic/ad-type data

12. Conclusion

This project demonstrated how data + ML unlock insights into YouTube performance. Focusing on views & engagement improves revenue potential. The model predicts earnings and supports data-driven strategy.