|  |  |
| --- | --- |
| **Project Title** | **Content Monetization Modeler** |
| **Skills take away From This Project** | **Regression models**  **Predictive Modeling Feature Engineering**  **Data Cleaning**  **Exploratory Data Analysis (EDA)**  **Regression Metrics (R², RMSE, MAE)**  **Data Visualization**  **Streamlit**  **Python**  **Pandas**  **Scikit-learn**  **Categorical Encoding**  **Outlier Detection**  **Missing Value Handling** |
| **Domain** | **Social Media Analytics** |

**Problem Statement:**

As video creators and media companies increasingly rely on platforms like YouTube for income, predicting potential ad revenue becomes essential for business planning and content strategy.

**Data Set:**

* **Name**: YouTube Monetization Modeler
* **Format**: CSV
* **Size**: ~122,000 rows
* **Source**: [Synthetic, created for learning purposes]
* **Target Variable**: ad\_revenue\_usd

**Data Set Explanation:**

Each row in the dataset represents performance metrics for a specific video on a specific day. The columns include:

* video\_id: Unique identifier
* date: Upload/report date
* views, likes, comments: Performance metrics
* watch\_time\_minutes, video\_length\_minutes: Engagement and content length
* subscribers: Subscriber count of channel
* category, device, country: Contextual info
* ad\_revenue\_usd: Revenue generated (target)

**Preprocessing Requirements**:

* Handle ~5% missing values in key columns.
* Remove ~2% duplicated records.
* Encode categorical variables (category, device, country).
* Normalize or scale features if necessary.

**Business Use Cases:**

* **Content Strategy Optimization**: Helps creators determine what type of content yields the highest returns.
* **Revenue Forecasting**: Media companies can predict expected income from future video uploads.
* **Creator Support Tools**: Can be integrated into platforms offering analytics services to YouTubers.
* **Ad Campaign Planning**: Enables advertisers to forecast ROI based on content performance metrics.

**Approach:**

* **Understand the Dataset**: Load and inspect the provided dataset.
* **EDA (Exploratory Data Analysis)**: Identify trends, correlations, and outliers.
* **Preprocessing**: Handle missing values, remove duplicates, and encode categorical variables.
* **Feature Engineering**: Create new features (e.g., engagement rate = (likes + comments) / views).
* **Model Building:** Experiment with 5 different regression models to predict ad\_revenue\_usd, and compare their performance to identify the most effective model.
* **Model Evaluation**: Use appropriate regression metrics to evaluate performance.
* **Streamlit App Development**: Build a basic interactive app to demonstrate predictions and visualizations.
* **Interpretation & Insights**: Provide insights from the model about which features influence revenue.
* **Documentation**: Clearly document findings and code.

**Results:**

* A trained and evaluated Regression model to predict ad revenue.
* A cleaned, well-processed dataset ready for future analysis.
* Insights about what drives YouTube ad revenue.
* **A Streamlit app that allows users to test the model interactively.**