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
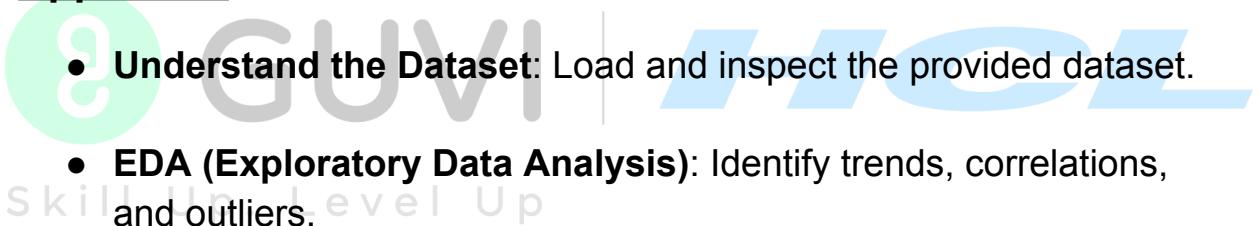
Problem Statement:

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. Your task is to build a Linear Regression model that can accurately estimate YouTube ad revenue for individual videos based on various performance and contextual features, and implement the results in a simple **Streamlit web application**.

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.**

Project Evaluation metrics:

- **R²** Score

- Root Mean Squared Error (**RMSE**)

- Mean Absolute Error (**MAE**)

- Code Quality & Documentation

- Insightfulness of Analysis and **EDA**

- Streamlit App Functionality and Usability

Technical Tags:

Linear Regression, EDA, Feature Engineering, Pandas, Scikit-learn, Outlier Detection, Missing Value Handling, Machine Learning, Data Visualization, Streamlit, Model Evaluation

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.

Project Deliverables:

- Jupyter Notebook or Python script with:

  - Full EDA

  - Preprocessing

  - Model building

  - Evaluation and insights

- Streamlit app showcasing:

- ○ Revenue prediction from user input

- ○ Basic visual analytics and model insights

- `README.md` with project overview and instructions

Project Guidelines:

- Follow clean coding practices (use functions, comments, etc.)

- Use Git/GitHub for version control and submission

- Avoid hardcoding file paths

- Include markdown explanations and error handling

- Ensure the Streamlit app is simple, intuitive, and functional

- The final notebook and app should run without errors top-to-bottom

Timeline:

Milestone	Deadline
Dataset Exploration & EDA	Day 1–2

Data Cleaning & Preprocessing	Day 3–4
Model Building & Tuning	Day 5–6
Streamlit App Development	Day 7
Documentation & Final Touches	Day 8

References:

Project Live Evaluation	🔷 **Project Live Evaluation**
EDA Guide	🔷 **Exploratory Data Analysis (EDA) G…**
Capstone Explanation Guideline	🔷 **Capstone Explanation Guideline**
GitHub Reference	🔶 **How to Use GitHub.pptx**
EDA	🔷 **EDA_INDEPTH**

PROJECT DOUBT CLARIFICATION SESSION (PROJECT AND CLASS DOUBTS)

About Session: The Project Doubt Clarification Session is a helpful resource for resolving questions and concerns about projects and class topics. It provides support in understanding project requirements, addressing code issues, and clarifying class concepts. The session aims to enhance comprehension and provide guidance to overcome challenges effectively.
Note: Book the slot at least before 12:00 Pm on the same day

Timing: Monday-Saturday (3:30 PM to 4:30 PM)

Booking link :https://forms.gle/XC553oSbMJ2Gcfug9

LIVE EVALUATION SESSION (CAPSTONE AND FINAL PROJECT)

About Session: The Live Evaluation Session for Capstone and Final Projects allows participants to showcase their projects and receive real-time feedback for improvement. It assesses project quality and provides an opportunity for discussion and evaluation.
Note: This form will Open only on Saturday (after 2 PM) and Sunday on Every Week

Timing: Monday-Saturday (05:30PM to 07:00PM)

Booking link : https://forms.gle/1m2Gsro41fLtZurRA

Created By:	Verified By:	Approved By:
Shadiya P P	Nehlath Harmain	Nehlath Harmain