Project Documentation: Top 500 YouTubers Analysis

Business Objective

To explore and understand key trends, patterns, and influencing factors among the top YouTube channels globally. The goal was to analyze which metrics contribute to subscriber growth and content performance using Exploratory Data Analysis (EDA).

Approach

- **Data Collection**:
  Scraped data from us.youtubers.me, resulting in an initial dataset of **1020 rows × 11 columns**.
- **Data Cleaning**:
  - **Removed** 509 duplicate rows.
  - **Handled** all null values.
  - **Engineered** new features, expanding the dataset to **502 rows × 21 columns**.
- **Exploratory Data Analysis (EDA)**:
  - **Univariate Analysis** to study individual variables (e.g., subscribers, views).
  - **Bivariate Analysis** to examine relationships (e.g., views vs. subscribers).
  - **Multivariate Analysis** to uncover deeper trends across multiple features.
- **Tools Used**:
  Python, Pandas, NumPy, Matplotlib, Seaborn, Plotly

Insights

- Entertainment and music channels dominate in subscriber count.
- A strong positive correlation exists between view count and subscriber numbers.
- Frequent uploads and consistent engagement often result in better growth.
- Some channels show high view counts but low subscribers, indicating viral or short-term spikes.

Conclusion

The project revealed meaningful insights into content trends and engagement strategies of top YouTubers. It also strengthened skills in data cleaning, feature engineering, and visualization. These findings can help content creators and marketers better understand audience behavior on YouTube.
