Unraveling YouTube Channel Dynamics

Northwest Missouri State University Section - 44517-02

Sushmitha Yelamakonda Abhilash Vallabhaneni Harika Madamanchi Anusha Bichal

# Project Idea

The project aims to perform an extensive analysis of YouTube channels, focusing on vital indicators such as subscribers, video views, content categories, and geographical distribution. Through the application of data science tools, our objective is to uncover hidden patterns, find emerging trends, and identify key factors that contribute significantly to the success of YouTube channels. This analysis helps content creators, stakeholders, and researchers better understand what makes YouTube channels popular. By exploring key metrics, we aim to contribute to the overall knowledge of what influences the performance of YouTube channels in the digital age.

# Project Description

”Unraveling YouTube Channel Dynamics” is all about deeply understanding what makes YouTube channels successful. We want to look closely at things like the number of subscribers, how many people view the videos, the types of content being created, and where in the world the audience is located. To achieve this, we’re using a powerful method called data science. It’s like a detective tool for numbers. We’re not just looking at the obvious stuff; we’re trying to find hidden patterns, discover new trends, and figure out the most important factors that make a YouTube channel successful.

# Tools and Technologies

The project will leverage various data science tools to conduct an extensive analysis of YouTube channels. These tools include:

## Jupyter Notebook

Used for interactive documentation, code execution, and visualization, providing a flexible environment for data exploration.

## Python Libraries

Essential Python libraries such as Pandas, NumPy, and Matplotlib will be em- ployed for data manipulation and numerical operations. These libraries enhance the capabilities of the analysis by providing efficient and effective data process- ing.

## SQL:

Technologies such as MySQL or PostgreSQL may be used, especially considering the scale of the dataset, to ensure efficient data storage and retrieval. SQL will be employed as a querying language for retrieving and working with data extracted from the dataset.

## Visualization Tools

Tools such as Tableau or Power BI may be employed for creating interactive and insightful visualizations. These tools enhance the presentation of findings, making them accessible to a broader audience.

## GIT:

Git Version control systems like Git may be employed to manage code changes collaboratively.

# High-Level Architecture

* **Data Import in Jupyter Notebook:** Import YouTube channel dataset directly into Jupyter Notebook.
* **Data Processing:** Process and clean the imported data for analysis.
* **SQL Operations:** Create a Framework and perform SQL operations directly on the dataset within Jupyter Notebook, allowing for flexible and efficient querying.
* **Data Analysis:** Conduct comprehensive analysis using Python Libraries, including Matplotlib for deep analysis.
* **Visualization:** Create interactive and insightful visualizations using tools like Tableau or Power BI.
* **Documentation:** Prepare a comprehensive and well-organized presenta- tion of the entire process and outcomes using LaTeX.

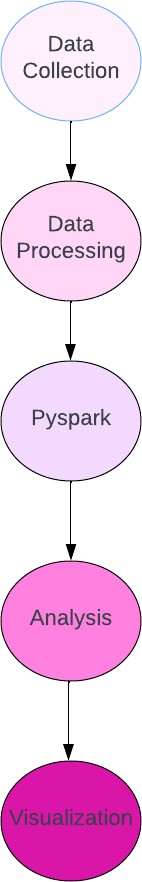


Figure 1: High-level architecture of the YouTube Channel Analysis Project.

# Goals

1. **Identify Channels with a High Variation in Monthly Earnings**
2. **Evaluate how efficiently each video generates revenue based on views.**
3. **Channels with High Engagement Relative to Subscribers**
4. **Analyze the performance of various channel categories by assess- ing the total number of video views they generated in the last 30 days.**
5. **Average Views per Video for Top 3 Education Channels**
6. **Top 5 Channels with the Most Monthly Earnings per Video**
7. **Identify channels that have experienced substantial growth over the past year by analyzing the percentage change in subscribers**
8. **Channels with the Highest Subscribers-to-Views Ratio**

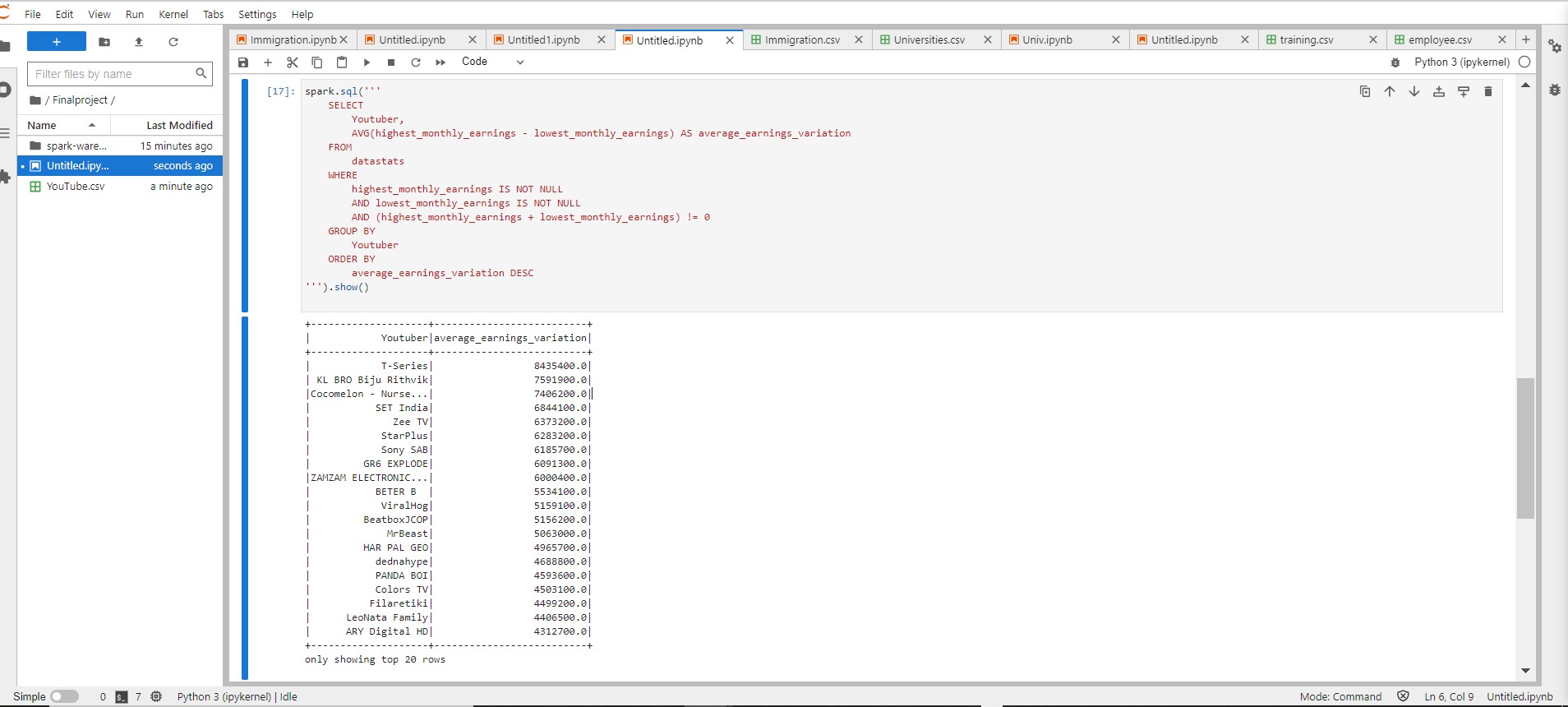
# Implementation Steps

1. **Download the Dataset from Kaggle:**
   * Navigate to the Kaggle dataset webpage.
   * Click on the ”Download” button to save the dataset file (commonly a .csv or .xlsx file) to your computer.
2. **Data Cleaning**
   * Clean an Excel dataset by removing empty rows, handling missing values, adjusting data types and save changes.
3. **Import Libraries and Load Dataset:**
   * Open your Python environment (e.g., Jupyter Notebook, Google Co- lab).
   * Import Pandas using import pandas as pd in your script or note- book.
   * Load the dataset with pd.read csv() for CSV files or pd.read excel()

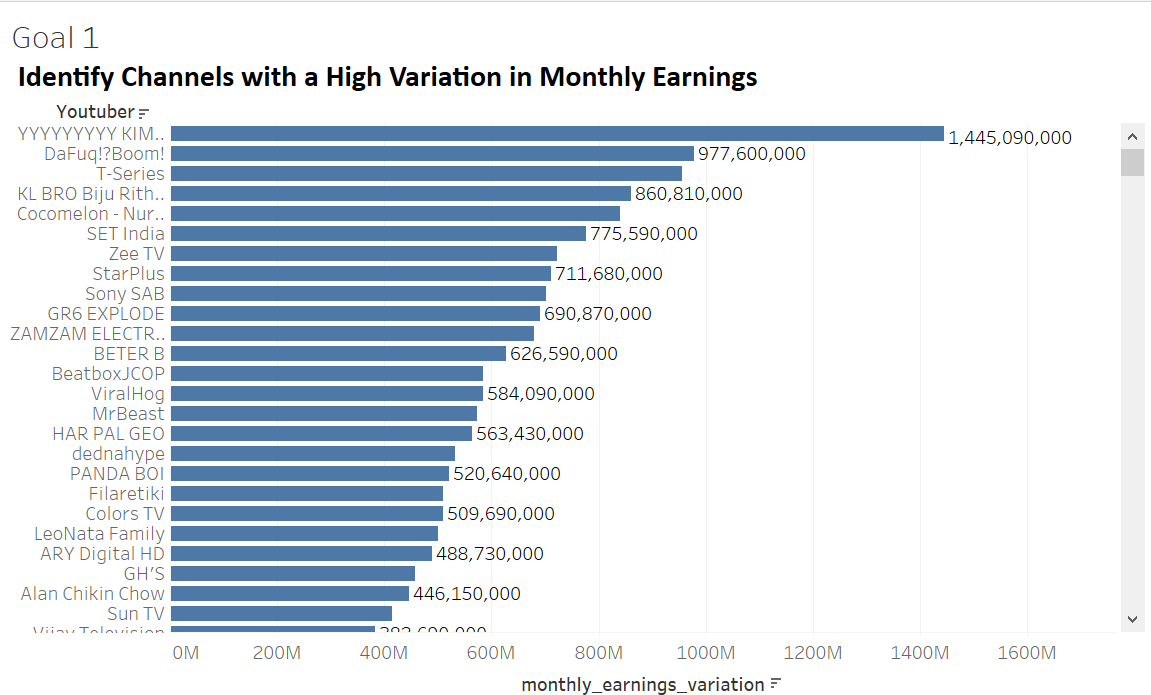
for Excel.

**Goal 1: Identify Channels with a High Variation in Monthly Earnings**

The goal of this query is to give us an idea of how much the monthly earnings of each YouTube channel typically change. The SQL query analyzes the YouTube channels in the dataset, calculating the average variation in monthly earnings for each channel. The results are then sorted in descending order based on the calculated average*earnings.*

**Screenshot:**

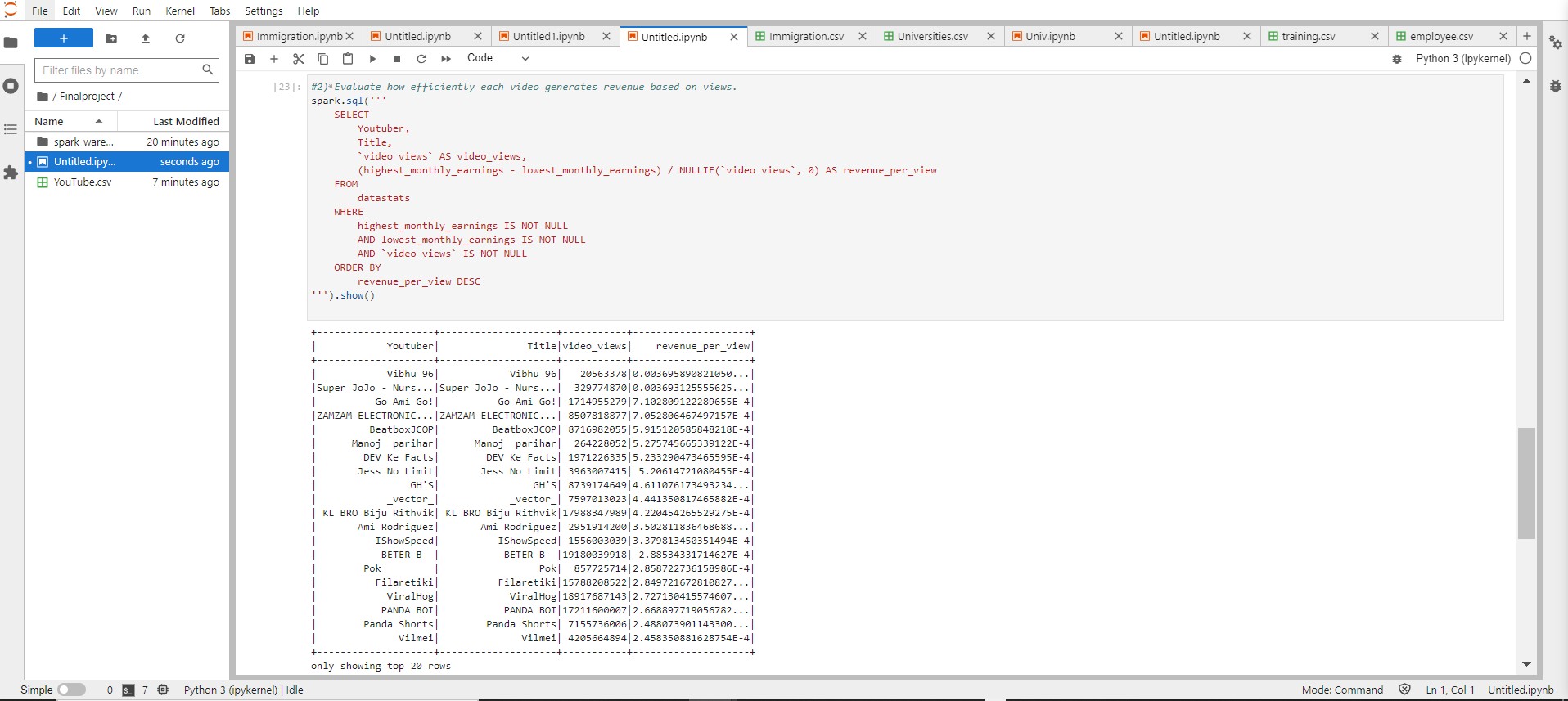
The Below Screenshot provides the resultant set showcasing channels with high monthly earnings variation.

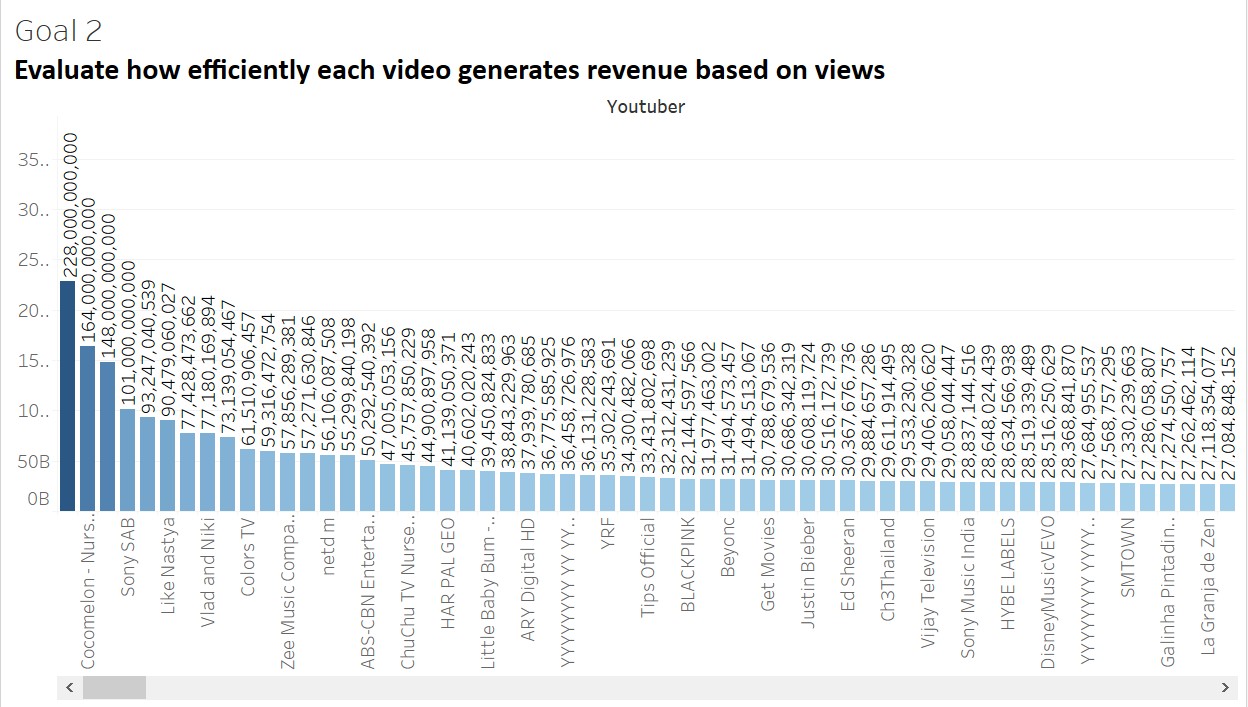
**Visualization:**

The output shows the top 20 YouTube channels with the highest average variation in monthly earnings. For Example, ”T-Series” has the highest average variation of 8,435,400, indicating a substantial fluctuation in monthly earnings. The list provides insights into channels with significant variations, which can be valuable for understanding the financial dynamics of these channels over time

**Goal 2: Evaluate how efficiently each video generates revenue based on views.**

The provided Spark SQL query evaluates the efficiency of each video in generating revenue based on views for YouTube channels in the dataset. The results are ordered by the calculated metric `revenue\_per\_view` in descending order.

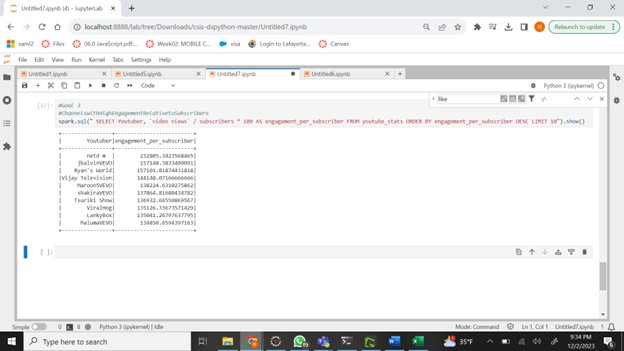
 **Screenshot:**

**Visualization:**

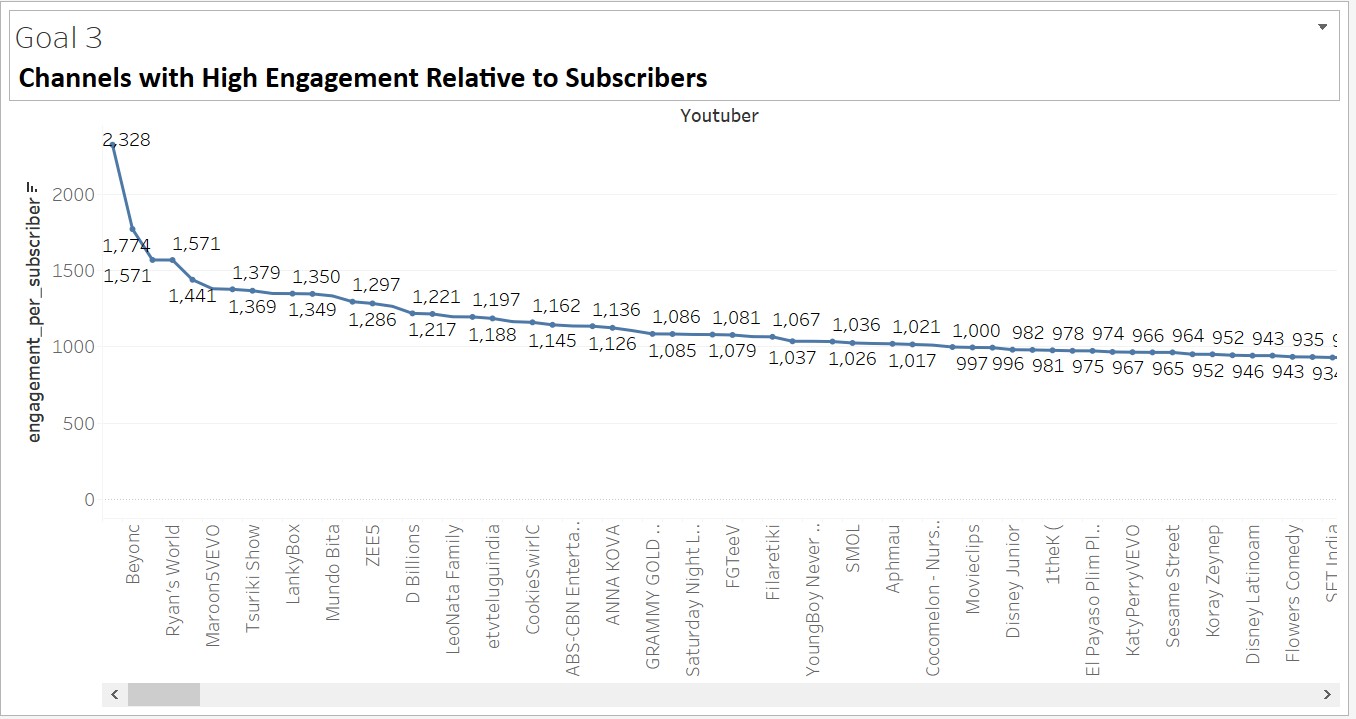
Visualization showcases that "Vibhu 96" has a revenue per viewof approximately 0.0037, indicating the revenue generated per view for that particular video. The list provides insights into videos that are relatively more efficient in generating revenue based on the number of views they receive.

**Goal 3: Channels with High Engagement Relative to Subscribers**

SQL query calculates ‘engagement per subscriber ‘, showing the per- centage of video views each subscriber. SQL query identifies YouTube channels with exceptionally high engagement relative to subscribers. The metric "engagement\_per\_subscriber" represents the percentage of video views per subscriber. The top 10 channels with the highest engagement are showcased, including "netd m" and "Ryan's World," suggesting a strong connection between content and subscriber activity.

**Screenshot:**

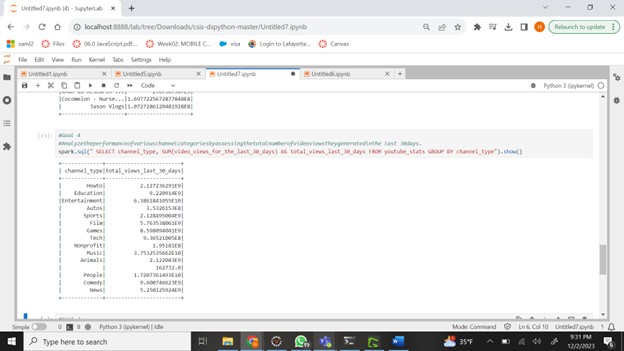
**Visualization:**



**Goal 4: Analyze the performance of various channel categories by assessing the total number of video views they generated in the last 30 days.**

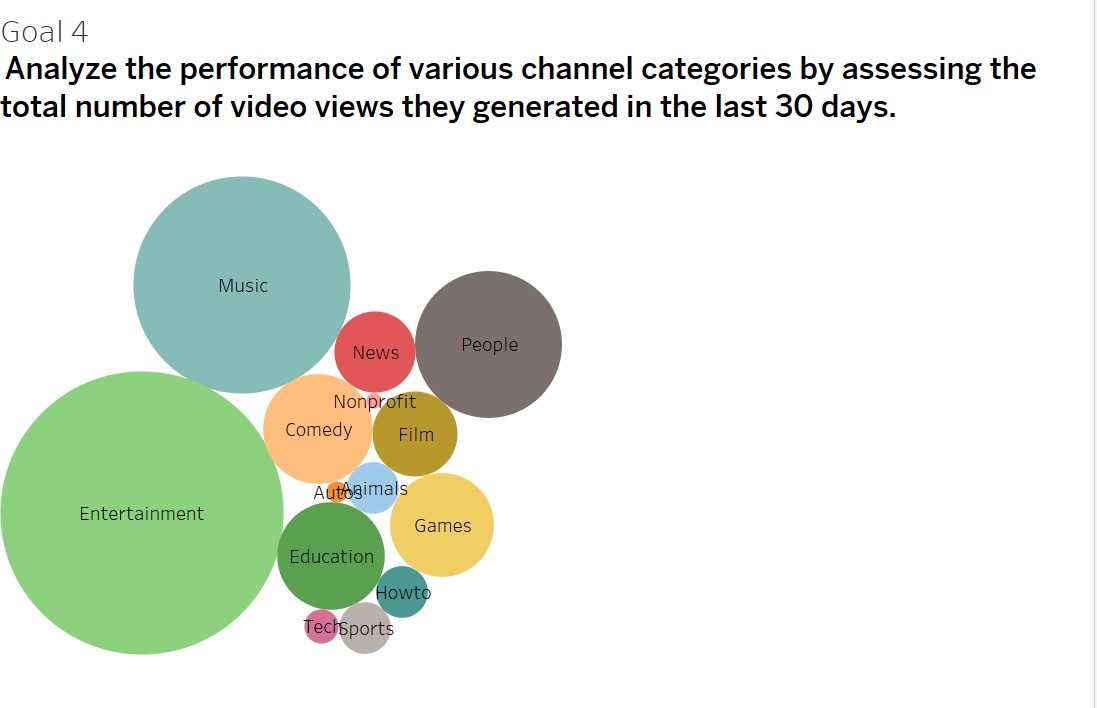
SQL query analyzes the performance of various YouTube channel categories by assessing the total number of video views they generated in the last 30 days. The results, grouped by "channel\_type," show the total views for each category. For example, the "Entertainment" category has generated approximately 61.56 billion views, while the "Education" category has accumulated around 9.22 billion views in the last 30 days.

**Screenshot:**



This Screenshot provides a summary of the overall views for different channel types.

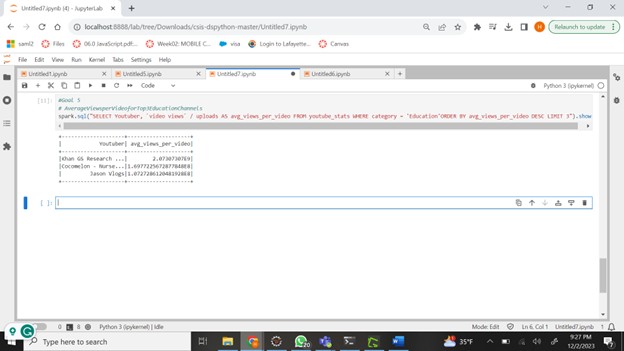
**Visualization:**



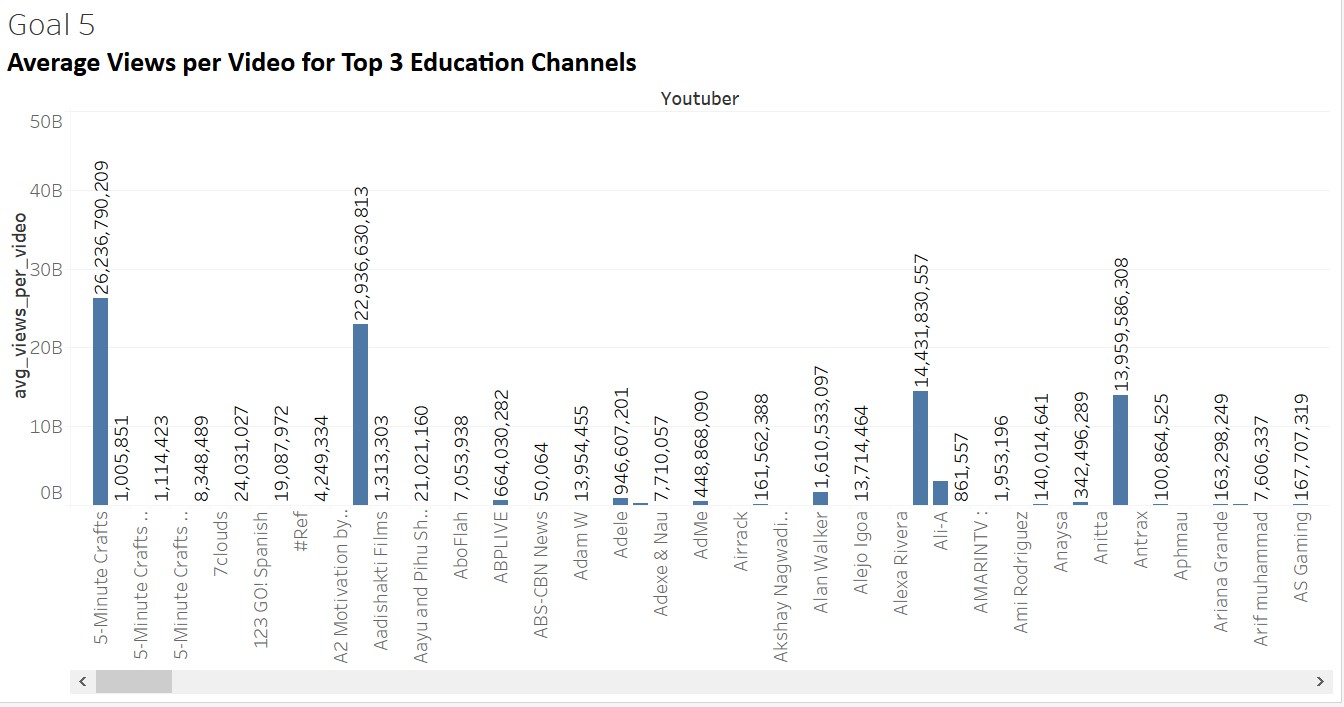
**Goal 5: Average Views per Video for Top 3 Education Channels**

This analysis focuses on YouTube channels in the "Education" category, looking at the top 3 channels. It calculates the average number of views each video receives for these channels. Notably, "Khan GS Research Centre" stands out with a very high average of around 2.07 billion views per video, indicating widespread viewership and popularity in educational content.

**Screenshot:**



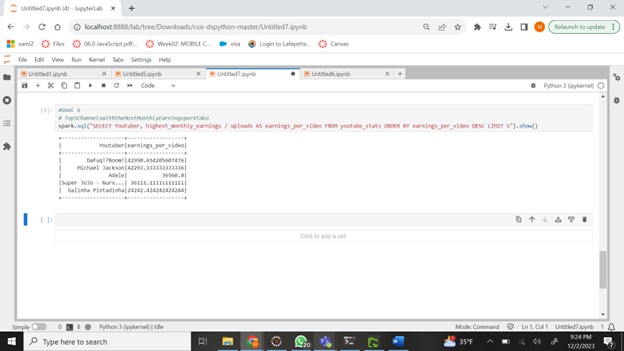
**Visualization:**



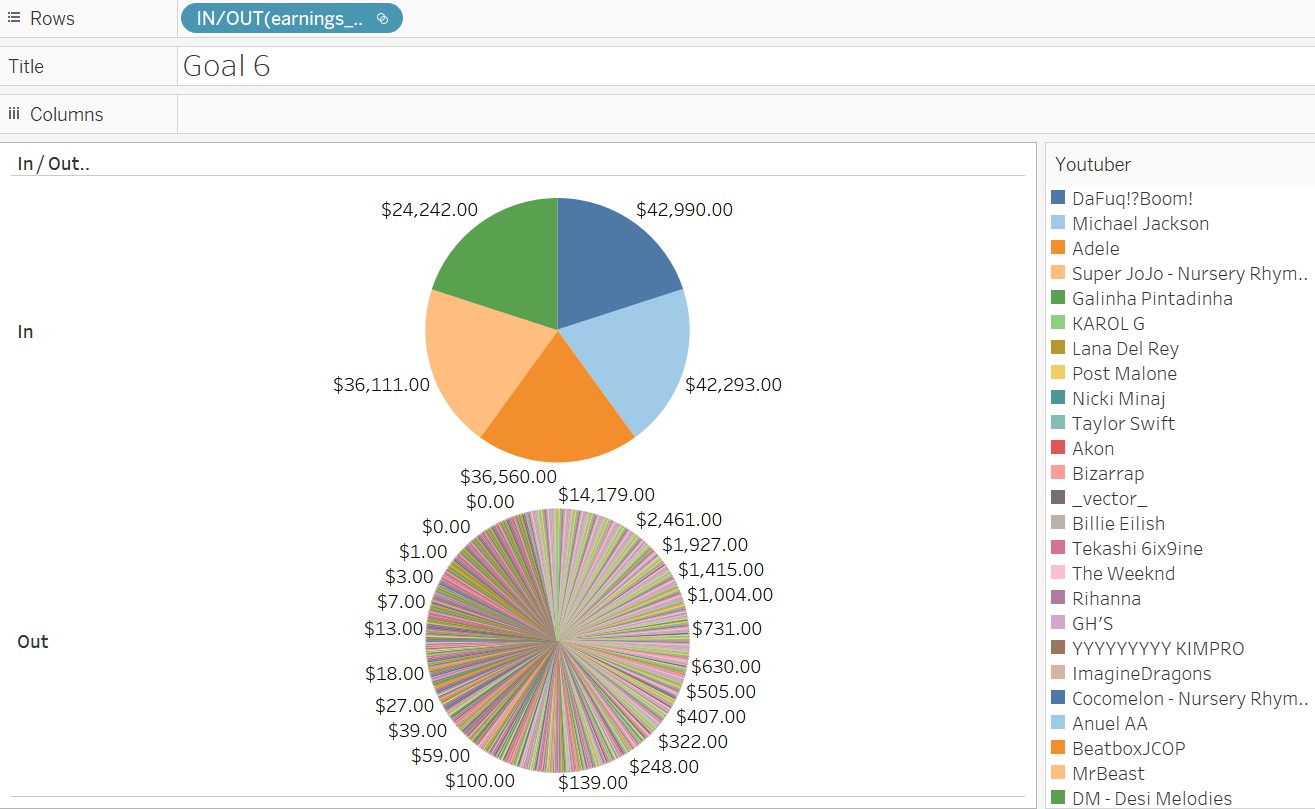
**Goal 6: Top 5 Channels with the Most Monthly Earnings per Video**

This analysis identifies YouTube channels with the highest earnings per video by calculating the ratio of highest monthly earnings to the number of uploads. The top 5 channels with the most significant earnings per video are highlighted, with "Michael Jackson" leading the list with an impressive earnings-per-video ratio of approximately 42,293.33.

**Screenshot:**



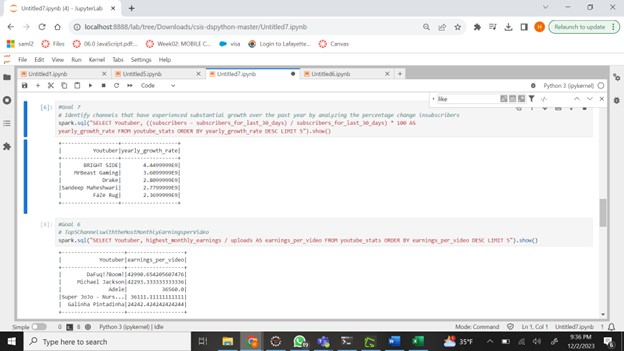
**Visualization:**



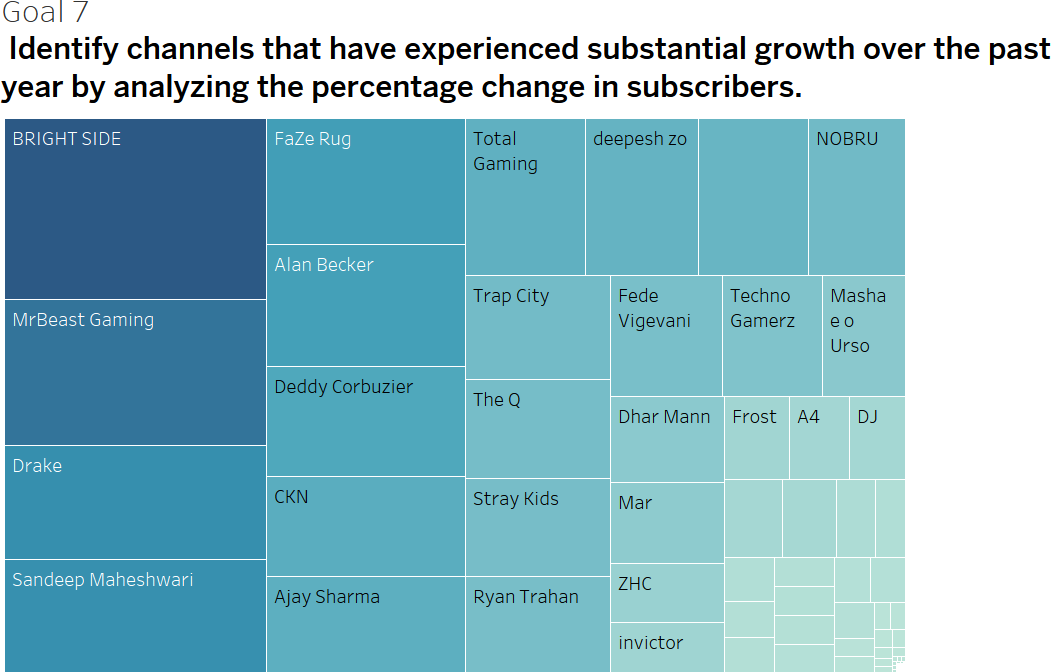
**Goal 7: Identify channels that have experienced substantial growth over the past year by analyzing the percentage change in subscribers.**

This analysis identifies YouTube channels that have experienced substantial growth over the past year by calculating the yearly growth rate in subscribers. The top 5 channels with the highest yearly growth rates are shown. For example, "BRIGHT SIDE" has an extraordinary yearly growth rate of approximately 4.45 billion percent, indicating a substantial increase in subscribers over the past year.

**Screenshot:**



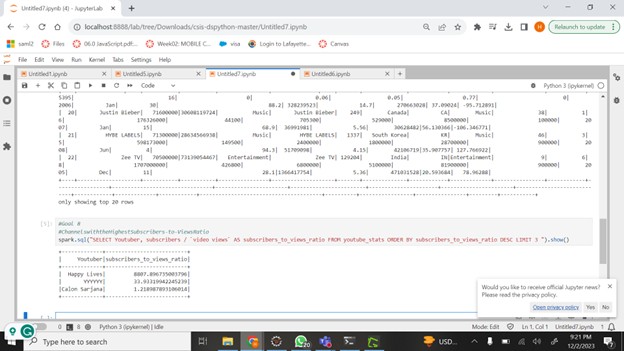
**Visualization:**



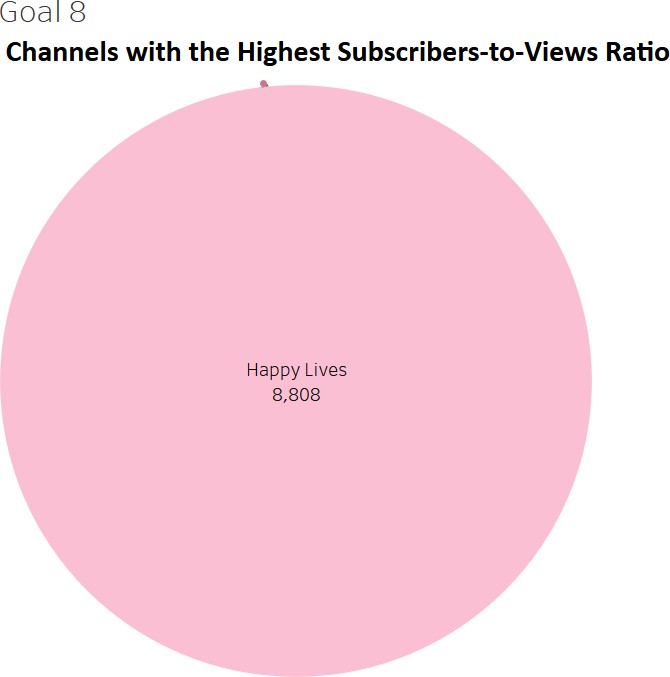
**Goal 7: Channels with the Highest Subscribers-to-Views Ratio**

The analysis explores channels with a high subscribers-to-views ratio, indicating strong audience engagement. In the top 3, "Happy Lives" stands out, suggesting a dedicated subscriber base despite a comparatively lower total view count.

**Screenshot:**



**Visualization:**



# Conclusions

In conclusion, the” Unraveling YouTube Channel Dynamics” project aims to uncover patterns and trends in YouTube channel success, addressing goals like

identifying high-earning channels, assessing subscriber engagement, and analyzing content category performance.The outlined high-level architecture ensures a systematic approach from data import to documentation using LaTeX. We looked at earnings, engagement, and growth, digging into the data with a keen eye for quality, speed, and security. This project not only find what makes YouTube channels successful but also sets the stage for future implementations. The project’s iterative nature, involving data processing and comprehensive analysis, reflects a meticulous strategy for extracting meaningful information.

# References

Github repo: [GitHub Link](https://github.com/sushmitha01/BigData) Kaggle Dataset: [DataSet Link](https://www.kaggle.com/datasets/nelgiriyewithana/global-youtube-statistics-2023)