FINAL PROJECT REPORT

1. Introduction

The "YouTube Trending Video Analytics" project aims to uncover meaningful insights from YouTube's trending video data across various countries. YouTube, being the largest video-sharing platform globally, offers a vast amount of data related to video views, likes, comments, and trends across regions and categories. Analyzing this data helps understand what type of content goes viral, how user engagement varies across countries, and which channels dominate the trending space.

This project was completed as part of a data analytics internship at Elevate Labs, focusing on end-to-end analysis—from data collection to visualization. The dataset used was sourced from Kaggle and consisted of multiple CSV files representing trending videos across various countries. Due to technical constraints in using SQL, the data aggregation and analysis were performed using Python (pandas), and visualization was done in Tableau.

2. Abstract

This project explores global YouTube trending data to uncover insights into video performance across different countries and categories. The dataset includes attributes like video titles, views, likes, comments, category IDs, regions, and trending dates.

Initially, the plan was to use SQL to perform the data analysis, but due to import issues with .sql files, we adapted and shifted the entire analysis pipeline to Python using the pandas library. The CSV files from multiple countries were combined into a single master dataset to simplify analysis. We then calculated various metrics such as total and average views, likes, and comment counts per category and region, identified the most trending channels, examined daily trending behavior, and computed view-to-like ratios.

The results were visualized using Tableau. The dashboard features four main charts:

1. **Regional Category Popularity Heatmap**
2. **Category-wise Segment Bubble Chart**
3. **Top Trending Channels Bar Chart**
4. **Views-to-Like Ratio Chart**

These charts collectively provide a clear picture of what content performs well in different parts of the world and how viewers engage with it. The dashboard is designed for ease of use, allowing users to extract insights at a glance.

3. Tools Used

This project utilized a range of tools and platforms to collect, clean, analyze, and visualize the data. Below is a list of the key tools used and their specific purposes:

1. Python (Jupyter Notebook)

- **Libraries Used:**
    - pandas – For data cleaning, merging, and manipulation.
    - matplotlib & seaborn – Used during exploratory data analysis for visual previews before moving to final visualization.
- Python played a central role in performing all SQL-like operations due to import errors with the original SQL setup.

2. Tableau

- Used for building the final interactive dashboard.

- Enabled the creation of a visually engaging summary of insights.

3. Microsoft Excel / Google Sheets

- Used occasionally to inspect and preprocess small portions of the dataset before importing into Python or Tableau.

4. Steps Involved in Building the Project

Phase 1: Defining the Project Scope

- The project aimed to analyze YouTube trending video data to uncover insights on content popularity across regions and categories.

Phase 2: Data Collection & Preparation

- Multiple regional YouTube datasets were merged into one master file, youtube_trending_master.csv.
- Data cleaning was done in **Python**, handling null values, date conversion, and ensuring uniformity in column names.

Phase 3: Data Analysis (SQL Operations in Python)

- Analysis included:
    - **Groupings** by region and category to calculate total and average views, likes, and comments.
    - Identifying the **most trending channels**.
    - Analyzing **daily trends** and calculating the **view-to-like ratio** for engagement.

Phase 4: Data Visualization (Tableau)

- Created four key visualizations:
    1. **Heatmap of Regional Category Popularity**
    2. **Category-wise Segment Bubble Chart**
    3. **Top Trending Channels (Bar Chart)**
    4. **Views-to-Like Ratio Chart**
- The dashboard was titled **"Inside YouTube: Visual Insights into Global Video Trends"** with detailed descriptions for each chart.

Improvisations & Challenges

- SQL import issues led to **all analysis being done in Python**.
- Focused on creating clear, actionable insights through the dashboard's visualizations.

Conclusion

This project analyzed YouTube trending data, providing insights into video popularity across regions and categories. Despite initial SQL import issues, we used Python to complete the analysis and built an informative Tableau dashboard. The dashboard highlights key trends, top channels, and engagement metrics. Future improvements could include real-time data integration for more up-to-date insights.