Introduction

YouTube has become one of the most powerful platforms for digital content creation and entertainment, influencing millions of viewers across the globe. Understanding the patterns behind trending videos helps creators, marketers, and analysts uncover what type of content gains the most traction.

This project, **"YouTube Trending Video Analytics,"** aims to explore global YouTube data to analyze trends in video performance, audience preferences, and sentiment patterns across regions. By leveraging data analytics tools such as Python, SQL, and Power BI, this study focuses on uncovering the relationship between video categories, audience engagement, and emotional tone in video titles.

Abstract

The project analyzes trending YouTube datasets collected from various countries to identify factors influencing video popularity. The dataset includes attributes such as channel names, categories, subscribers, views, uploads, and sentiment of video titles.

Data preprocessing and exploratory analysis were performed using **Python**, while **Power BI** was utilized for visualization and storytelling. The project also integrates **sentiment analysis** to understand how the tone of video titles affects audience engagement.

The results reveal that entertainment and music-related videos dominate global trends, with positive sentiment titles generally attracting more views and engagement. Through region-wise and category-wise dashboards, the project presents clear visual insights into YouTube content performance.

Tools and Technologies Used

Python (Pandas, Matplotlib, Seaborn, TextBlob): Data cleaning, preprocessing, visualization, and sentiment analysis

SQL: Ranking and aggregation of categories based on average views

Power BI: Dashboard creation, interactivity, and regional comparison

Jupyter Notebook: Python-based data analysis environment

CSV Dataset: Source file containing YouTube channel and video statistics.

Steps Involved in Building the Project

Step 1: Data Collection

The dataset containing YouTube channel details (subscribers, views, uploads, country, earnings, etc.) was obtained from publicly available sources.

Step 2: Data Cleaning and Preprocessing (Python)

- Removed duplicates and missing values

- Standardized numeric columns like subscribers, video_views, and uploads
- Converted date columns (created_date, created_month, created_year) into proper formats
- Ensured data consistency across multiple regions

Step 3: Sentiment Analysis

- Used **TextBlob** to analyze video titles and assign sentiment polarity (Positive, Neutral, Negative)
- Created a new column title_sentiment to represent this sentiment classification

Step 4: Exploratory Data Analysis

- Visualized category and regional distributions using Matplotlib and Seaborn
- Identified top-performing YouTubers and categories

Step 5: Data Visualization (Power BI)

Created four interactive dashboards:

1. **Overview Dashboard** – Key metrics (total channels, views, uploads, average sentiment)
2. **Regional Insights Dashboard** – Comparison of countries by views, uploads, and earnings
3. **YouTuber Performance Dashboard** – Top 10 creators by subscribers and views
4. **Sentiment Insights Dashboard** – Impact of sentiment on views and subscribers

Step 6: Insights and Storytelling

Combined quantitative and visual insights to identify global trends, high-performing categories, and regional differences in YouTube content popularity.

Conclusion

The **YouTube Trending Video Analytics** project successfully demonstrates how data analytics and visualization techniques can be used to uncover meaningful insights from large-scale social media data. The analysis revealed that **Entertainment and Music** are the most dominant categories globally, while **positive sentiment titles** tend to generate more audience engagement.

By integrating **Python for data preparation**, **SQL for ranking**, and **Power BI for visualization**, the project highlights the complete workflow of a real-world data analytics pipeline. This approach not only enhances understanding of global content trends but also provides a framework for future predictive modeling and recommendation systems in the digital media domain.