

# YouTube Trending Analytics

Revo Berliana





# Revo Berliana

## Education

*Information System – Gunadarma  
University*

## Working

*IT Operation Engineer*

# Overview Project

## 1. Introduction

- Project Background
- Problem Statement
- Project Overview

## 2. Data Platform & Understanding

- Data Platform
- Sturcture Project

## 3. Transformation & Data Modelling

- Transformation Proses
- Visualization Dashboard

## 4. Conclusion


On the left side of the slide, there are three overlapping geometric shapes: a large black parallelogram at the top, a medium-sized light orange parallelogram in the middle, and a smaller dark orange parallelogram at the bottom. All shapes are slanted to the right.

# Project Background

## Project Background

YouTube has become a major platform for content creators, brands, and marketers to reach a global audience. Understanding trends and video performance is a key challenge, especially on a large scale.

Manual data tracking is inefficient, which is why an automated system is needed to fetch, store, and analyze YouTube data

A large black parallelogram is positioned on the left side of the slide. Overlapping its bottom edge are two smaller parallelograms: a light orange one in front and a darker orange one behind it.

# Problem Statement

## Problem Statement

This project aims to automate and batch the process of collecting, storing, and transforming YouTube API data for analysis. The main challenges include:

- Automation & Consistency
- Structured Data Management

## Project Goals

- Automate daily data extraction from YouTube API using Airflow (Docker).
- Store structured data in BigQuery for easy access.
- Transform data with dbt to ensure quality and usability.
- Enable visualization and analysis in Google Data Studio.

A large black parallelogram is positioned on the left side of the slide. Below it, two overlapping parallelograms in shades of orange and yellow are also positioned on the left, creating a layered, geometric effect.

# Data Platform Understanding





**Data Sources**



Apache  
**Airflow**



**Processing Layer**



**Storage Layer**



**Visualization**

## 📁 Final Project

- └ 📁 dags/
  - └ 📄 youtube-api-load-to-bigquery.py (DAG fetch data, load to bigquery)
- └ 📁 data/ (Stores temporary and raw JSON files)
- └ 📁 logs/
- └ 📁 youtube\_trending/models/ (Contains dbt models for data transformation)
  - └ 📁 staging/ (Cleans and structures raw YouTube data)
    - └ 📄 stg\_youtube\_data.sql (Processes trending video data)
  - └ 📁 marts/ (Fact tables for analytics and reporting)
    - └ 📄 fact\_video\_performance.sql (Aggregated video performance)
    - └ 📄 fact\_channel\_performance.sql (Aggregated channel insights)
- └ 📁 scripts/
- └ 📄 profiles.yml (dbt connection config for BigQuery)
- └ 📄 docker-compose.yaml (Docker setup for running Airflow locally)
- └ 📄 requirements.txt (Python dependencies for the project)
- └ 📄 .env
- └ 📄 .gitignore
- └ 📄 youtube-api-key.json (API key for YouTube Data API access)
- └ 📄 gcp-key.json (Service account key for BigQuery authentication)

A large black parallelogram is positioned on the left side of the slide. Overlapping its bottom edge are two orange parallelograms, one in a lighter shade and one in a darker shade, creating a layered effect.

# Data Understanding

## Data Sources & Collection

- Source: YouTube API (Trending videos)
- Collected Data(endpoint): Video ID, title, views, likes, comments, channel info, category, duration, etc.
- Extraction Method: Airflow DAG (fetches data daily, saves as JSON)


## Data Processing & Storage

Processing:

- Cleans and normalizes data (removes nulls, standardizes tags)
- Converts ISO 8601 duration to seconds using isodate library for duration

Storage:

- BigQuery

A large black parallelogram is positioned on the left side of the slide. Overlapping its bottom edge are two orange parallelograms, one in a lighter shade and one in a darker shade, creating a layered effect.

# Transformation Consideration & Data Modelling

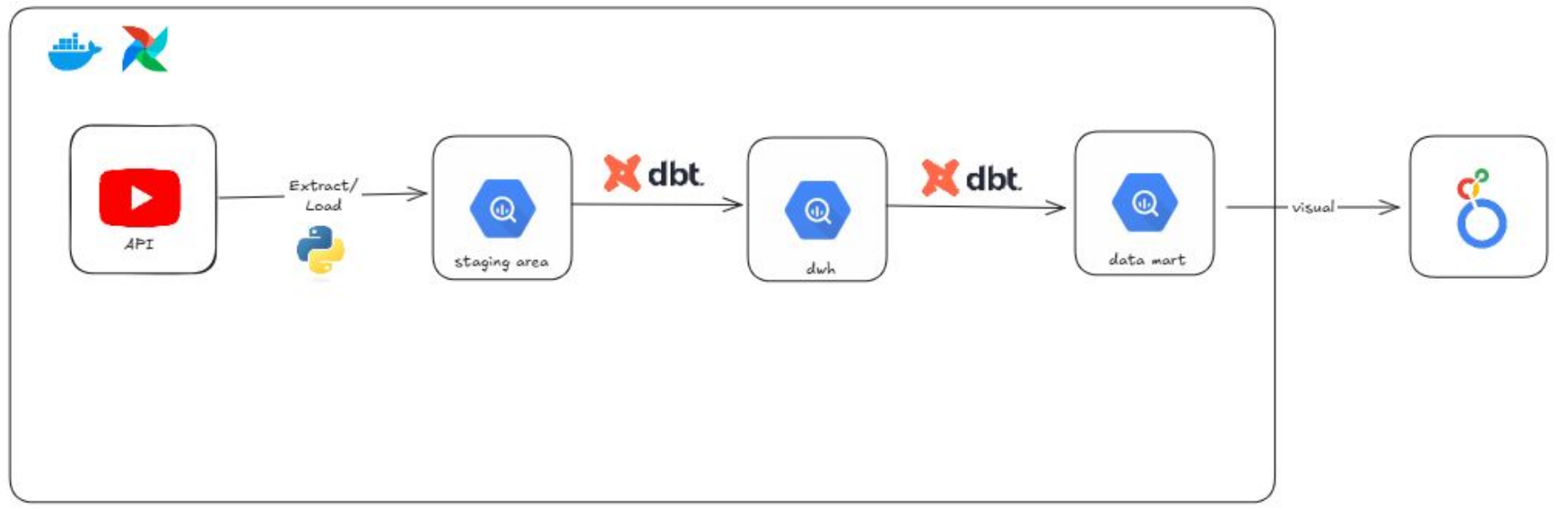
## Transformation Process

### 1. Fact Tables

- Fact Table (fact\_channel\_performance) > ratio of likes and comment and engagement rate
- Fact Table (fact\_video\_performance) > tren to days

### 2. Optimization

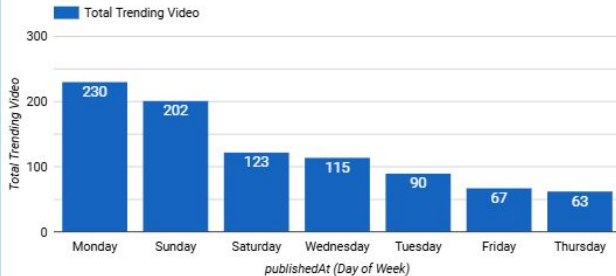
- Partitioning fact\_video\_performance by (publishedAt)
- SAFE\_DIVIDE(ratio) to prevent errors
- Materialized views for reporting



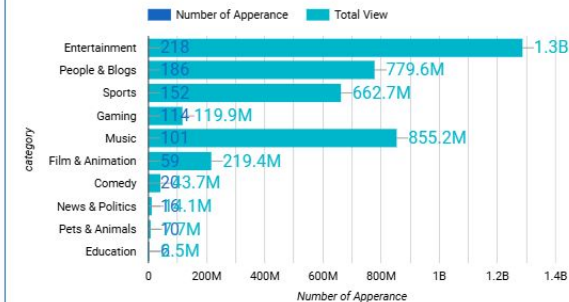
Engagement Rate of Channel

channelTitle	total_views	total_likes	total_comments	avg_like_ratio	Engagement Rate
XODIAC Official	1,004,004	146,640	46,088	0.15	0.19
SEVENTEEN	3,535,893	563,694	34,002	0.16	0.17
BLACKPINK	9,858,060	1,390,238	130,488	0.14	0.15
Mikazuki Arion (AKA ...	2,913,651	330,363	31,641	0.11	0.12
VIVINOS	45,063,741	4,995,711	427,123	0.12	0.12
Aisar Khaledd	10,324,020	1,047,764	172,633	0.11	0.12
SMTOWN	6,842,403	673,168	37,687	0.1	0.1
JISOO	228,093,521	21,345,514	1,224,179	0.1	0.1
SBSKPOP X INKIGAYO	20,391,312	1,884,456	53,156	0.09	0.1
Aqeela Calista	4,233,376	373,794	22,422	0.09	0.09

Trending Video by Publication

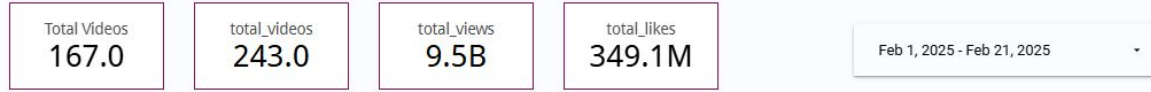


Top 10 Category Based on Trending Video Appearance

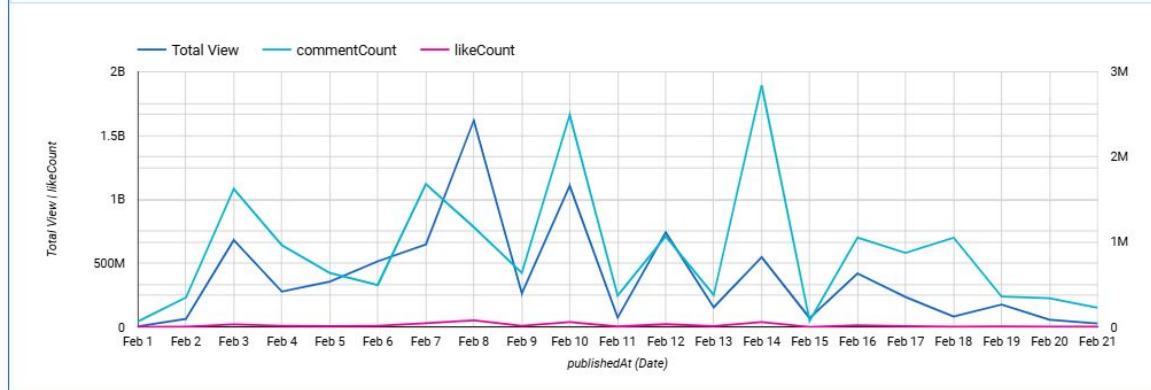




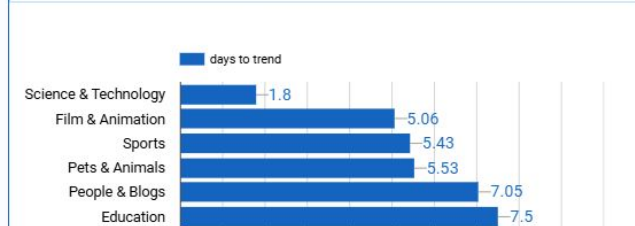
## Trending Youtube Dashboard



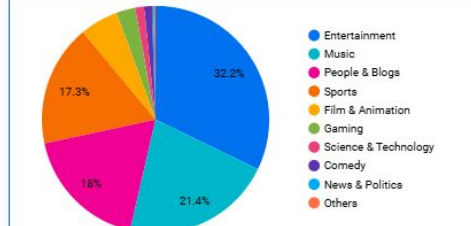
Tren Views based on Publish date



Average Days to Trend per Category



Proportion Video by Category




A large black parallelogram on the left side of the slide, with two overlapping orange parallelograms positioned below it, creating a modern, abstract geometric design.

# Conclusion & Recommendation

## Conclusion

The platform efficiently processes and presents YouTube trending data using Airflow, BigQuery, and dbt. It enables structured storage, transformation, and visualization for engagement analysis.

 Final Takeaway: The system works well for batch analytics but can be improved with real-time processing with ksqlDB , clustering category, and improve dbt models(dbt snapshot-> historical changes, dbt test check missing & inconsistent.



A large, stylized graphic on the left side of the slide. It consists of a blue outline of a person's head and shoulders. Inside the head is a series of concentric circles: a small light orange circle, a medium orange circle, and a large orange circle. The body is a large blue circle, and inside it is a large orange circle, which contains a smaller orange circle in the center.

**Terima  
Kasih.**