YOUTUBE VIDEO ENGAGEMENT ANALYSIS – MINI DATA PIPELINE & KPI DASHBOARD

Author: Amarachi F. Onyedinma-N
Challenge: Day 17 –Python Challenge @ ZION TECH HUB
Tools Used: Python, Matplotlib, Streamlit, Altair, Pandas

INTRODUCTION

With the increasing popularity of online video content, especially in education, it's critical to track performance metrics that reflect not just popularity (views) but also engagement and retention. This mini project explores the performance of five educational YouTube videos using a simplified analytics pipeline.

The goal is to design a lightweight but effective ETL process that extracts raw video data, transforms it to compute key performance indicators (KPIs), and visualizes these metrics in a simple dashboard. This exercise simulates real-world analytics tasks for content strategy and marketing insights.

To expand interactivity, a lightweight web application was also built using **Streamlit**, offering a real-time dashboard experience powered by **Altair** for visualization and **Pandas** for data transformation.

STATEMENT OF THE PROBLEM

Content creators and analysts often struggle to go beyond surface-level video metrics. Views alone don't indicate how much value a video provides or how well it retains audience attention. There's a need to quantify deeper insights like retention, engagement, and conversion to guide better content decisions.

OBJECTIVES

• To simulate a lightweight ETL pipeline using Python.
• To compute critical KPIs such as retention rate, like rate, and conversion rate.
• To design a custom engagement score that reflects both watch time and user interactions.
• To visualize video performance for strategic analysis.
• To generate a summary report accessible to technical and non-technical stakeholders.
• To present KPIs interactively via a **Streamlit dashboard** using **Altair charts and metric cards**.

DATA DESCRIPTION

The mock dataset consists of 5 educational videos, each with metadata including:
• video_id: Unique identifier
• title: Name of the video
• views: Number of total views
• likes: Total likes received
• watch_time: Total minutes watched

METHODOLOGY

The process follows the ETL (Extract, Transform, Load) framework:
• **Extract:** A static list of video data (views, likes, watch time) is loaded.
• **Transform:** KPIs and engagement scores are computed using simple formulas.
• **Load:** A KPI report is generated and saved, along with a visualization for stakeholder review.

ETL PROCESS

• **Extract:** A mock dataset of 5 educational YouTube videos with fields:

– video_id, title, views, likes, watch_time (in minutes)

• **Transform:**

– Calculated key metrics:

• Total & average watch time

• Like rate = likes / views

• Conversion rate = likes / watch time

• Retention rate = total watch time / (views × avg video length)

• Engagement score = (likes × 2) + (watch_time × 0.3)

• **Load:**

– KPIs were printed and saved to a file: `kpi_report_v2.txt`

– Visualization saved as: `engagement_scores.png`

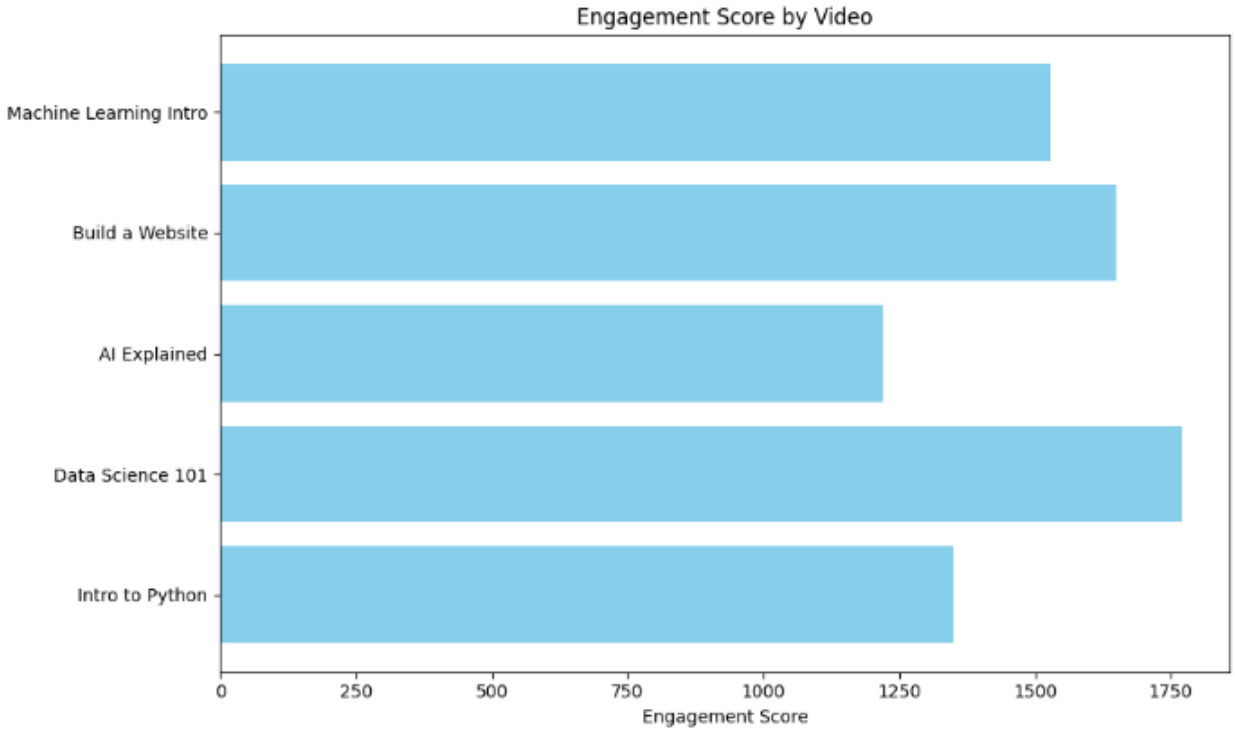
– An interactive version was built using **Streamlit**, featuring **Altair bar charts**, **highlight cards**, and filters for a smoother exploration of the dataset.

KEY PERFORMANCE INDICATORS (KPIS)

Metric	Value
Average Watch Time	2680.00 minutes
Like Rate	0.22
Conversion Rate	0.05 (likes per min watched)
Retention Rate	0.60
Most Popular Video	"Data Science 101" (1800 views)
Highest Engagement Score	"Data Science 101" (Score: 1756.0)

These metrics help identify which videos keep users engaged and which are most effective at converting viewers to active participants.

VISUALIZATION OUTPUT



The horizontal bar chart was created using matplotlib to show the Engagement Score by Video and saved as: `engagement_scores.png`.

This visual helps stakeholders quickly identify high-performing content based on both viewership and interaction metrics.

The interactive version in **Streamlit** enhances this by enabling users to dynamically explore engagement scores using **Altair** visualizations, improving decision-making for educators and content strategists.

Insight:

"Data Science 101" stands out with the highest engagement and views, making it the top performer. In contrast, "AI Explained" shows lower engagement, indicating room for improvement in content clarity or delivery.

Recommendation:

Focus on expanding high-performing topics like Data Science to sustain engagement, while refining lower-performing videos to improve overall viewer experience.

CONCLUSION

This mini data-pipeline project shows how a simple ETL process can turn raw YouTube statistics into useful insights for better content strategy. Instead of just focusing on views, the dashboard highlights engagement as the real measure of success.

1. “Data Science 101” stood out across all key metrics, engagement score, like rate, and viewer retention, likely because the topic is in high demand and the video delivered it in a clear, structured, and relatable way.

2. In contrast, videos like “AI Explained” and “Intro to Python” had lower engagement, possibly due to complex concepts being presented too broadly, or lacking interactivity and pacing that keeps viewers engaged.

3. Retention overall is solid but uneven, highlighting where some videos could improve in storytelling, clarity, or delivery style.

4. The workflow is lightweight but scalable, making it easy to apply to more content or even full channel reviews.

When combined with **Streamlit's real-time capabilities**, this approach creates a solid foundation for scalable, interactive content analytics.

Even with a small dataset, this project shows how tracking the right metrics, not just view counts, can lead to smarter, data-driven content decisions.

RECOMMENDATIONS

1. **Build on what’s working:** The content team should consider creating more videos around top-performing topics like Data Science and Web Development. For example, “Data Science 101” could be expanded into a full series or broken into shorter episodes. Repurposing this content into other formats, such as blog posts, short clips, or LinkedIn carousels, can also extend its reach.

2. **Improve weaker videos:** Revisit videos like “AI Explained” and “Intro to Python” to identify areas for improvement. For “AI Explained,” simplifying complex concepts or breaking it into smaller segments might help. For “Intro to Python,” adding engaging visuals, practical examples, or mini-projects could boost viewer interest and retention.

3. **Segment the audience:** Understanding who is watching (beginners vs. advanced learners) can help tailor content more effectively. The team could analyze viewer patterns and create targeted videos that meet the needs of specific audience segments.

4. **Focus on retention strategies:** Since watch time plays a big role in engagement, the team should aim to strengthen video structure, start with a strong hook, maintain a logical flow, and close with a clear call-to-action. These techniques can help improve full video completions.

5. **Promote high-performing content strategically:** Use successful topics in titles and thumbnails to draw attention to both new and existing videos. Sharing snippets or highlights of popular content across platforms like LinkedIn, Twitter, or Instagram can help attract a broader audience.

6. **Encourage viewer feedback:** At the end of each video, prompt viewers to leave comments or rate the content. Asking what topics, they’d like to see next not only boosts interaction but also helps the team plan future content based on real audience input.