UNITED STATES INTERNATIONAL UNIVERSITY

CLASS DSA 3050A

GROUP 4

Howard Muchaki - 668712

Patrick Keegan - 668964

Shashank Shukla - 666831

ANALYSIS OF YOUTUBE ANALYTICS DATASET

[**1. Business case**](#_30ucqc4tn404) **3**

[1.1 Problem statement](#_vkl1vuflqu6) 3

[**2. Data architecture**](#_cdmi4hoq67sy) **4**

[**3. Key insights**](#_2zfy8wttbxq3) **5**

[**4. ROI Projections**](#_zaoojfxdnrk5) **6**

[**5. Recommendation and conclusion**](#_2oidvee76bhm) **7**

# 1. Business case

A popular YouTuber approached our team seeking guidance on effectively managing their channel. Their primary goals were to minimise subscriber loss, enhance viewer retention, and maximise their return on investment (ROI) through both ad revenue and premium subscriptions. To achieve these objectives, we discussed strategies such as optimising content for engagement, analysing viewer demographics to tailor uploads, and employing effective marketing techniques to promote exclusive content for premium members. By focusing on fostering a loyal community and delivering consistent value, the YouTuber aims to build a sustainable and profitable channel while creating meaningful connections with their audience.

## 1.1 Problem statement

YouTube channels experience fluctuations in subscriber counts based on various operational strategies and content management approaches. Understanding the factors that contribute to unsubscriptions is crucial for content creators, as it impacts both audience retention and revenue generation. This process involves analysing specific elements such as video quality, consistency of uploads, viewer interaction, and the relevance of content to the target audience. Additionally, it's important to evaluate the effects of unsubscriptions on overall revenue streams, including ad revenue, sponsorship opportunities, and merchandise sales. A comprehensive understanding of viewer engagement metrics—such as watch time, likes, comments, and shares—can provide valuable insights into subscriber behaviour and preferences, ultimately guiding creators in optimising their channels to foster loyalty and growth.

## 1.2 Objectives

This study aims to identify the factors that contribute to customer churn within the channelling process. Additionally, it seeks to analyse engagement trends over time related to the channel.

# 

# 

# 

# 2. Data architecture

The provided data is organised in a star schema, a widely used data architecture for data warehousing and business intelligence. This structure is ideal for creating dashboards and reports because it simplifies the querying process and improves performance. The architecture consists of a central fact table and multiple dimension tables.

* **Fact Table:** “**fact\_video\_performance.csv”** serves as the central fact table. It contains the core metrics (facts) of interest for video performance, such as Views, Likes, Shares, Watch Time, and Estimated Revenue. Each row in this table corresponds to a single video's performance metrics.
* **Dimension Tables:** These tables provide context to the facts. The provided data includes the following dimension tables:
  + **dim\_channel.csv**: A very small table that provides details about the YouTube channel itself, such as Channel\_ID, Channel\_Name, Country, and Subscribers\_Total.
  + **dim\_date.csv**: A table that breaks down dates into various components like Day, Month, Quarter, and Year. This is crucial for time-series analysis.
  + **dim\_video.csv**: This table contains descriptive attributes for each video, such as Video Duration, Video Publish Time, Month, Year, and Video Thumbnail CTR (%).

The tables have been constructed after performing data mining using the Apriori Algorithm, which further utilises association rules(in-built function) that extract the frequent itemsets in a dataset. This gives us all the impact each variable has on one another in our dataset, including the interpretation of what makes a YouTube Channel or Video good for the content creators and finding the target audience to then be aware and reduce some of the concerns our client had. From viewership to subscriber retention.

Further, we will look at the star schema and find what has been of help to us

# 3. Key insights

The day of the week that exhibits the highest levels of engagement, measured by likes, shares, and comments, is Tuesday. Analysis indicates that two seasons, specifically spring and winter, experience minimal churn, characterised by a churn probability of less than 0.3. Conversely, the remaining two seasons demonstrate a moderate churn probability, ranging from greater than 0.5 to not exceeding 0.8. Channel growth experienced an upward trajectory over the first three years, reaching its zenith in 2018, while the retention rate remained stable from the onset of the COVID-19 pandemic until 2024.

Factors influencing churn include the publishing season, with summer and fall showing the lowest retention rates. Additionally, the day of the week plays a significant role, as both Tuesday and Sunday are associated with heightened interaction with content. This increased engagement subsequently leads to a high revenue per 1,000 views, with Tuesday generating revenue of $8.04 and YouTube ad revenue amounting to $1,150. Furthermore, August is identified as the month with the least churn, attributed to the majority of school seasons being in recess during this period.

In terms of revenue, Tuesdays and Fridays record the highest average revenue, with the Fall season generating the most revenue overall. From our analysis, the YouTuber can maximise earnings by prioritising posts on Tuesdays and Fridays and focusing content around the Fall and Summer seasons, while using the Winter season for strategic planning or upskilling.

Additionally, ad impressions peak within the first three days of publishing. To capitalise on this, the YouTuber should consider investing in sponsored video visibility during this period and employ strategies such as teasers and countdowns before publishing. These actions can help maximise views and, consequently, ad impressions, further driving revenue growth.

# 4. ROI Projections

# 5. Recommendation and conclusion

To effectively reduce churn, it's recommended to adjust the upload schedule. Transition from daily uploads to specific days, focusing on Tuesdays, Wednesdays, and Sundays, while increasing the volume of videos. During the fall season, consider posting more frequently on Sundays and Wednesdays to further lower the probability of churn. In the winter months, aim to enhance content availability by uploading more on Sundays, Tuesdays, and Fridays.