



University of New Haven

TAGLIATELA COLLEGE OF ENGINEERING

Electrical & Computer Engineering and Computer Science



Distributed And Scalable Data Engineering

Spring 2024

Team 1

CONTENTS

Project Title2

Abstract3

Introduction.....3

Executive Summary.....4

Highlights of the Project5

Available Research5

Methodology6

Data Pipeline6

Results7

Discussions8

Conclusion9

The Predictive Analytics Dashboard provides insights on video content trends and viewer engagement patterns. It analyzes trending topics, formats, and viewer preferences to help creators align their content strategy. Additionally, it examines engagement metrics to optimize content for maximizing reach and growing the audience effectively.



Ashish Agarwal
Harikrishna Para
Raja Reddy Dondeti
Vaishnavi Kukkala

Abstract

The project aims to develop a Predictive Analytics Dashboard that provides insights into video content trends and viewer engagement patterns on YouTube. It leverages the YouTube Data API v3 to extract semi-structured data and analyze it using various AWS technologies. The dashboard offers two main components: Trending Videos Analysis and Engagement Patterns. The former identifies popular trends in video topics, formats, and content styles, keeping creators informed about what's capturing viewers' attention, allowing them to adapt their content strategy accordingly. The latter analyzes viewer engagement metrics based on video length, upload frequency, and publishing times, helping understand how audience preferences are evolving and enabling creators to maximize engagement and reach a wider audience.

Introduction

In the ever-evolving world of online video content, creators often struggle to keep up with changing audience preferences and adapt their content strategy accordingly. The Predictive Analytics Dashboard project aims to address this challenge by providing a comprehensive solution that leverages data analytics and the YouTube Data API v3.

The project team consists of experienced professionals in the fields of data analysis, research, engineering, and science, each bringing their unique expertise to the table. The primary objective is to develop a dashboard that offers valuable insights into trending video topics, formats, and content styles, as well as viewer engagement patterns.

By understanding what captures viewers' attention and how audience preferences are changing, creators can make informed decisions and adapt their content strategy to align with these trends. The dashboard will also enable creators to maximize engagement and reach a wider audience by analyzing metrics such as video length, upload frequency, and publishing times.

Executive Summary

The Predictive Analytics Dashboard project addresses the challenges faced by video creators in keeping up with evolving audience interests, optimizing content, and developing long-term strategies. By leveraging data analytics and the YouTube Data API v3, the project provides valuable insights into:

- Trending videos
- Content styles
- Engagement patterns

The dashboard empowers creators to:

- Make informed decisions.
- Produce more engaging content.
- Grow their audience effectively.

The project follows the CRISP-DM methodology and utilizes various services, including:

- Youtube API
- Amazon Event Bridge
- AWS Lambda
- Amazon S3
- CloudWatch
- Docker
- Amazon DynamoDB
- Amazon EC2
- Windows
- Power BI

These technologies ensure:

- Scalability
- Secure storage
- Monitoring

Highlights of Project

1. Identifies current popular trends in video topics, formats, and content styles.
2. Analyzes viewer engagement metrics like video length, upload frequency, and publishing times.
3. Helps understand how audience preferences are evolving over time.
4. Allows creators to adapt their content strategy to align with trends.
5. Enables creators to maximize engagement and reach a wider audience.

Review of available research

There is extensive research in the field of video content analysis and audience engagement prediction. Several studies have focused on leveraging data analytics and machine learning techniques to identify trending topics, analyse user behaviour, and predict engagement metrics.

One relevant study, "Predicting Viewer Engagement on YouTube using Multimodal Data" (Zhao et al., 2019), proposed a deep learning model that combines visual, auditory, and textual features to predict video engagement metrics such as likes, dislikes, and comments. Another study, "Trending YouTube Video Statistics" (Khan, 2017), analysed trending video data to identify patterns and correlations between various video attributes and their popularity.

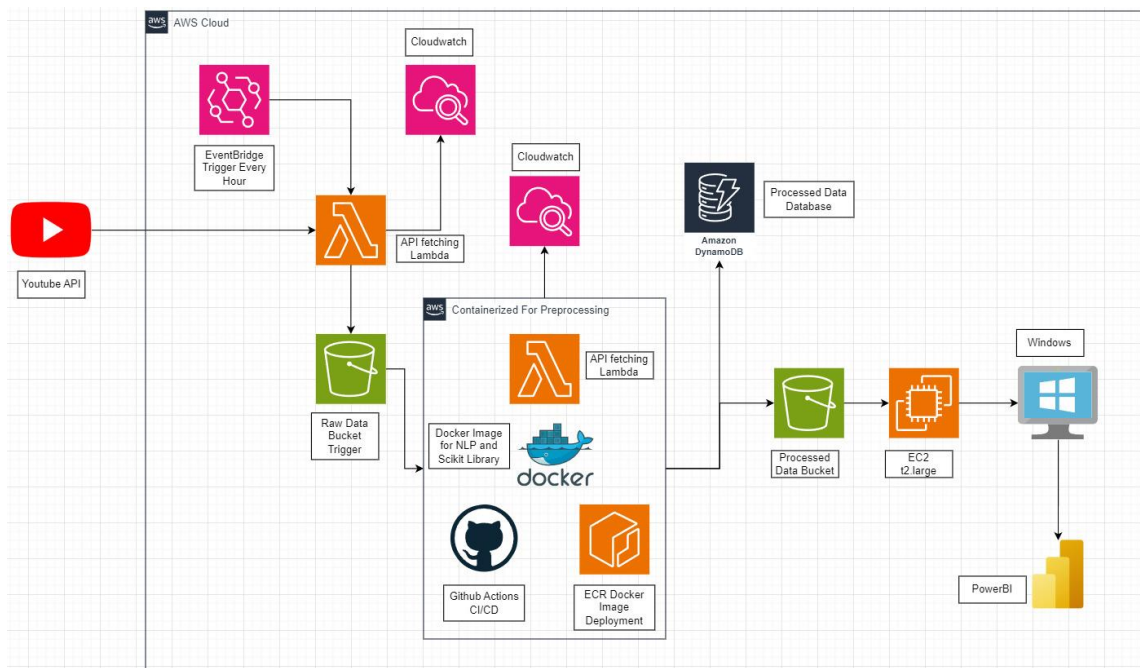
Additionally, research has been conducted on optimizing content strategies based on audience preferences and engagement patterns. For instance, "Optimizing Video Content for Viewer Engagement" (Li et al., 2020) explored techniques for recommending optimal video lengths, upload frequencies, and publishing times to maximize viewer engagement.

Methodology

The Predictive Analytics Dashboard project follows the CRISP-DM (Cross-Industry Standard Process for Data Mining) methodology, which is a widely adopted framework for data mining and analytics projects. The project execution phases include:

- Data Extraction:** Extracting semi-structured data from the YouTube Data API v3.
- Data Transformation:** Cleaning, formatting, and preprocessing the extracted data.
- Data Loading:** Loading the transformed data into a secure and scalable data storage solution (e.g., AWS DynamoDB, S3).
- Secure Storage:** Implementing secure storage and access controls for the data.
- Monitoring:** Setting up monitoring and logging mechanisms to ensure data integrity and system performance.
- Data Analysis:** Performing exploratory data analysis, statistical modelling, and machine learning techniques to uncover insights and trends.
- Data Visualization:** Creating interactive visualizations and dashboards (e.g. PowerBI) to present the findings in a user-friendly manner.

Data Pipeline



Results

Visual 1

The pie chart shows the distribution of the sum of views and likes across the different video categories. Unboxing videos account for the largest portion, followed by educational videos and comedy videos.

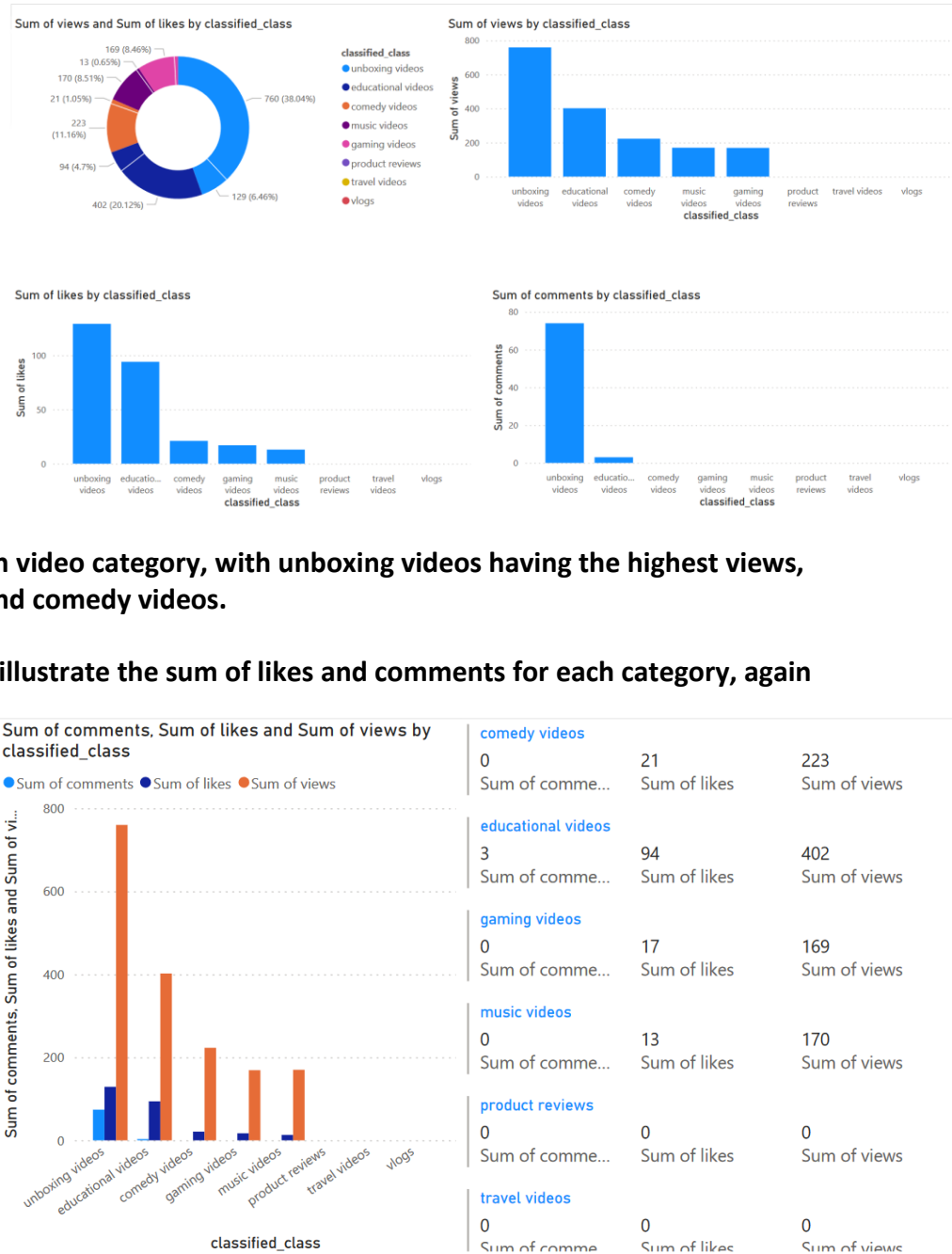
The bar chart on the right displays the sum of views for each video category, with unboxing videos having the highest views, followed by educational videos and comedy videos.

The smaller charts at the bottom illustrate the sum of likes and comments for each category, again showing unboxing videos and educational videos leading in terms of likes, while unboxing videos have the highest number of comments.

Visual 2

This bar chart provides a clear comparison of the sum of comments, likes, and views for each video category. It confirms the insights from the previous visualizations, with unboxing videos and educational videos leading in terms of views and likes, while unboxing videos also have the highest number of comments.

The table on the right shows the actual values for each metric (sum of comments, likes, and views) for each video category.



Discussion

The further discussions section highlights potential areas for enhancing and expanding the Predictive Analytics Dashboard project:

1. Integrating data from other video-sharing platforms to gain a more comprehensive understanding of audience behavior and content trends across different platforms.
2. Developing personalized content recommendations based on individual viewer preferences and viewing history to improve engagement and retention.
3. Incorporating sentiment analysis techniques to analyze viewer comments and feedback, providing insights into audience sentiments and aiding content strategy optimization.
4. Exploring features that facilitate collaboration and community building among creators and their audiences, fostering increased engagement and loyalty.
5. Addressing ethical concerns related to data privacy, algorithmic biases, and responsible use of audience data, ensuring the project's ethical and responsible implementation.

These enhancements and considerations aim to contribute to the continuous improvement and evolution of the project, keeping it relevant and valuable in the rapidly changing online video content landscape.

Conclusion

In conclusion, the Predictive Analytics Dashboard project represents a comprehensive solution to address the challenges faced by video creators in keeping up with evolving audience interests, optimizing content, and developing long-term strategies. By leveraging the power of data analytics, the YouTube Data API, and cutting-edge AWS technologies, this project empowers creators with valuable insights into trending video topics, formats, content styles, and engagement patterns. The dashboard equips creators with the knowledge and tools necessary to make informed decisions, adapt their content strategies, maximize engagement, and reach a wider audience. As the online video landscape continues to evolve rapidly, the Predictive Analytics Dashboard stands as a versatile and scalable platform, poised to remain relevant and valuable by continuously incorporating advancements in data analysis techniques, emerging technologies, and industry best practices. Ultimately, this project represents a significant step towards enhancing the success and growth of video creators, fostering a more engaging and fulfilling experience for both creators and viewers alike.

For more information: Visit the Github Repo for the project [here](#)!