TEAM ENIGMA

Project Streaming Analytics on YOUTUBE

Members Presenting:
Abhishek , Jagan

Motivation

Motivation:

- The availability of a large collection of films from around the world is the key argument for choosing YouTube.
- Using a streaming analytics platform will enable to make faster and more informed decisions.
- Businesses can use data-driven decision-making to generate real-time insights and projections in order to improve their performance.
- Content Creators and Users gain a big advantage with the analysis of videos that are being uploaded on YouTube.

Problem Statement And Research Questions

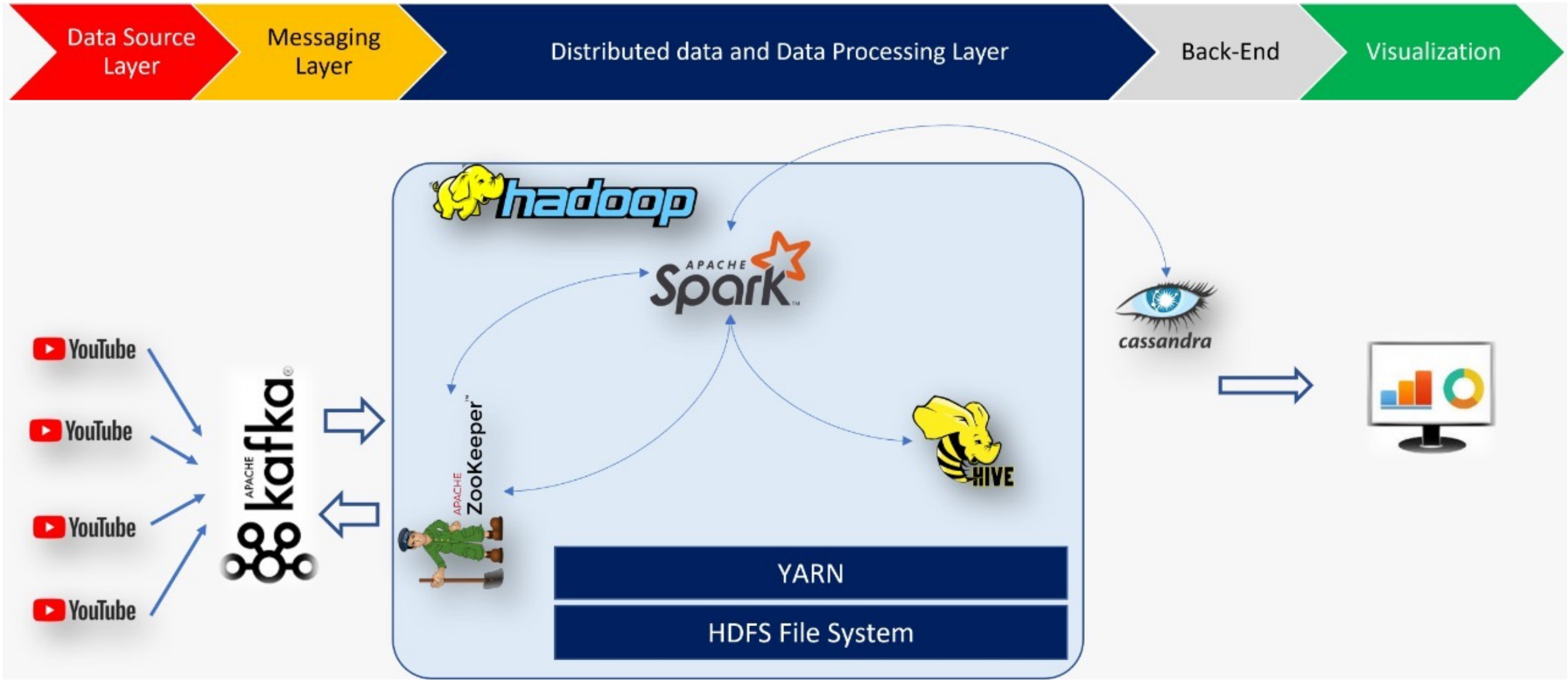
Problem Statement

- Traditionally, querying and retrieving on the Datawarehouse is a costly operation.
- In the area of Streaming, there are no significant improvements.
- Especially, processing and streaming data objects like videos, gaming, health care etc.
- Our objective is to create and implement an analytical platform on top of the YouTube videos we collect in order to gain valuable insights and uncover hidden patterns in the streaming world.

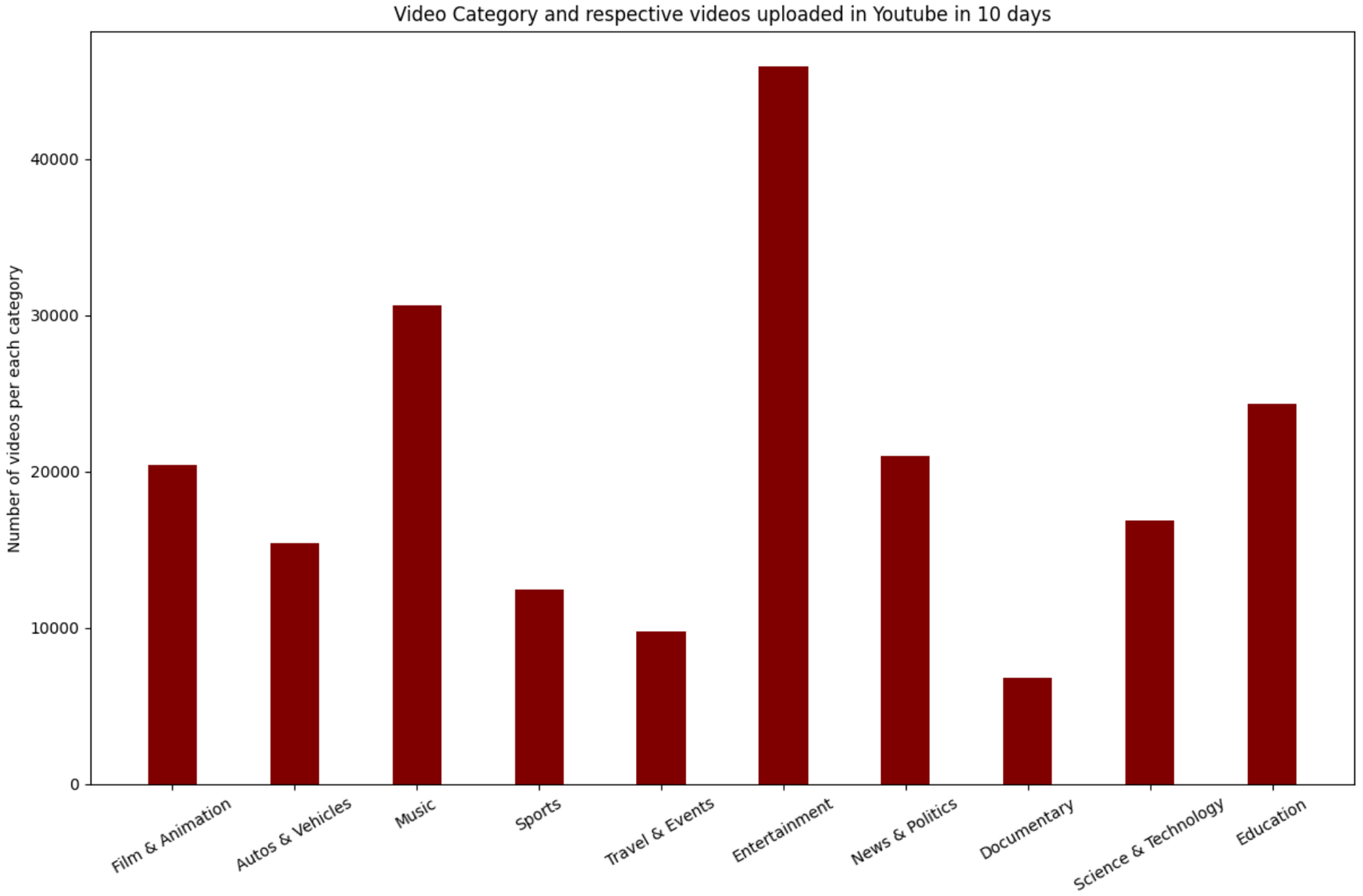
Research Questions

- How the Big data Architecture is useful in the context of Streaming Analytics?
- How to Perform Streaming video analytics on YouTube?
- How to design and implement a Scalable framework that can handle and process Big data?

Approach


Data Source Layer
Messaging Layer
Distributed data and Data Processing Layer
Back-End
Visualization
YouTube
YouTube
YouTube
YouTube
APACHE kafka
hadoop
APACHE Spark
APACHE ZooKeeper
HIVE
YARN
HDFS File System
cassandra

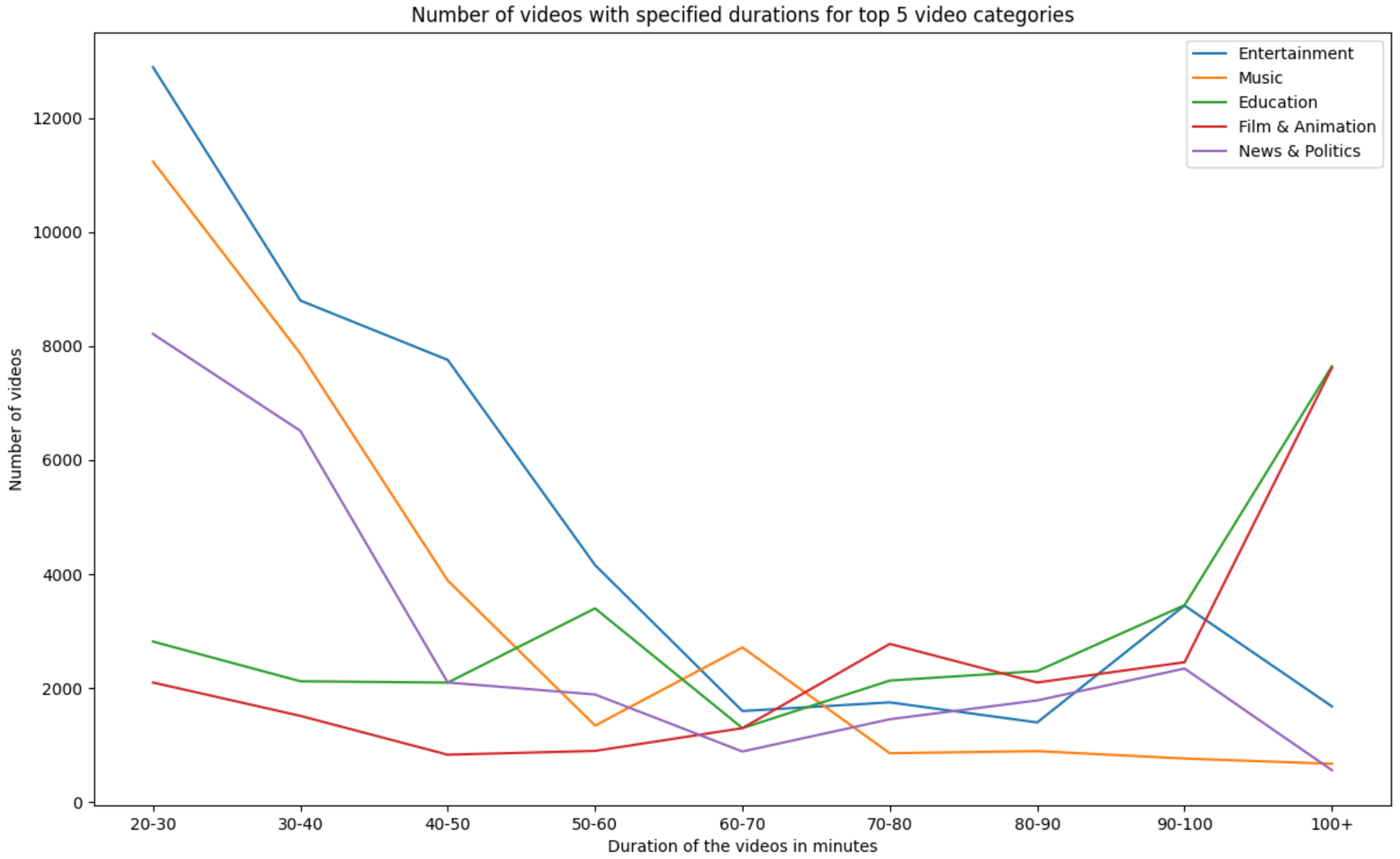
Final Results



Plot between categories and number of videos per category :

- After analysis of the collected videos from the APIs, it is observed that the videos related to entertainment are obtained the most among all the different categories.

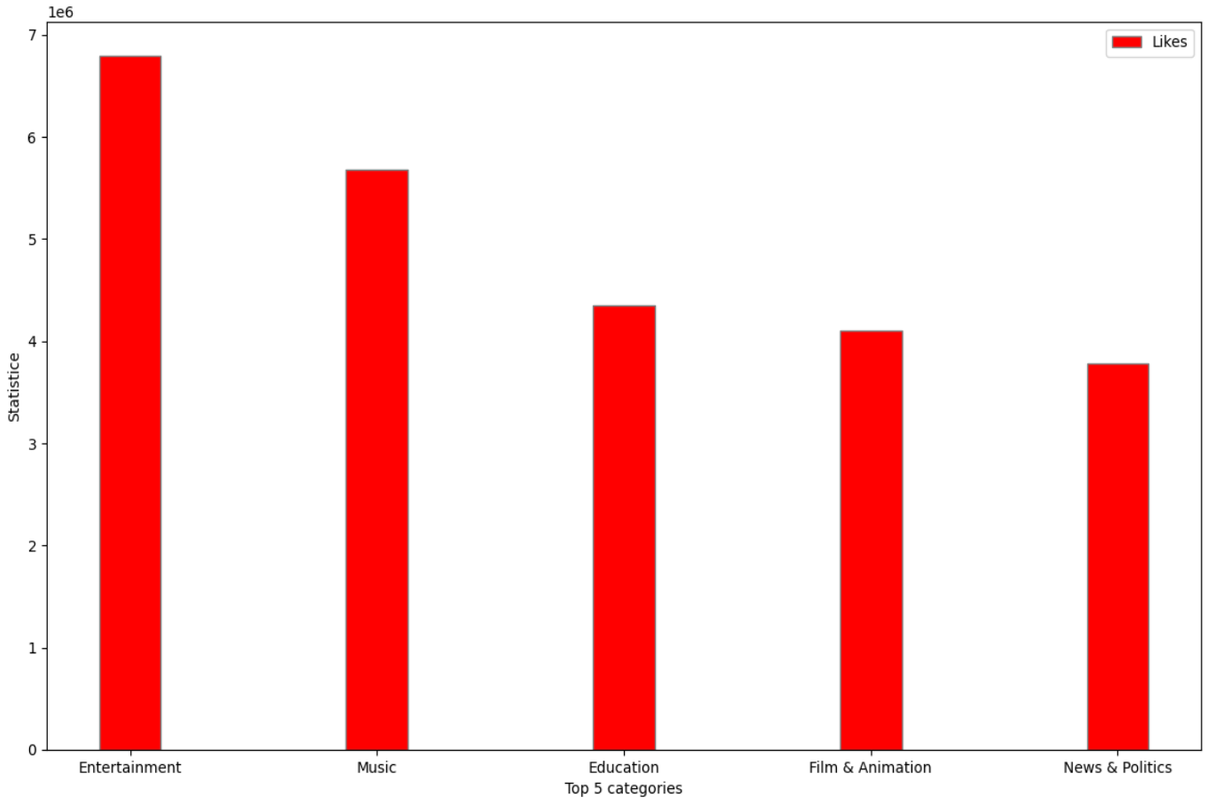
Final Results



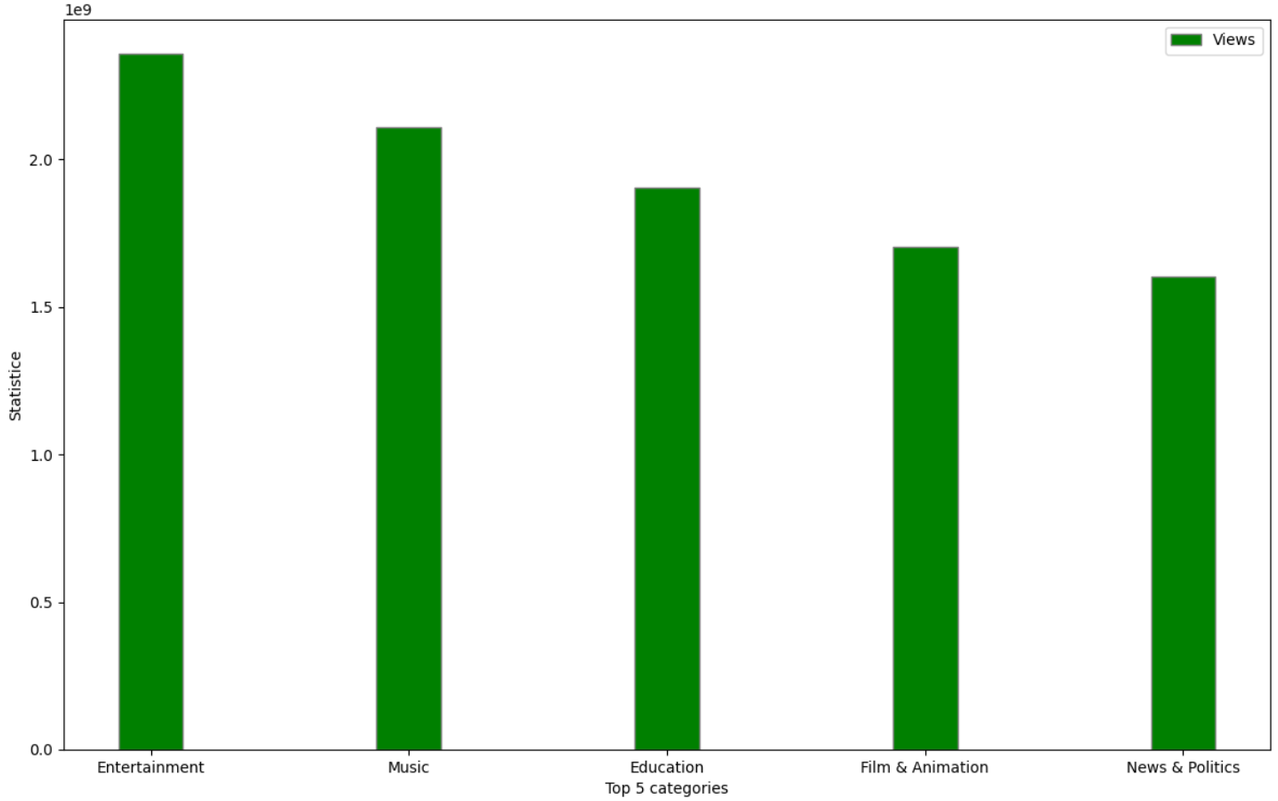
Plot between duration of videos and videos per category :

- After analysis of the collected videos from the APIs, it is observed that the there are more lengthy videos in Film & Animation and and Education and less lengthy videos in other categories.

Final Results



Plot between number of likes and top 5 categories



Plot between number of views and top 5 categories

Real World Impact & Future Research Opportunities

Potential Real-World Impact of the project's findings:

- This will be a major helping hand to the video creators because these insights will help them to better understand what type of videos are receiving reach and engagement.
- Our end goal through this project is to demonstrate the impact of videos, so viewers can pick up the top trending videos and can watch them.

Future Research Opportunities:

- Application of Machine Learning and a Natural Language Processing techniques on the Video Titles and Description to understand the videos better.
- Creation of a dashboard to visualize the analytics.
- Extending it to a recommendation system which aligns to the user's interests.