ANALYZING THE POPULARITY OF GAMING VIDEOS

Supervised by: L. Reem Alqifari

INTRODUCTION

Gaming content dominates a significant portion of YouTube, drawing millions of views daily. This project aims to analyze the factors that influence the popularity of gaming videos, helping creators, marketers, and platforms understand what drives user engagement. By exploring engagement metrics such as likes, comments, and views, we seek to uncover patterns behind why certain videos go viral while others are overlooked.

OBJECTIVES

We aim to answer the following questions:

- How does the upload time impact the video's popularity?
- Is there a correlation between the number of comments and the number of views?
- Which gaming channels are the most popular in terms of engagement metrics?
- What is the relationship between video duration and the number of views?
- Do videos with certain tags (e.g., "funny," "gameplay," "walkthrough") receive more engagement?

DATA COLLECTION

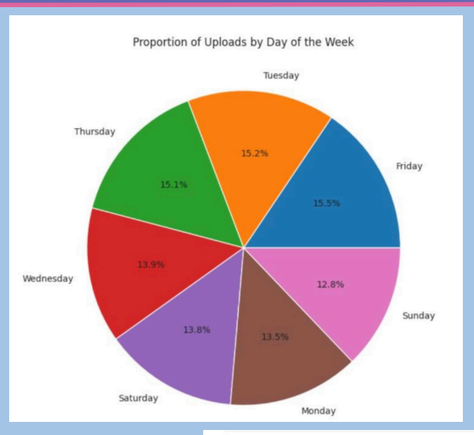
We collected data using the YouTube Data API v3, filtering for videos under the Gaming category.
Each row in the dataset represents a single gaming video and includes features such as:

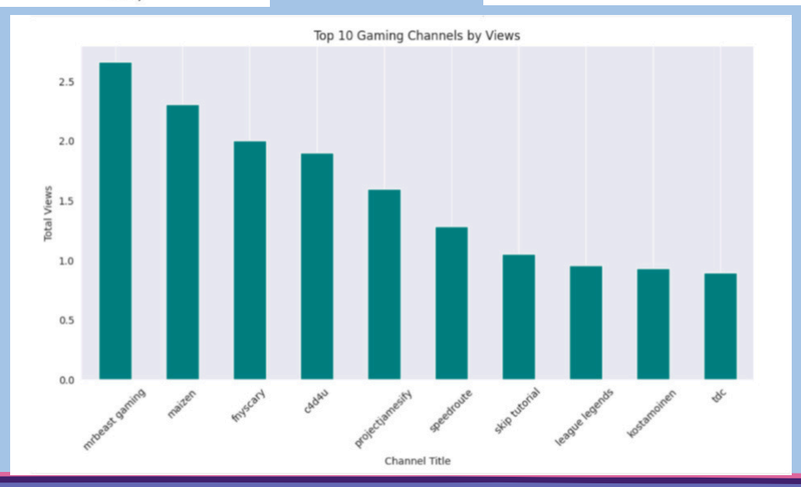
- title, tags, description (text features)
- views, likes, comment_count (engagement metrics)
- published_at (upload time)
- duration

DATA ANALYSIS

To understand our data, we performed exploratory data analysis on key attributes related to video performance, and we found that:

• Videos uploaded between 6 PM and 9 PM UTC, especially on Fridays and Saturdays, receive the highest average views.
• There is a moderate positive correlation between comments and views, suggesting that user interaction enhances video visibility.
• The top 10% of gaming channels account for nearly half of all views, indicating a high concentration of success among a few dominant creators.





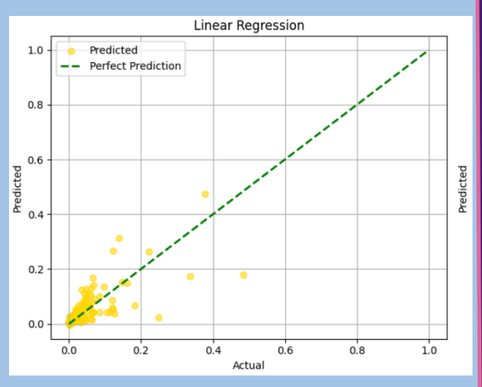
MODELS AND FINDINGS

To predict the number of views on gaming videos, we developed and evaluated three regression models using features such as likes, comment count, and video duration:

- Linear Regression (baseline model)
- Random Forest Regressor
- XGBoost Regressor

To evaluate model performance, we used:

- Mean Squared Error (MSE)
- R^2 Score (Coefficient of Determination)



We concluded that Linear Regression was the top performer, providing the best accuracy and data fit. More complex models like Random Forest and XGBoost underperformed, suggesting that the baseline model was more suitable for our data.

CONCLUSIONS

- Engaging gaming videos with high numbers of likes and comment counts tend to attract the most views, highlighting audience interaction as a key driver of popularity. Among the models tested, Linear Regression performed best in predicting video views, indicating that straightforward relationships between features like likes, comments, and duration are effective for modeling.

- Clustering analysis using KMeans (optimal at 4 clusters) further revealed that viral videos are typically short, highly liked, and not overloaded with tags. On the other hand, videos that are too long or overly tagged tend to receive less engagement. These findings suggest that creators should focus on concise, interactive, and well-targeted content to maximize visibility and user engagement on YouTube.