King Saud University
College of Computer and Information Sciences
Department of Information Technology

جامعة الملك سعود
King Saud University
كلية علوم الحاسب والمعلومات
قسم تقنية المعلومات

IT 362 Course Project
Semester-2, 1446H

Analyzing the popularity of gaming videos

Phase 1: Data Collection Research and Assessment

Group members:

Group#:	**7**	
Section#:	**66849**	
Group Members	Name	ID
	Tarfah Bin Moammar	**444200611**
	Rawan Al-Batati	**443204566**
	Haya Al-Fayez	**444200858**
	Dalal Al-Yousef	**444203019**
	Haya Al-Ibrahim	**444200657**

Supervised by: Dr. Reem Al-qifari

1.Introduction:

The rapid expansion of online video platforms, particularly YouTube, has profoundly reshaped the entertainment landscape, making video game content one of the site's most popular and fastest-growing categories. Ranging from gameplay highlights and live streams to esports tournaments and commentary, these videos draw millions of viewers to YouTube every day.

The surging popularity of gaming videos on YouTube raises important questions about what drives their widespread appeal and how they shape gaming culture, social interaction, and the broader entertainment industry. To explore these issues, this study centers on the question: **What are the key elements that determine the success of video game content on YouTube?**

By examining this question, the research aims to identify and evaluate the main factors such as video length, tags, or timing of publication that contribute to the popularity of gaming videos on YouTube, offering insights into how and why certain videos achieve remarkable reach and engagement.

2.Data sources:

This project's dataset is sourced entirely from the **YouTube Data API**, which provides publicly available metadata and statistics for YouTube videos. Because the API offers comprehensive information such as titles, tags, descriptions, and publication dates across a large number of gaming videos, **no additional external data sources were used**. The richness of these YouTube-provided features was deemed sufficient for analyzing key factors related to video popularity, including durations, tags, channel information, and engagement metrics.

Nominal (Categorical):

- **channel_title** (Name of the channel that uploaded the video.)
- **Description** (Description text provided by the video uploader.)
- **title** (Title of the video.)
- **tags** (Keywords or tags associated with the video.)

Ratio (Numeric):

- **comment_count** (Number of comments on the video.)
- **duration** (Length of the video in ISO 8601 duration format (e.g., PT10M5S for 10 minutes, 5 seconds).
- **likes** (Number of likes.)
- **views** (Number of views.)

All data was collected exclusively through the YouTube Data API's search.list and videos.list endpoints, focusing on gaming related videos (Category ID 20) published from 2021 to 2024. Approximately **1800 observations (videos)** were compiled, each corresponding to a single YouTube gaming video.

Potential biases:

1- Representation Bias

- We exclusively used the "Gaming" category (Category ID 20) with certain date ranges (2021–2024). Thus, the dataset may underrepresent smaller or niche channels whose videos might not be labeled correctly or might fall outside these date ranges.

- The YouTube Data API's search results often prioritize more recent or more popular videos, which could skew the dataset toward higher-profile channels or trending content.

2. Measurement Bias

- Not all videos publicly display tags or engagement metrics (likes/comments can be disabled). As a result, some videos may be excluded or appear incomplete, potentially biasing the dataset toward creators who make these features publicly accessible.

3. Historical Bias

- The platform frequently updates its recommendations and monetization policies. Videos from different years (2021 vs. 2024) might have faced different algorithmic conditions, thus influencing engagement metrics.

- Larger or more established channels may have consistently higher viewership due to existing subscriber bases. This can perpetuate an inequality where smaller channels struggle to gain visibility, creating a skew toward already-popular creators.

3.Objectives:

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

4.Method:

1.Upload Time vs. Popularity

- Convert the published_at timestamp to a more analyzable format (e.g., day of the week, hour of upload)
- Compare engagement metrics (views, likes, comment count) across different upload times.
- use visualization (e.g., bar plots, heatmaps) to detect patterns.

2. Video Duration vs. Views

- Convert the duration (ISO 8601) to a numeric value in seconds or minutes.
- Plot and compute the correlation between duration and views.
- Potentially group durations into categories (short, medium, long) to see if there are thresholds where view counts cluster.

3. Comments vs. Views

- Evaluate the relationship between comment_count and views.
- Use correlation analyses and scatterplots to see if higher comment volume consistently accompanies higher view counts.

4. Effect of Tags on Engagement

- Identify prevalent tags across the dataset.
- For each tag, compare average views, likes, and comments for videos containing that tag vs. those that do not.
- Conduct statistical tests (e.g., chi-square, t-tests) or regression analyses to see if certain tags significantly boost engagement.

5. Popular Channels by Engagement

- Group videos by channel_title.
- Aggregate engagement metrics (sum of views, likes, comments) per channel.
- Rank channels by total or average engagement to identify top performers.

5.Challenges:

1. **Challenge:** The YouTube Data API enforces quotas, which limit how many requests can be made in a day. Large-scale data collection may be throttled.

Recommendation:
Spread out data collection over multiple days to stay within quota limits.

2. **Challenge**: The search.list endpoint often returns a limited set of results (e.g., ~500–600) for a given query and date range, which can bias the dataset toward more recent or more popular videos.

Recommendation:
Use multiple date ranges and varied parameters to broaden coverage.

3. **Challenge**: YouTube often hides or restricts tags to channel owners; many videos do not have publicly visible tags. This makes it harder to analyze the effect of tags on engagement.

Recommendation:
Focus on a subset of videos where tags are available.

4. **Challenge**: Many YouTube videos do not publicly display all fields (e.g., hidden likes, disabled comments, or private tags). This can reduce the usable sample size.

Recommendation:
Collect a larger initial sample to compensate for rows missing crucial fields.