Ameesh Daryani, Jose Baca, Ohitha Busireddy

CNIT 372 Group 10 Final Report

Application Motivation

Our motivation behind developing this application is centered on empowering users with data-driven insights, aiming to facilitate informed choices, enhance community engagement, and ultimately provide a more enriching YouTube experience. Personalized Insights are a key focus, offering users a profound understanding of their YouTube viewing habits to support well-informed decisions in content consumption. Additionally, the application aids in Content Discovery, helping users explore new content, genres, and creators aligned with their interests, thereby expanding and diversifying their YouTube experience. Time Management is another critical aspect, as users can leverage insights into their daily YouTube usage to effectively manage screen time and foster a healthy balance in online activities. Furthermore, the application promotes a Global Perspective by enabling users to explore popular videos and creators from different countries, providing a diverse and global outlook on content.

Database Description

For our analysis, we have utilized two Kaggle datasets available at the following links: YouTube Statistics and Global YouTube Statistics 2023. The datasets collectively contribute to the creation of three essential tables—VIDEOSTATS, COMMENTS, and GLOBALSTATS—each serving a specific purpose in our analysis.

In the VIDEOSTATS table, we capture unique video metrics such as title, video ID, and engagement metrics like likes, comments, and views. This table forms the core of our analysis, providing detailed insights into individual video performance. The COMMENTS table is intricately linked to VIDEOSTATS, offering additional insights into comment metrics, including comment ID, content, likes, and sentiment, enriching our understanding of user interactions. Finally, the GLOBALSTATS table provides comprehensive information about YouTuber rankings, earnings, and demographic data. This table is linked to both VIDEOSTATS and COMMENTS through the YouTuber name, creating a holistic view of content creators and their impact.

Questions/Solutions:

1)

How many different types of videos do users watch?

This type of question can help consumers figure out their own usage behaviors.

Knowing what types of videos they watch can help students make decisions about what other topics they may be interested in.

```
PROCEDURE get_num_video_types(p_num_video_types OUT NUMBER) IS
  BEGIN
    SELECT COUNT(DISTINCT video_type) INTO p_num_video_types FROM
VIDEOSTATS;
  END get_num_video_types;
```

Users can gain a deep understanding of their YouTube viewing habits by knowing the diversity of video types they've engaged with. They can use this information to make informed decisions about their content consumption.

2)

What videos do average users watch most frequently?

Knowing the frequency with which users watch videos can help them to understand what genres and types of content speak most to them, allowing them to look for more of the same content, or maybe even branch out and look for more variety.

```
PROCEDURE get_top_video_by_average_views(p_title OUT VARCHAR2,
p_average_views OUT NUMBER) IS
  BEGIN
    SELECT title, AVG(views) INTO p_title, p_average_views
    FROM VIDEOSTATS
    GROUP BY title
    ORDER BY average_views DESC
    FETCH FIRST 1 ROWS ONLY;
  END get_top_video_by_average_views;
```

Helps users to identify the most consistently popular video based on average views, aiding consumers in discovering highly engaging and well-received content.

3)

Which comments have the most likes within what time period since they were posted?

Users that interact with youtube videos through comments often attempt to gain likes from other interactors. If the question is how to get the most peer reactions, this would be the first question to ask towards that goal.

```
PROCEDURE get_most_liked_comment(p_commentid OUT NUMBER, p_videoid
```

```
OUT NUMBER, p_commentcontent OUT VARCHAR2, p_likes OUT NUMBER,
p_sentiment OUT VARCHAR2) IS
  BEGIN
    SELECT commentid, videoid, commentcontent, likes, sentiment
INTO p_commentid, p_videoid, p_commentcontent, p_likes,
p_sentiment
    FROM COMMENTS
    ORDER BY likes DESC
    FETCH FIRST 1 ROWS ONLY;
  END get_most_liked_comment;
```

Highlights the most liked comment, providing insight into user sentiment and the community's positive interactions around a video.

4)

How much youtube do users watch a day?

This question can help users track and understand their usage. If, for example, a user wanted to cut down on their youtube usage, this would be the first question they would need answered to get toward that goal.

```
  PROCEDURE get_total_watch_time(p_videoid OUT NUMBER,
p_total_watch_time OUT NUMBER) IS
  BEGIN
    SELECT videoid, SUM(views) INTO p_videoid, p_total_watch_time
    FROM VIDEOSTATS
    GROUP BY videoid;
  END get_total_watch_time;
```

Calculates the total watch time for each video, assisting consumers in understanding the overall popularity and impact of specific content.

5)

What are the most popular genres of videos?

Knowing this can help USERS find new genres of videos or find new Youtubers based on genres.

```
  PROCEDURE get_genre_count(p_keyword OUT VARCHAR2, p_genre_count OUT
NUMBER) IS
  BEGIN
```

```
    SELECT keyword, COUNT(*) INTO p_keyword, p_genre_count
    FROM VIDEOSTATS
    GROUP BY keyword
    ORDER BY genre_count DESC;
  END get_genre_count;
```

Show the count of videos for each genre, enabling consumers to explore and find content that aligns with their specific interests.

6)

Which YouTubers have the most comments and likes on their videos?

Users may want to know which Youtubers get a lot of comments and may want to watch Youtubers that have more engaging audiences as they can also participate in the community

```
PROCEDURE get_youtuber_stats(p_youtuber OUT VARCHAR2, p_comment_count OUT
NUMBER, p_like_count OUT NUMBER) IS
  BEGIN
    SELECT g.youtuber, COUNT(c.commentid), SUM(v.likes) INTO p_youtuber,
p_comment_count, p_like_count
    FROM GLOBALSTATS g
    LEFT JOIN VIDEOSTATS v ON g.title = v.title
    LEFT JOIN COMMENTS c ON v.videoid = c.videoid
    GROUP BY g.youtuber
    ORDER BY comment_count DESC, like_count DESC;
  END get_youtuber_stats;
```

Provide statistics for each YouTuber, including comment and like counts, allowing consumers to gauge the popularity and engagement level of different content creators.

7)

What are the most popular videos by country?

Users trying to diversify the videos that they watch would ask this question as they would want to see a variety of popular videos from different countries.

```
 PROCEDURE get_country_view_count(p_videoid OUT NUMBER, p_country
OUT VARCHAR2, p_view_count OUT NUMBER) IS
  BEGIN
    SELECT v.videoid, COALESCE(g.country, 'United States'),
COUNT(*) INTO p_videoid, p_country, p_view_count
```

```
    FROM VIDEOSTATS v
    LEFT JOIN GLOBALSTATS g ON v.title = g.title
    GROUP BY v.videoid, g.country
    ORDER BY view_count DESC;
  END get_country_view_count;
```

Display the view count for each video in a specific country, offering consumers insights into the global popularity of videos.

8)

What is the average duration of the most popular videos?

This question involves understanding video duration trends, which can be essential for content creators and consumers alike. It helps creators optimize their video length for engagement and viewers to manage their time effectively.

```
PROCEDURE get_average_popularity(p_average_popularity OUT NUMBER) IS
  BEGIN
    SELECT AVG(vs.views / vs.comments) INTO p_average_popularity
    FROM VIDEOSTATS vs
    JOIN GLOBALSTATS gs ON vs.videoid = gs.videoid
    ORDER BY (vs.views / vs.comments) DESC
    FETCH FIRST 10 ROWS ONLY;
  END get_average_popularity;
```

Identify the video with the maximum likes, helping consumers discover exceptionally well-received content on YouTube.

9)

How many likes does the most popular video currently have?

Involves basic aggregation to find the maximum number of likes, which is valuable for users curious about trending content.

```
PROCEDURE get_max_likes(p_max_likes OUT NUMBER) IS
  BEGIN
    SELECT MAX(likes) INTO p_max_likes FROM VIDEOSTATS;
  END get_max_likes;
```

Calculate the average popularity of videos based on the ratio of views to comments, aiding consumers in identifying content with a balanced and engaging audience interaction.

10)

What day of the week sees the highest number of new video uploads?

Involves analyzing temporal patterns, helping both content creators and users know the most active days for new content.

```
PROCEDURE get_day_of_highest_uploads(p_day_of_highest_uploads OUT VARCHAR2,
p_upload_count OUT NUMBER) IS
  BEGIN
    SELECT TO_CHAR(publishedat, 'Day'), COUNT(*) INTO
p_day_of_highest_uploads, p_upload_count
    FROM VIDEOSTATS
    GROUP BY TO_CHAR(publishedat, 'Day')
    ORDER BY upload_count DESC
    FETCH FIRST 1 ROWS ONLY;
  END get_day_of_highest_uploads;
```

Determine the day with the highest number of video uploads, allowing consumers to plan their viewing based on peak content release times.

Team

Ameesh Daryani:

- Created and set up Github repo, created and filled out readme, and downloaded and remediated data from online sources before creating a database loading file uploaded to the repository. Helped to create 4 questions and reasonings. Refactored all solutions to be in procedures, and all procedures to be included in a package. Wrote Application Motivation and Database Description sections in the report, as well as formatted questions and code snippets 1 through 10

Ohitha Busireddy

- I created three crucial questions and provided detailed reasoning behind them. These questions laid the groundwork for our database design and subsequent analysis. Additionally, I successfully solved five (Q1, Q2, Q3, Q8, Q10) out of the ten questions, showcasing a deep understanding of our dataset and the ability to derive meaningful insights. Furthermore, I took the

initiative to craft the application's motivation, highlighting its significance in empowering content creators, analysts, and YouTube enthusiasts with valuable insights from video data.

Jose Baca

- I helped create 3 crucial questions and the detailed reasonings behind them. In addition to helping shape the questions, I was actively involved in the selection process of crucial datasets, Additionally, I helped provide solutions for a few of the ten questions.