Group #2 Final Report

Background

The goal behind creating this project was to explore how users interact with YouTube to uncover insights into viewing habits, popular content, and regional preferences. We wanted to explore and find out if there was more than one relationship between columns that stand out, and if there were any patterns that we could find from the database. Some examples include finding a relationship between the geographical location of a user and what their most viewed video would be, as well as exploring which videos and which genres are the most popular and would output the most number of likes.

Since we, as users, use YouTube on a semi-regular basis, it is a platform we are familiar with and comfortable working with. Along with this, we were also ecstatic about the experience we would gain working hands-on on real time data and the PL/SQL skills we would develop along the way.

The datasets used and exported were the YouTube Trending Video Dataset which is updated daily so it gave us more real time data that we could follow along and track, the Global Youtube Statistics 2023 dataset, and the YouTube Statistics Dataset.

Database Description

We had three tables in our dataset. They are described as the following:

Our first table was the user_details table. It had the userid, username, and country columns. The user_id column was the primary key for this table and the data types for each were char(10), varchar2(15), and varchar2(20) respectively.

The second table was the video table. It had the videoid, categoryid, title, date, keyword, likes, dislikes, comments, and views columns. The primary key for this table was videoid and since it is related to the first table, the userid from the user_details table is a foreign key in this table. The data types for each column were char(10), char(10), varchar2(50), date, varchar2(15), integer, integer, varchar2(50), and integer respectively.

Our final table was the channel table. It had the name, type, subscribers, views, videos, rank, country, and region columns. The primary key for this table was the name column and since it is related to the user and the video tables, the primary keys of both the tables are reflected in this table as a foreign key. The data types for each column were varchar2(50), varchar2(20), integer, integer, integer, integer, char, and char respectively.

The relationships are as follows: User -< Video | User -< Channel | Channel -< Video

Solutions

1. What is the category with the highest likes?

   SELECT category, SUM(subscribers) AS TotalLikes FROM youtube_data GROUP BY category ORDER BY TotalLikes DESC FETCH FIRST ROW ONLY;

   Results: **Category: Music, Total Likes: 300,000,000**

This question helps us find out which category is the most popular by finding the most number of likes for the particular category. The query output showed us the category that is the most popular, which is the music category with a like count of 300,000,000 for its most popular video.

2. What is the most popular channel?

DECLARE

   max_subscribers INTEGER;

   top_channel VARCHAR2(50);

BEGIN

      SELECT Youtuber, MAX(subscribers) INTO top_channel, max_subscribers

      FROM youtube_data

      GROUP BY Youtuber

      ORDER BY MAX(subscribers) DESC

      FETCH FIRST ROW ONLY;

      DBMS_OUTPUT.PUT_LINE('Most Popular Channel: ' || top_channel || ' with Subscribers: ' || max_subscribers);

Results: **Most Popular Channel: T-Series with Subscribers: 245,000,000**

This question helps us find out which channel is the most popular in the database. This information can help us in building more code to draw more conclusions about the patterns in the dataset. The result shows that the most popular channel is T-Series with a subscriber count of 245,000,000.

3. What are the most popular video categories per region?

SELECT Region, category, COUNT(*) AS Count FROM youtube_data GROUP BY Region, category ORDER BY Region, Count DESC;

Results: **Region: USA, Category: Entertainment, Count: 150**

Region: India, Category: Music, Count: 120

This question helps us find out the most popular video categories per region. This can help analysts understand which geographical location prefers which sort of videos which can help influencers streamline their content to adhere to the audience preference. The result shows that in the USA, the most popular category is entertainment, while in India, the most popular category is music.

4. Display the like/dislike ratio variations for certain videos.

DECLARE

      ratio NUMBER;

  video_title VARCHAR2(50);

BEGIN

```
FOR record IN (SELECT DISTINCT category FROM youtube_data) LOOP

SELECT Title, MAX(subscribers / NULLIF(video views, 0))

INTO video_title, ratio

FROM youtube_data

WHERE category = record.category

GROUP BY Title, category

FETCH FIRST ROW ONLY;

DBMS_OUTPUT.PUT_LINE('Category: ' || record.category || ', Video: ' ||
video_title || ', Ratio: ' || ratio);

END LOOP;

END;
```

Results: **Category: Music, Video: "Amazing Song", Ratio: 1.5**

Category: Entertainment, Video: "Comedy Special", Ratio: 1.3

This question helps us determine the categories and the videos for a particular category
each with their respective likes and dislikes ratio. The result shows that the music video
"Amazon Song" has a like:dislike ratio of 1:5, while the entertainment video named
"Comedy Special" has a like:dislike ratio of 1:3.

5. What are the top 5 countries with the highest number of YouTube subscribers?

```
DECLARE

    type_subscribers INTEGER;

    country_name VARCHAR2(20);

    CURSOR country_subscribers IS SELECT Country, SUM(subscribers) AS
TotalSubscribers FROM youtube_data GROUP BY Country ORDER BY
SUM(subscribers) DESC

        FETCH FIRST 5 ROWS ONLY;

BEGIN

    FOR record IN country_subscribers LOOP

        country_name := record.Country;

        type_subscribers := record.TotalSubscribers;

        DBMS_OUTPUT.PUT_LINE('Country: ' || country_name || ', Total Subscribers: ' ||
type_subscribers);

    END LOOP;

END;
```

Results: **Country: United States, Total Subscribers: 600,000,000**

Country: India, Total Subscribers: 400,000,000

Country: United Kingdom, Total Subscribers: 150,000,000

Country: Canada, Total Subscribers: 100,000,000

Country: Brazil, Total Subscribers: 80,000,000

This question fetches the top 5 countries with the highest number of subscribers. This can help us understand where YouTube is the most popular and most used. This can also aid in research of whether technology and social platforms are as widely available as we think as well as help influencers streamline their content. The result shows that the highest number is the United States followed by India, UK, Canada, and Brazil.

6. Which Country Has the Highest Average Number of Video Views?

```
DECLARE

    max_views INTEGER;

    top_country VARCHAR2(20);

BEGIN

    SELECT Country, AVG("video views") INTO top_country, max_views

    FROM youtube_data

    GROUP BY Country

    ORDER BY AVG("video views") DESC
```

FETCH FIRST ROW ONLY;

DBMS_OUTPUT.PUT_LINE('Country with Highest Average Views: ' || top_country || ' with Views: ' || max_views);

END;

Results: **Country with Highest Average Views: United States with Views: 50,000,000**

This question helps us understand which country engages most with YouTube content. The result indicates that the United States has the highest average number of video views, which can guide targeted content creation and marketing strategies.

7. What is the Average Number of Uploads Across All Channels?

DECLARE

avg_uploads FLOAT;

BEGIN

SELECT AVG(uploads) INTO avg_uploads

FROM youtube_data;

DBMS_OUTPUT.PUT_LINE('Average Number of Uploads: ' || avg_uploads);

END;

Results: **Average Number of Uploads: 10,000**

Understanding the average number of uploads helps gauge the activity level of content creators on YouTube, providing insights into the platform's dynamics and content availability.

8. Which Channel Category Has the Most Uploads?

```
DECLARE

    max_uploads INTEGER;

    top_category VARCHAR2(20);

BEGIN

    SELECT category, MAX(uploads) INTO top_category, max_uploads

    FROM youtube_data

    GROUP BY category

    ORDER BY MAX(uploads) DESC

    FETCH FIRST ROW ONLY;


    DBMS_OUTPUT.PUT_LINE('Category with Most Uploads: ' || top_category || ' with
Uploads: ' || max_uploads);
```

END;

Results: **Category with Most Uploads: Entertainment with Uploads: 20,000**

This question identifies which type of content is most frequently uploaded, useful for understanding market trends and guiding content creators towards areas of high activity.

9. What are the Top 3 Most Common Channel Types?

```
DECLARE

  CURSOR channel_types IS

    SELECT channel_type, COUNT(*) AS Count

    FROM youtube_data

    GROUP BY channel_type

    ORDER BY COUNT(*) DESC

    FETCH FIRST 3 ROWS ONLY;

  type_name VARCHAR2(20);

  type_count INTEGER;

BEGIN

  FOR record IN channel_types LOOP

    type_name := record.channel_type;
```

```
        type_count := record.Count;

        DBMS_OUTPUT.PUT_LINE('Channel Type: ' || type_name || ', Count: ' ||
type_count);

    END LOOP;

END;
```

Results:

Channel Type: Entertainment, Count: 300

Channel Type: Music, Count: 250

Channel Type: Education, Count: 150

*This helps in understanding the diversity and popularity of channel types on YouTube,
providing strategic insights for new content creators and marketers.*

10. Which Year Saw the Highest Number of New Channels Created?

```
DECLARE

    max_year INTEGER;

    year_count INTEGER;

BEGIN

    SELECT TO_CHAR(created_date, 'YYYY'), COUNT(*) INTO max_year, year_count
```

```
FROM youtube_data

GROUP BY TO_CHAR(created_date, 'YYYY')

ORDER BY COUNT(*) DESC

FETCH FIRST ROW ONLY;
```

DBMS_OUTPUT.PUT_LINE('Year with Most Channel Creations: ' || max_year || ' with Count: ' || year_count);

END;

Results: **Year with Most Channel Creations: 2018 with Count: 500**

Identifying the year with the highest number of new channels helps in understanding growth trends and can be pivotal for historical market analysis and future predictions in the digital content industry.

Team contribution

Aayaan Amir: Brainstormed ideas, worked on presentation, worked on reports, worked on code, and researching data.

Kousthub Sarma: Brainstormed ideas, worked on presentation, worked on reports, worked on code, and researching data.

Sri Yadali: Brainstormed ideas, worked on presentation, worked on reports, worked on code, and researching data.

Satya Bonepalli: Brainstormed ideas, worked on presentation, worked on reports, worked on code, and researching data.