Group Project Final Report

CNIT 372

Team 9

Alexis Chaffin,

Sean Maloney,

Natsu Yamamoto,

Collin Johnson

Submitted To: Prof. Tianyi Li

Submitted On: 15 December 2023

github Link: https://github.com/mocha-jfif/CNIT372Project

Background

Our goal is to determine how successful a content creator on YouTube is based on metrics such as their subscriber count and view count concerning the average video length they produce. This data can be analyzed to see how much any one factor affects the success of a creator, and if there are any combinations of factors that correlate to the success of a channel. This provides an insight into what content is more popular—short-form, such as TikTok, YouTube shorts, and so on, or long-form, like traditional YouTube videos.

Database Description

This database contains three tables. The three tables are the Influencer Table (named GP_Influencers), the Video Table (GP_VIDEOS), and the Interaction Table (GP_INTERACTIONS). Each of these tables are linked with the primary key “Video_Title”, which is the title of a video with a type of varchar2.

The Influencer Table stores data about the influencer and their channel, such as a video title, the name of the creator, the gender of the creator, the total number of subscribers, the total number of views, the total number of videos created, the total number of playlists, and the date the channel was created. Figure 1 shows the datatypes for each of these entities.

COLUMN_NAME	DATA_TYPE	NULLABLE	DATA_DEFAULT	COLUMN_ID	COMMENTS
VIDEO_TITLE	VARCHAR2(256 BYTE)	No	(null)	1	(null)
CREATOR_NAME	VARCHAR2(64 BYTE)	Yes	(null)	2	(null)
CREATOR_GENDER	VARCHAR2(16 BYTE)	Yes	(null)	3	(null)
TOTAL_SUBSCRIBERS	NUMBER	Yes	(null)	4	(null)
TOTAL_VIEWS	NUMBER	Yes	(null)	5	(null)
TOTAL_VIDEOS	NUMBER	Yes	(null)	6	(null)
TOTAL_PLAYLISTS	NUMBER	Yes	(null)	7	(null)

Figure 1: The GP_INFLUENCERS Table

The Video Table stores data about specific videos, such as the total number of views a video has received, the duration time of each video, the duration of the video in seconds, whether it has subtitles, whether there is a description, the total number of hashtags, the maximum quality

of the video, and of course, the video title. Figure 2 shows the datatypes for each of these entities.

COLUMN_NAME	DATA_TYPE	NULLABLE	DATA_DEFAULT	COLUMN_ID	COMMENTS
VIDEO_TITLE	VARCHAR2(256 BYTE)	No	(null)	1	(null)
TOTAL_VIEWS	NUMBER	Yes	(null)	2	(null)
DURATION_TIME	TIMESTAMP(6)	Yes	(null)	3	(null)
DURATION_SECONDS	NUMBER	Yes	(null)	4	(null)
DATE_UPLOADED	DATE	Yes	(null)	5	(null)
HAS_SUBTITLES	VARCHAR2(3 BYTE)	Yes	(null)	6	(null)
HAS_DESCRIPTION	VARCHAR2(3 BYTE)	Yes	(null)	7	(null)
TOTAL_HASHTAGS	NUMBER	Yes	(null)	8	(null)
MAX_QUALITY	NUMBER	Yes	(null)	9	(null)

Figure 2: The GP_VIDEOS Table

The Interaction Table stores data regarding the interactions per video, such as the total number of likes on the video, the number of comments, the date of the last comment, whether the video has premiered, how many weekly posts, and the video title. Figure 3 shows the datatypes for each of these entities.

COLUMN_NAME	DATA_TYPE	NULLABLE	DATA_DEFAULT	COLUMN_ID	COMMENTS
VIDEO_TITLE	VARCHAR2(256 BYTE)	No	(null)	1	(null)
TOTAL_LIKES	NUMBER	Yes	(null)	2	(null)
TOTAL_COMMENTS	NUMBER	Yes	(null)	3	(null)
LAST_COMMENT_DATE	DATE	Yes	(null)	4	(null)
HAS_PREMIERED	VARCHAR2(3 BYTE)	Yes	(null)	5	(null)
WEEKLY_POSTS	NUMBER	Yes	(null)	6	(null)

Figure 3: The GP_INTERACTIONS Table

Team Contribution

Our team had 4 members: Collin Johnson, who is a senior in CIT; Sean Maloney, who is a junior in CIT; Natsu Yamamoto, who is a junior in CIT; and Alexis Chaffin, who is a senior in CIT. Each member contributed to the project. Alexis completed questions 1 and 2. Sean completed questions 3 and 4, Collin completed questions 7, 8, and 9, and Natsu completed questions 5, 6, and 10. Each team member wrote the sections pertaining to their questions in the report.

Questions and Solutions

1. Which influencer (or creator) has the shortest and longest average video length, and how many subscribers do they have?
    a. **Justification:** This question is an important basis for any assumptions about the popularity of a channel based on video length. This will look at the average video length related to how many subscribers they have, which is then used for analysis in Question 2. This is useful for both data analysts and content creators trying to grow their channels.
    b. **Solution**: Create a function called “mindur” to return the minimum average duration of a video. Next, create a function called “maxdur” to return the maximum average duration of a video. Using PL/SQL, deploy these two functions to return channel statistics of channels that have the maximum average duration and the minimum average duration. The code is shown in Figure Q1.
    c. **Results:**

    Influencer with maximum average duration (in seconds): your story tv

    Average length: 34980

    Number of subscribers: 163000

    - - - - - - - - - - - -

    Influencer with minimum average duration (in seconds): Acvideos

    Average length: 42

    Number of subscribers: 229000

    d. **Insights**: Not applicable; analyzed in Question 2.

```
create or replace function maxdur return number is
    maxavgduration number;
begin
    select
        max(avg(v.duration_seconds))
    into
        maxavgduration
    from
        gp_influencers i
    join
        gp_videos v on i.video_title = v.video_title
    group by
        i.creator_name;

    return maxavgduration;
end;

declare
    mincreate varchar2(64);
    maxcreate varchar2(64);
    mindur number;
    maxdur number;
    subcountmax number;
    subcountmin number;
begin
    select max(avgduration), max(creator_name)
    into maxdur, maxcreate
    from (
        select avg(v.duration_seconds) as avgduration, c.creator_name
        from gp_influencers c
        join gp_videos v on c.video_title = v.video_title
        group by c.creator_name
    );

    select min(avgduration), min(creator_name)
    into mindur, mincreate
    from (
        select avg(v.duration_seconds) as avgduration, c.creator_name
        from gp_influencers c
        join gp_videos v on c.video_title = v.video_title
        group by c.creator_name
    );

    select total_subscribers
    into subcountmax
    from gp_influencers
    where creator_name = maxcreate;

    select total_subscribers
    into subcountmin
    from gp_influencers
    where creator_name = mincreate;

    dbms_output.put_line('Influencer with maximum average duration (in seconds): ' || maxcreate);
    dbms_output.put_line('Average length: ' || maxdur);
    dbms_output.put_line('Number of subscribers: ' || subcountmax);
    dbms_output.put_line('- - - - - - - - - - - -');
    dbms_output.put_line('Influencer with minimum average duration (in seconds): ' || mincreate);
    dbms_output.put_line('Average length: ' || mindur);
    dbms_output.put_line('Number of subscribers: ' || subcountmin);
end;
```

Figure 4: Question 1 Code Block

2. By finding the influencer with the highest subscriber count and the lowest subscriber count and using the previous influencers found in Question 1, how do these influencers compare to those who have the highest and lowest metrics? (Subscribers, views, etc.)
    a. **Justification:** This is an important comparison tool to determine if video length does contribute to success. Comparing the influencers based on the number of subscribers or views they have provided an indicator of whether video length contributes directly to success or not.
    b. **Solution:** Create a function called “high_sub” to return the creator’s name with the highest subscriber count by selecting the maximum total subscribers from the GP_Influencers table. Next, create a function called “low_sub” to return the creator’s name with the lowest subscriber count by selecting the minimum total subscribers from the GP_Influencers table. Using PL/SQL, deploy these two functions to return channel names of channels that have the maximum subscriber count and the minimum subscriber count. The code is shown in Figure Q2.
    c. **Results:**

        Prime has the highest subscriber count of: 89800000.

        Tesery has the lowest subscriber count of: 242.
    d. **Insights:** Neither of the creators that were found in Question 1 are present in Question 2, and the number of subscribers is nowhere near each other. This begins to suggest that video length is not a primary factor in a channel’s success, but it is yet to be proven false.

```
create or replace function high_sub return varchar2 is
    v_creator_name varchar2(255);
    highest_subcount number;

begin
    select
        max(total_subscribers)
    into
        highest_subcount
    from
        gp_influencers;

    select creator_name
    into v_creator_name
    from gp_influencers
    where total_subscribers = highest_subcount and rownum = 1;

    return v_creator_name || ' has the highest subscriber count of: ' ||
highest_subcount;
end;

create or replace function low_sub return varchar2 is
    v_creator_name varchar2(255);
    lowest_subcount number;

begin
    select
        min(total_subscribers)
    into
        lowest_subcount
    from
        gp_influencers;

    select creator_name
    into v_creator_name
    from gp_influencers
    where total_subscribers = lowest_subcount and rownum = 1;

    return v_creator_name || ' has the lowest subscriber count of: ' ||
lowest_subcount;
end;

declare
    high_res varchar2(2500);
    low_res varchar2(2500);
```

Figure 5: Question 2 Code Block

3. Find the average viewer count, number of likes, and number of comments on short videos (less than 2 minutes) and long videos (more than 10 minutes). Which type of video do people interact with more?
    a. **Justification:** This helps determine how well short-form content does on YouTube. This can help content creators find out what kind of content will garner more views and interactions, and therefore help their channel grow more. Short-form content has become exceedingly popular thanks to new social media trends, such as TikTok and Instagram. Long-form content has historically been more popular than short-form, and longer videos generally provide more information on any given topic. We plan to use subqueries to get a list of all short and long videos, and then get average values from those lists.
    b. **Solution:** Create a procedure named "Get_Average_Popularity_Measurements" that calculates the average statistics of videos of either greater than or less than a developer-specified length. Cursors and FOR loops were utilized to find sums of views, likes, and comments on relative videos to use when calculating the average values.
    c. **Results:**

        Influencers with average video durations of 600 seconds or greater have 848569155 views, 253481 likes, and 12766 comments on average.

        - - - - - - - - - - - -

        Influencers with average video durations of 120 seconds or less have 2733504562 views, 364421 likes, and 209403 comments on average.

d. **Insights:** We found that videos that are shorter than 2 minutes garner substantially more views, likes, and comments on YouTube. This suggests that short-form content has become more popular with the consumer market than long-form content. However, it is important to note that this could be an indirect result of a viewer's ability to watch many short videos in the span of one long video, which can increase the amount of likes given to short videos in comparison with longer ones. Regardless, this data does demonstrate that shorter videos are interacted with more than longer videos.

```
CREATE OR REPLACE PROCEDURE Get_Average_Popularity_Measurements (p_comparison IN VARCHAR2, p_duration IN NUMBER)
AS
    -- Define variables.
    v_total_views NUMBER := 0;
    v_total_likes NUMBER := 0;
    v_total_comments NUMBER := 0;
    v_total_videos NUMBER := 0;

    v_average_views NUMBER;
    v_average_likes NUMBER;
    v_average_comments NUMBER;

    -- Define cursors.
    CURSOR c_videos_greater IS
        SELECT GP_Videos.Video_Title, GP_Videos.Total_Views, GP_Interactions.Total_Likes, GP_Interactions.Total_Comments
        FROM GP_Videos INNER JOIN GP_Interactions
            ON GP_Videos.Video_Title = GP_Interactions.Video_Title
        WHERE Duration_Seconds >= p_duration;

    CURSOR c_videos_less IS
        SELECT GP_Videos.Video_Title, GP_Videos.Total_Views, GP_Interactions.Total_Likes, GP_Interactions.Total_Comments
        FROM GP_Videos INNER JOIN GP_Interactions
            ON GP_Videos.Video_Title = GP_Interactions.Video_Title
        WHERE Duration_Seconds <= p_duration;
BEGIN
    -- Loop through c_videos_greater if the user specifies the "greater" comparison.
    IF LOWER(p_comparison) = 'greater' THEN
        FOR video IN c_videos_greater LOOP
            -- Add the current video's views, likes, and comments to the totals and increment the amount of videos taken into consideration thus far.
            v_total_views := v_total_views + video.Total_Views;
            v_total_likes := v_total_likes + video.Total_Likes;
            v_total_comments := v_total_comments + video.Total_Comments;

            v_total_videos := v_total_videos + 1;
        END LOOP;

        -- Check if the cursor was empty.
        IF v_total_videos != 0 THEN
            -- Calculate and print the average numbers of views, likes, and comments.
            v_average_views := ROUND((v_total_views / v_total_videos), 0);
            v_average_likes := ROUND((v_total_likes / v_total_videos), 0);
            v_average_comments := ROUND((v_total_comments / v_total_videos), 0);

            DBMS_OUTPUT.PUT_LINE('Influencers with average video durations of ' || p_duration || ' seconds or greater have ' || v_average_views || ' views, ' || v_average_likes || ' likes, and ' || v_average_comments || ' comments on average.');
        ELSE
            -- Print an error if the cursor was empty.
            DBMS_OUTPUT.PUT_LINE('ERROR: There are no videos that match the given criteria.');
        END IF;
    -- Loop through c_videos_less if the user specifies the "less" comparison.
    ELSIF LOWER(p_comparison) = 'less' THEN
        FOR video IN c_videos_less LOOP
            -- Add the current video's views, likes, and comments to the totals and increment the amount of videos taken into consideration thus far.
            v_total_views := v_total_views + video.Total_Views;
            v_total_likes := v_total_likes + video.Total_Likes;
            v_total_comments := v_total_comments + video.Total_Comments;

            v_total_videos := v_total_videos + 1;
        END LOOP;

        -- Check if the cursor was empty.
        IF v_total_videos != 0 THEN
            -- Calculate and print the average numbers of views, likes, and comments.
            v_average_views := ROUND((v_total_views / v_total_videos), 0);
            v_average_likes := ROUND((v_total_likes / v_total_videos), 0);
            v_average_comments := ROUND((v_total_comments / v_total_videos), 0);

            DBMS_OUTPUT.PUT_LINE('Influencers with average video durations of ' || p_duration || ' seconds or less have ' || v_average_views || ' views, ' || v_average_likes || ' likes, and ' || v_average_comments || ' comments on average.');
        ELSE
            -- Print an error if the cursor was empty.
            DBMS_OUTPUT.PUT_LINE('ERROR: There are no videos that match the given criteria.');
        END IF;
    -- Print an error message if the comparison argument is not recognized.
    ELSE
        DBMS_OUTPUT.PUT_LINE('ERROR: Comparison argument not recognized. Please enter "greater" or "less" as the comparison argument.');
    END IF;
END Get_Average_Popularity_Measurements;
/
```

Figure 6: Question 3 Code Block

4. Find the average subscriber count for content creators who primarily make short videos (less than 2 minutes) and long videos (more than 10 minutes). Which group of content creators have more subscribers on average?
    a. **Justification:** This helps us learn how much the average viewer in interested in a specific creator's short-form content. Modern social media usually offers content from an extremely wide range of creators in a feed that plays short videos one after another, so viewers are constantly exposed to different creators and subjects. Content creators who make longer videos typically focus on a few specific topics of interest, and their content is based on those topics. We can use PL/SQL to loop through all videos made by specific content creators and calculate their average subscriber counts.
    b. **Solution:** Create a procedure named "Get_Average_Subscribers" that calculates the average subscribers of creators who specialize in creating content either greater than or less than a developer-specified length. Cursors and FOR loops were utilized to find sum of subscribers on relative videos to use when calculating the average value.
    c. **Results:**

Influencers with average video durations of 600 seconds or greater have 7626484 subscribers on average.

- - - - - - - - - - - -

Influencers with average video durations of 120 seconds or less have 8855957 subscribers on average.

d. **Insights:** The results of this question demonstrate that while creators of short videos do in fact have more subscribers than creators of long videos, the difference is fairly slim and could very well be considered negligible to many creators. This data suggests that the average length of videos produced by a creator and the number of consumers subscribed to their channel are not directly related.

```
CREATE OR REPLACE PROCEDURE Get_Average_Subscribers (p_comparison IN VARCHAR2, p_duration IN NUMBER)
AS
    -- Define variables.
    v_total_subscribers NUMBER := 0;
    v_total_creators NUMBER := 0;

    v_average_subscribers NUMBER;

    -- Define cursors.
    CURSOR c_creators_greater IS
        SELECT GP_Influencers.Creator_Name, GP_Influencers.Total_Subscribers
        FROM GP_Influencers INNER JOIN GP_Videos
            ON GP_Influencers.Video_Title = GP_Videos.Video_Title
        WHERE Duration_Seconds >= p_duration;

    CURSOR c_creators_less IS
        SELECT GP_Influencers.Creator_Name, GP_Influencers.Total_Subscribers
        FROM GP_Influencers INNER JOIN GP_Videos
            ON GP_Influencers.Video_Title = GP_Videos.Video_Title
        WHERE Duration_Seconds <= p_duration;
BEGIN
    -- Loop through c_creators_greater if the user specifies the "greater" comparison.
    IF LOWER(p_comparison) = 'greater' THEN
        FOR creator IN c_creators_greater LOOP
            -- Add the current creator's subscribers to the total and increment the amount of subscribers taken into consideration thus far.
            v_total_subscribers := v_total_subscribers + creator.Total_Subscribers;

            v_total_creators := v_total_creators + 1;
        END LOOP;

        -- Check if the cursor was empty.
        IF v_total_creators != 0 THEN
            -- Calculate and print the average number of subscribers.
            v_average_subscribers := ROUND((v_total_subscribers / v_total_creators), 0);

            DBMS_OUTPUT.PUT_LINK('Influencers with average video durations of ' || p_duration || ' seconds or greater have ' || v_average_subscribers || ' subscribers on average.');
        ELSE
            -- Print an error if the cursor was empty.
            DBMS_OUTPUT.PUT_LINK('ERROR: There are no influencers with videos that match the given criteria.');
        END IF;
    -- Loop through c_creators_less if the user specifies the "less" comparison.
    ELSIF LOWER(p_comparison) = 'less' THEN
        FOR creator IN c_creators_less LOOP
            -- Add the current creator's subscribers to the total and increment the amount of subscribers taken into consideration thus far.
            v_total_subscribers := v_total_subscribers + creator.Total_Subscribers;

            v_total_creators := v_total_creators + 1;
        END LOOP;

        -- Check if the cursor was empty.
        IF v_total_creators != 0 THEN
            -- Calculate and print the average number of subscribers.
            v_average_subscribers := ROUND((v_total_subscribers / v_total_creators), 0);

            DBMS_OUTPUT.PUT_LINK('Influencers with average video durations of ' || p_duration || ' seconds or less have ' || v_average_subscribers || ' subscribers on average.');
        ELSE
            -- Print an error if the cursor was empty.
            DBMS_OUTPUT.PUT_LINK('ERROR: There are no influencers with videos that match the given criteria.');
        END IF;
    -- Print an error message if the comparison argument is not recognized.
    ELSE
        DBMS_OUTPUT.PUT_LINK('ERROR: Comparison argument not recognized. Please enter "greater" or "less" as the comparison argument.');
    END IF;
END Get_Average_Subscribers;
/
```

Figure 7: Question 4 Code Block

5. Are more short videos (2 minutes or less) being uploaded in recent years compared to long videos (10 minutes or more)?
    a. **Justification:** This helps us learn how the period of when a video was uploaded can influence the length of the content. Short videos and hashtags are relatively newer concepts on YouTube compared to long videos, comments, and likes, so it is possible that shorter videos were uploaded more in recent years and longer videos occurred more in earlier years. As the popularity of short videos has increased tremendously, it is also possible that there are more short videos than long videos in the 2020's.
    b. **Solution:** Create a procedure named video_date that uses a for loop to loop through each video that is less than 1 minute long and finds the years where short videos were uploaded and number of videos uploaded that year. The procedure uses another for loop to loop through videos that are over 10 minutes long, which are considered long videos. The procedure prints the data and separates them into the two different categories (short videos and long videos).
    c. **Results:**

        Video upload years for short videos

        Year uploaded: 2022 Number of videos: 36

        Year uploaded: 2019 Number of videos: 5

        Year uploaded: 2008 Number of videos: 1

        Year uploaded: 2013 Number of videos: 1

        Year uploaded: 2020 Number of videos: 2

Year uploaded: 2005 Number of videos: 1

Year uploaded: 2018 Number of videos: 2

Year uploaded: 2021 Number of videos: 6

Year uploaded: 2017 Number of videos: 1

Year uploaded: 2012 Number of videos: 1

Year uploaded: 2015 Number of videos: 3

Year uploaded: 2006 Number of videos: 1

- - - - - - - - - - - -

Video upload years for long videos

Year uploaded: 2022 Number of videos: 274

Year uploaded: 2019 Number of videos: 31

Year uploaded: 2013 Number of videos: 1

Year uploaded: 2020 Number of videos: 52

Year uploaded: 2016 Number of videos: 6

Year uploaded: 2018 Number of videos: 20

Year uploaded: 2014 Number of videos: 2

Year uploaded: 2021 Number of videos: 78

Year uploaded: 2017 Number of videos: 13

Year uploaded: 2012 Number of videos: 2

Year uploaded: 2015 Number of videos: 6

Year uploaded: 2011 Number of videos: 1

d. **Insights:** The results show that there are far more long videos than short videos in general. Because of this, there are much more long videos than short videos even in recent years. There is a decent number of longer videos in previous years while there are less than 10 short videos every year other than 2022. There was a major increase in short videos in 2022. This shows that short videos have slowly started to become more popular and they are more common in the past year. However, as there are many more long videos than short videos, we can see that long videos are more popular despite the trends of short videos.

```
create or replace procedure video_date
is
begin
      DBMS_OUTPUT.PUT_LINE('Video upload years for short videos');
      DBMS_OUTPUT.PUT_LINE('-----------------------------------');
      FOR vid IN (
      select to_char(date_uploaded, 'YYYY') as v, count(to_char(date_uploaded, 'YYYY')) as s
       from gp_videos
       where duration_seconds <= 120
       group by to_char(date_uploaded, 'YYYY'))
   LOOP
      DBMS_OUTPUT.PUT_LINE('Year uploaded: ' || vid.v|| ' ' || 'Number of videos: ' || vid.s);
   END LOOP;

   DBMS_OUTPUT.PUT_LINE(' ');
   DBMS_OUTPUT.PUT_LINE('Video upload years for long videos');
   DBMS_OUTPUT.PUT_LINE('-----------------------------------');

   FOR vid in
        (
       select to_char(date_uploaded, 'YYYY') as v, count(to_char(date_uploaded, 'YYYY')) as l
       from gp_videos
       where duration_seconds >= 600
       group by to_char(date_uploaded, 'YYYY'))
    LOOP
        DBMS_OUTPUT.PUT_LINE('Year uploaded: ' || vid.v || ' ' || 'Number of videos: ' || vid.l);
    END LOOP;

end;
```

Figure 8: Question 5 Code Block

6. Find the most subscribed channel with videos typically made in the intermediate range of time (2-9 minutes). How does this compare with subscriber count for longer and shorter videos?
    a. **Justification:** This helps us understand the differences between short and long form content versus the bridge between the two and how these differences affect view count in corresponding videos. Also attempts to answer the question of which length of videos are usually preferred by the maximum amount of people.
    b. **Solution:** Create a procedure that first finds the number of subscribers among creators who make videos with an average video duration between 2 and 9 minutes. The program then sorts the number of subscribers into descending order and fetches the first row, which is the most subscribed channel. The procedure then finds the name of the creator using that information. We also made a copy of the gp_videos table and added a row there that classified each video into short, long, or intermediate. We created a trigger that would fire when a table was uploaded.
    c. **Results:**

        Creator name: Bhushan Kumar

        Number of Subscribers: 225000000
    d. **Insights:** In Question 1, we found the subscribers of the channels with the longest and shortest average video lengths. The highest number of subscribers of the influencer who makes videos in the intermediate range of time is much higher than the influencer with the highest average video duration. It is also much higher than the influencer with the shortest average video duration. We can see from this

that it's possible many people prefer videos that aren't too long or too short. It is likely that long videos are too long for people to stay engaged and interested, but short videos don't have enough content for people to want to subscribe to the channels that make them. Videos made in the intermediate range may be the most interesting to them. Using the copy of gp_videos, we also counted the number of intermediate length videos there were and compared it to the number of short and long videos. We found out that the number of long videos was the highest and intermediate length videos was the second highest. We can see from this that while intermediate length videos aren't the most common, they are still much more common than short videos, which suggests that people like those kinds of videos more than short videos.

```
create or replace procedure subscribed_channels
is
    subscriber_count NUMBER;
    creatorname VARCHAR2(64);

begin

    select total_subscribers
    into subscriber_count
    from gp_influencers join gp_videos on gp_influencers.video_title = gp_videos.video_title
    group by total_subscribers
    having avg(duration_seconds) between 121 and 599
    order by total_subscribers desc
    fetch first 1 row only;

    select creator_name
    into creatorname
    from gp_influencers join gp_videos on gp_influencers.video_title = gp_videos.video_title
    where total_subscribers = subscriber_count and rownum = 1
    group by creator_name
    having avg(duration_seconds) between 121 and 599;

    DBMS_OUTPUT.PUT_LINE('Creator name: ' || creatorname);
    DBMS_OUTPUT.PUT_LINE('Number of Subscribers: ' || subscriber_count);


end;
```

Figure 9: Question 6 Code Block

```sql
create table gp_videos_copy as select * from gp_videos;

CREATE OR REPLACE TRIGGER GP_PROJECT_INSERT_TRIGGER
    AFTER UPDATE ON GP_VIDEOS_COPY
BEGIN
    DBMS_OUTPUT.PUT_LINE('Table Updated.');
END;

alter table gp_videos_copy add video_form VARCHAR2(20);

update gp_videos_copy
set video_form = 'short'
where duration_seconds <= 120;

update gp_videos_copy
set video_form = 'long'
where duration_seconds >= 600;

update gp_videos_copy
set video_form = 'intermediate'
where duration_seconds between 121 and 599;
```

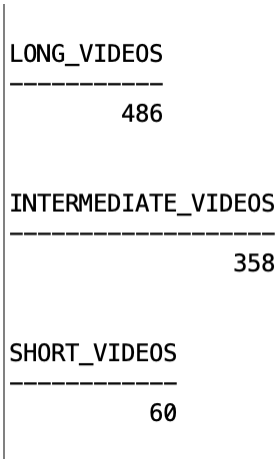
Figure 9.1: Copy of gp_videos and Trigger Creation with Queries to Add the Row video_form and Update the Table

```
Table Updated.

Table Updated.

Table Updated.
```

Figure 9.2: Question 6 Trigger Result



LONG_VIDEOS
486

INTERMEDIATE_VIDEOS
358

SHORT_VIDEOS
60

Figure 9.3: Results Showing Number of Videos for Each Type of Video

7. Find the newest channels with a large number of subscribers (greater than 100,000). What length of video do they make on average?
    a. **Justification:** This helps us find which lengths of videos are popular among large, newer creators of videos. This will help us analyze users' interaction with YouTube and trends related to lengths of videos. This will also lead to other potential questions having to do with subscriber count, views on videos, and lengths of videos that can further our insights gained. Content creators can use this information to tailor their content to what is currently most popular.
    b. **Solution:** Create a procedure named "FindNewChannelsWithSubscribers" that calculates the average video length of a content creator's videos when that creator's channel has been created recently. It uses joins to join information from two different tables and then a loop structure to print each occurrence.
    c. **Results:**

        Creator Name: Top Generality

        Total Subscribers: 255000

        Average Video Length: 1029

        Creator Name: Aperture

        Total Subscribers: 16000000

        Average Video Length: 1215

        Creator Name: Acvideos

        Total Subscribers: 229000

Average Video Length: 785

Creator Name: Chandoo

Total Subscribers: 296000

Average Video Length: 831

Creator Name: English With Ronnie · Englishlessons4U With Engvid

Total Subscribers: 4450000

Average Video Length: 543

d. **Insights:** The most conclusive result to be drawn from this output is that the newer channels with many subscribers are creating videos around 10 minutes long usually. When attempting to create videos of their own, new creators might be tempted to follow this model for video creation.

```
CREATE OR REPLACE PROCEDURE FindNewChannelsWithSubscribers
AS
BEGIN
    FOR ChannelInfo IN (
        SELECT
            Creator_Name,
            Total_Subscribers,
            AvgVideoLength
        FROM (
            SELECT
                GI.Creator_Name,
                GI.Total_Subscribers,
                AVG(GV.Duration_Seconds) OVER (PARTITION BY GI.Video_Title) AS AvgVideoLength,
                RANK() OVER (PARTITION BY GI.Video_Title ORDER BY GI.Channel_Created DESC) AS rnk
            FROM
                GP_Influencers GI
            JOIN GP_Videos GV ON GI.Video_Title = GV.Video_Title
            WHERE
                GI.Total_Subscribers > 100000
        )
        WHERE
            rnk = 1
    )
    LOOP
        DBMS_OUTPUT.PUT_LINE('Creator Name: ' || ChannelInfo.Creator_Name);
        DBMS_OUTPUT.PUT_LINE('Total Subscribers: ' || ChannelInfo.Total_Subscribers);
        DBMS_OUTPUT.PUT_LINE('Average Video Length: ' || ChannelInfo.AvgVideoLength);
        DBMS_OUTPUT.PUT_LINE('----------------------');
    END LOOP;
END FindNewChannelsWithSubscribers;
/
```

Figure 10: Question 7 Code Block

8. Find the total views that creators have achieved when creating long and short form content (less than two minutes vs greater than 10 minutes). Which group is cultivating a higher total number of views?
    a. **Justification:** This question will attempt to answer a question we had about whether long form content is gaining less views than short form content. Short form content platforms have become increasingly popular among consumers in recent years and it will be valuable for content creators to know whether this content is more successful than longer videos that were more common in the past.
    b. **Solution:** Create a procedure called “CalculateTotalViewsByDuration” which uses NVL() and SUM() to select videos under two minutes and videos over ten minutes into two separate variables. It then calculates the total views for each and outputs them to the user.
    c. **Results:**

        Total Views for Short Form Content: 808294369

        Total Views for Long Form Content: 5343225145
    d. **Insights:** It is conclusive in this dataset that the long form content views are receiving higher amounts of views. More would need to be done to explain why this is, but it is likely that the number of views could be attributed to a higher total number of videos, or that larger creators may be responsible for longer form content.

```
CREATE OR REPLACE PROCEDURE CalculateTotalViewsByDuration
AS
    vShortFormViews NUMBER := 0;
    vLongFormViews NUMBER := 0;
BEGIN

    SELECT
        NVL(SUM(GV.Total_Views), 0)
    INTO
        vShortFormViews
    FROM
        GP_Videos GV
    WHERE
        GV.Duration_Seconds < 120;


    SELECT
        NVL(SUM(GV.Total_Views), 0)
    INTO
        vLongFormViews
    FROM
        GP_Videos GV
    WHERE
        GV.Duration_Seconds > 600;


    DBMS_OUTPUT.PUT_LINE('Total Views for Short Form Content: ' || vShortFormViews);
    DBMS_OUTPUT.PUT_LINE('Total Views for Long Form Content: ' || vLongFormViews);

END CalculateTotalViewsByDuration;
/
```

Figure 11: Question 8 Code Block

9. Find the average video length of a video with more than 500,000 views, and the average video length of a video with less than that. Do longer or shorter videos get more views?
    a. **Justification:** This question is useful because it looks at the larger videos in the dataset and gives an average video length of all of those videos so we can analyze what that is. It then will do the same for videos that are smaller and compute a result to compare the original average to. This will allow for the video length data and number of views to be correlated easily by providing an easily digestible output to analyze.
    b. **Solution:** Create a new procedure called "CalculateAverageVideoLength" which selects videos with both more and less than 500,000 views into different variables based on their average duration in seconds. It them prints the results of these averages for each corresponding group.
    c. **Results:**

        Average Video Length for Videos with More Than 500,000 Views:
        2474.25172413793103448275862068965517241 4 seconds

        Average Video Length for Videos with Less Than or Equal to 500,000 Views:
        970.765822784810126582278481012658227848 seconds

    d. **Insights:** This is an interesting result because it shows that longer videos once again are garnering more views. This time we can see that the average video length is nearly three times as long for videos with more than 500,000 views than it is for videos with less than or equal to 500,000 views. Based off this query, we can more conclusively say that creators may want to increase their video length to attract more views.

```
CREATE OR REPLACE PROCEDURE CalculateAverageVideoLength AS
    v_more_than_500k_avg_duration NUMBER;
    v_less_than_or_equal_500k_avg_duration NUMBER;
BEGIN

    SELECT AVG(Duration_Seconds)
    INTO v_more_than_500k_avg_duration
    FROM GP_Videos
    WHERE Total_Views > 500000;


    SELECT AVG(Duration_Seconds)
    INTO v_less_than_or_equal_500k_avg_duration
    FROM GP_Videos
    WHERE Total_Views <= 500000;


    DBMS_OUTPUT.PUT_LINE('Average Video Length for Videos with More Than 500,000 Views: ' || v_more_than_500k_avg_dur
    DBMS_OUTPUT.PUT_LINE('Average Video Length for Videos with Less Than or Equal to 500,000 Views: ' || v_less_than_
END CalculateAverageVideoLength;
/
SET SERVEROUTPUT ON;
EXEC CalculateAverageVideoLength;
```

Figure 12: Question 9 Code Block

10. Can hashtags affect the popularity of a video? Do they help videos get more views or go more viral?
    a. **Justification:** This question can help us understand how factors like hashtags can affect the success of a video. Videos with hashtags may have more views as it is likely easier to access for non-subscribers, which means that it has a bigger chance of going viral.
    b. **Solution:** Create a procedure called count_hashtags that finds the average number of hashtags of videos with over 500,000 views and the average number of hashtags of videos with less than 500,000 views. Print the average number of hashtags for both types of videos.
    c. **Results:**

        Videos with over 500,000 views

        1.086882453151618398637137989778534923 34

        - - - - - - - - - - - -

        Videos with under 500,000 views

        .930599369085173501577287066246056782334 4

    d. **Insights:** This procedure shows that videos with over 500,000 views have a higher average number of hashtags than videos with less than 500,000 views. This suggests that hashtags do have an influence on more successful videos. However,

the difference is quite minimal. There is not a significant difference between the two averages. From this, we can conclude that while having more hashtags could make the chance of a video having more views higher, it does not have a big effect.

```
create or replace procedure count_hashtags
is
    hashtag_count NUMBER;
    low_hashtag_count NUMBER;

begin
    DBMS_OUTPUT.PUT_LINE('Videos with over 500,000 views');
    DBMS_OUTPUT.PUT_LINE('-------------------------------');

    select avg(total_hashtags)
    into hashtag_count
    from gp_videos
    where total_views > 500000;

    DBMS_OUTPUT.PUT_LINE(hashtag_count);


   DBMS_OUTPUT.PUT_LINE(' ');
   DBMS_OUTPUT.PUT_LINE('Videos with under 500,000 views');
   DBMS_OUTPUT.PUT_LINE('--------------------------------');

   select avg(total_hashtags)
   into low_hashtag_count
   from gp_videos
   where total_views <= 500000;

   DBMS_OUTPUT.PUT_LINE(low_hashtag_count);

end;
```

Figure 13: Question 10 Code Block