Milestone 1 | YouTube Trending Analysis

INTRODUCTION: Welcome to your first Milestone. Milestones are a great opportunity for you to practice your skills, both in using SQL, but also in interpreting the information that comes out of the queries you write like a true data analyst.

In this Milestone, you'll take on the role of a Junior Data Analyst at a digital media consultancy that helps brands and content creators optimize their online presence. Your manager has asked you to conduct an initial analysis using YouTube trending video data. She wants insights on how factors like views, likes, content category, and publication timing influence a video's likelihood of trending. Your findings will help inform data-driven recommendations for a client on how to maximize content's reach and engagement.

HOW IT WORKS: Follow the prompts in the tasks below to investigate your data. Post your answers in the provided boxes: the **yellow boxes** for the queries you write, and **blue boxes** for text-based answers. When you're done, export your document as a pdf file and upload it to HQ for feedback from The Accelerator Team. Please don't *ever* remove (paste your query below 👇) or (write your **answer** below 👇). These help your Evaluator!

SQL APP: Here's that link to our specialized SQL app, where you'll write your SQL queries and interact with the data.

— Data Set **Description**

The YouTube Trending Videos (`youtube.trending`) consists of 6351 videos that were listed in the Trending Videos in the United States, recorded between November 2017 and June 2018. There are 16 columns in the dataset; we'll be working with the following columns in this SkillBuilder: `title`, `channel_title`, `views`, `likes`, and `dislikes`.

— **Task 1:** Top User Engagement

To start, you've been asked to look at the videos with the highest levels of user engagement, in terms of likes, dislikes, and comments.

A. Write a query that returns these columns: title, channel_title, views, likes, dislikes, and comment_count. Run your query to ensure it returns the desired data then copy the query into the box below once you've verified that it does what you want.

(paste your query below 👇)

```
SELECT title, channel_title, views, likes, dislikes,
comment_count
FROM youtube.trending
```

B. Add `ORDER BY` to your query in order to find the video that has the highest number of likes. What is the name of that video? Post your query into the yellow box, and the name of the most-liked video in the blue box.

(paste your query below 👇)

```
SELECT title, channel_title, views, likes, dislikes,
comment_count
FROM youtube.trending
ORDER BY likes DESC
```

(write your **answer** below 👇)

BTS (방탄소년단) 'FAKE LOVE' Official MV

C. Modify the ORDER BY line in your query to find the video that has the highest number of dislikes. What is the name of that video? (As before, a query goes in the yellow box, a text answer in the blue box.)

(paste your query below 👇)

```
SELECT title, channel_title, views, likes, dislikes,
comment_count
FROM youtube.trending
ORDER BY dislikes DESC
```

(write your **answer** below 👇)

So Sorry.

D. Modify the ORDER BY line in your query once more to find the name of the video with the highest number of comments? What is the video?

(paste your query below 👇)

```
SELECT title, channel_title, views, likes, dislikes,
comment_count
FROM youtube.trending
ORDER BY comment_count DESC
```

(write your **answer** below 👇)

So Sorry.

Try this prompt: What are some factors that might explain *why* a YouTube video with a high number of dislikes also has a high number of comments?

– **Task 2:** Comments Counts Large and Small

Your employer wants to go further into the patterns of user engagement via comments on top trending videos.

A. Continuing from the queries of the previous task, modify the query to return only the top 10 videos in terms of comment count.

(paste your query below 👇)

```
SELECT title, channel_title, views, likes, dislikes,
comment_count
FROM youtube.trending
ORDER BY comment_count DESC
LIMIT 10
```

B. How many comments are on the 10th-most-commented video? What is the ratio of this comment count to the top commented video (Task 1D)?

HINT: The ratio is obtained by dividing the first number by the second. This should be done with a calculator outside of SQL using what was returned from the part A query, and not with a new SQL query.

(write your **answer** below 👇)

371864 comments on the 10th-most-commented-video. We have ratio Top 1 = 3.661 x Top10

C. Let's dig deeper down the rankings. What is the number of comments on the 100th-ranked video? Use the `OFFSET` keyword to skip past the top 99 videos so that the first row returned will be the 100th rank. (In other words, don't just change the `LIMIT` to 100 and check the last row returned.)

(paste your query below 👇)

```
SELECT title, channel_title, views, likes, dislikes,
comment_count
FROM youtube.trending
ORDER BY comment_count DESC
LIMIT 1
OFFSET 99
```

(write your **answer** below 👇)

WINNER - 'EVERYDAY' M/V ; 53665 comments

D. Okay, let's take one more step down the rankings. What is the number of comments on the 1000th-ranked video?

Try this prompt: I need to find the 1000th-ranked video in a dataset ordered by comment count. How do I determine what values to use for OFFSET and LIMIT to retrieve only the 1000th row?

(paste your query below 👇)

```
SELECT title, channel_title, views, likes, dislikes,
comment_count
FROM youtube.trending
ORDER BY comment_count DESC
LIMIT 1
```

```
OFFSET 999
```

(write your **answer** below 👇)

7155

— Task 3: Analyzing and Reflecting

A. Let's reflect on the data we just looked at. In each step from part B through D, we looked at the 10th, 100th, and 1000th most-commented videos, a 10-fold increase in rank number. How different are the videos from one another at the top rankings compared to those in the middle rankings in terms of comment count? Write a sentence or two to summarize what your takeaways are.

Try this prompt: I'm analyzing a dataset of YouTube videos ranked by comment count. I've compared the 10th, 100th, and 1000th most-commented videos. What kinds of patterns or trends should I look for in the comment counts as I move further down the rankings? Can you suggest any questions I should ask myself to better understand the differences between these rankings?

(write your **answer** below 👇)

We can see that the higher the rank, the bigger count of comments these videos have. The engagement rate by number of comments decreases exponentially when we move to the lower rank; as dropping from 10th to 100th and 100th to 1000th have a consistent reduction rate of 86-87%.

— LevelUp

A. Do you observe the same pattern you did with comment count with likes, dislikes, or views? Write several SQL queries to find out.

(write your **answer** below 👇)

Similar to engagement by comment count, the decline (of likes/dislikes) from top ranking videos to middle-ranking ones are also very steep. The number of likes drops by 78% and 87% respectively from top 10th to 100th and 100th to 1000th. The decline rate is even steeper when sorting data with most disliked videos, as the number of dislikes drop by around 84.7% from 10th to 100th and 89.6% from 100th to 1000th.

— Submission

Great work completing your first Milestone! To submit your completed Milestone, you will need to download / export this document as a PDF and then upload it to the Milestone submission page for feedback from The Accelerator Team. You can find the option to download as a PDF from the File menu in the upper-left corner of the Google Doc interface.