[Course](https://lms.codinginvaders.com/courses/course-v1:CodingInvaders+DATEST+1/course/) [Project 4 - SQL](https://lms.codinginvaders.com/courses/course-v1:CodingInvaders+DATEST+1/course/#block-v1:CodingInvaders+DATEST+1+type@chapter+block@b1099f73750e4cb0b9590d31491ad3e2) [Kickoff SQL project 2](https://lms.codinginvaders.com/courses/course-v1:CodingInvaders+DATEST+1/course/#block-v1:CodingInvaders+DATEST+1+type@sequential+block@dcfee7c3662245968388aae03917c977) Project Canvas

**Business Problem:**

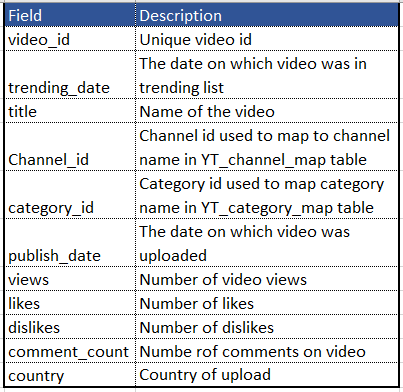
YouTube’s trending videos vary by location across the world. The impact of likes, dislikes, comments may differ based on the countries. The demographics of countries itself is a big factor in the way videos are consumed in various countries. The videos can be trending if video views are above a certain level. Your manager wants to present an analysis how average trending periods of videos vary across countries. How the likes, dislikes, comments impact the duration of trending videos. Your manager has also asked for some interesting insights from the data which could be done through exploratory data analysis.

**Dataset concept:** There are 3 tables available for the analysis -

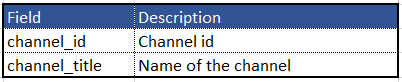
1. **YT\_trending\_videos:** This table has video level information along with dates on which videos were trending along with metrics such as comments, likes, views, etc.
2. **YT\_channel\_map:** This table has channel\_id mapping with the channel title which is also interpreted as channel name
3. **YT\_category\_map:** This table has category\_id mapping. The videos on YouTube are mapped to a category based on the type of video such as Movie, trailer, animation, etc.

**Data Dictionary:**

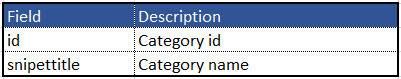
**YT\_trending\_videos**

****

**YT\_channel\_map**

****

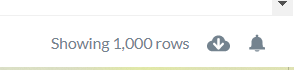
**YT\_category\_map**

****

**How to submit the reports?**

Most of the question consists of creating a report on the project. It is followed by a simple question to verify if the report was generated correctly or not. While answers to questions need to be submitted, the report should be downloaded in the excel. Consolidate all the reports in one excel, where Sheet name should be the **assignment number**. Save the excel by the name SQL\_P2\_<firstname>\_<lastname>.xlsx

When the report is executed in Metabase, it can be downloaded by clicking on the button on the right bottom corner as mentioned below.





[Course](https://lms.codinginvaders.com/courses/course-v1:CodingInvaders+DATEST+1/course/) [Project 4 - SQL](https://lms.codinginvaders.com/courses/course-v1:CodingInvaders+DATEST+1/course/#block-v1:CodingInvaders+DATEST+1+type@chapter+block@b1099f73750e4cb0b9590d31491ad3e2) [SQL Project 2 Task](https://lms.codinginvaders.com/courses/course-v1:CodingInvaders+DATEST+1/course/#block-v1:CodingInvaders+DATEST+1+type@sequential+block@a882629140c445698b14598e07517591) Assignments

## **Assignments**

Bookmark this page

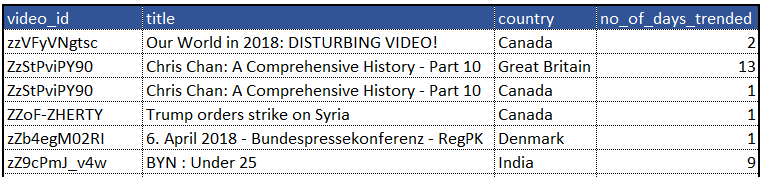
**Brief for Assignment 1, 2, & 3:**

We need to understand the distribution of duration of trending videos. The videos can trend on multiple days. The popular videos such as new songs, trailer trend longer compared to one of the viral videos. Hence it is important to understand the distribution of duration of videos trending by their respective category. We will need to run a few sql queries to analyze the distribution.

### Assignment 1

1/1 point (graded)

Create a report for overall distribution of duration of trending videos by each country. The report will have video\_id, video name (title), country, No\_of\_days\_trended. Sample report:



**Task:** Save the report with assignment number in the Google work sheet as mentioned previously.

**Question**: How many videos have trended for more than 5 days in the US?

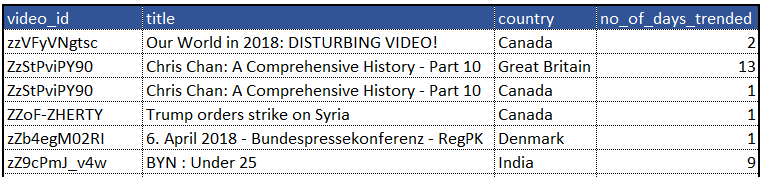
/\*1 point possible (graded)

Create a report for overall distribution of duration of trending videos by each country. The report will have video\_id, video name (title), country, No\_of\_days\_trended\*/

SELECT video\_id, title, country, COUNT(trending\_date) AS No\_of\_days\_trended

FROM yt\_trending\_videos

GROUP BY video\_id, title, country;



/\* Question: How many videos have trended for more than 5 days in the US?\*/

SELECT COUNT(\*)

FROM

( SELECT video\_id, title, country, COUNT(trending\_date) AS No\_of\_days\_trended

FROM yt\_trending\_videos

WHERE country = "US"

GROUP BY video\_id, title, country

HAVING No\_of\_days\_trended > 5)B;

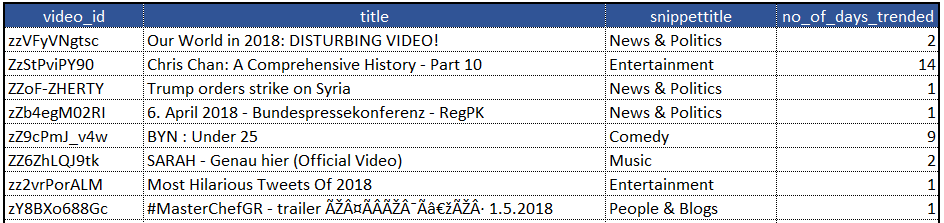


### Assignment 2

1 point possible (graded)

Create a report for overall distribution of duration of trending videos by each category. The report will have video\_id, category title, No\_of\_days\_trended.

Sample report:



**Task:** Save the report with assignment number in the Google work sheet as mentioned previously.

**Question:** Which category has the highest average trending period?

1. Hint (1 of 1): Join is required

/\*Assignment 2\*/

/\*1 point possible (graded)

Create a report for overall distribution of duration of trending videos by each category. The report will have video\_id, category title, No\_of\_days\_trended.\*/

SELECT \* FROM yt\_category\_map;

SELECT \* FROM yt\_channel\_map;

SELECT \* FROM yt\_trending\_videos;

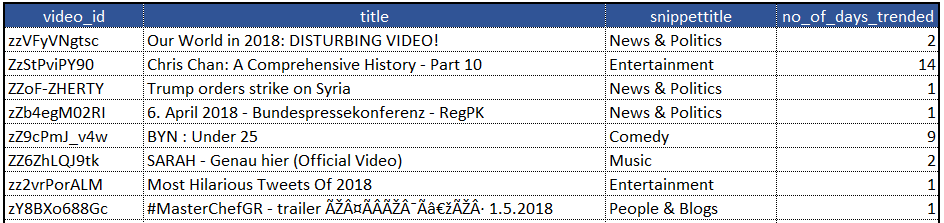
SELECT video\_id, title, snippettitle, COUNT(trending\_date) AS No\_of\_days\_trended

FROM yt\_trending\_videos A

JOIN yt\_category\_map B

ON A.video\_id = B.id

GROUP BY video\_id, title, country;



/\* Question: Which category has the highest average trending period?\*/

SELECT snippettitle AS category\_title,

Avg(no\_of\_days\_trended) AS avg\_trending\_days

FROM (SELECT video\_id,

title,

snippettitle,

Count(trending\_date) AS no\_of\_days\_trended

FROM yt\_trending\_videos a

INNER JOIN yt\_category\_map b

ON a.category\_id = b.id

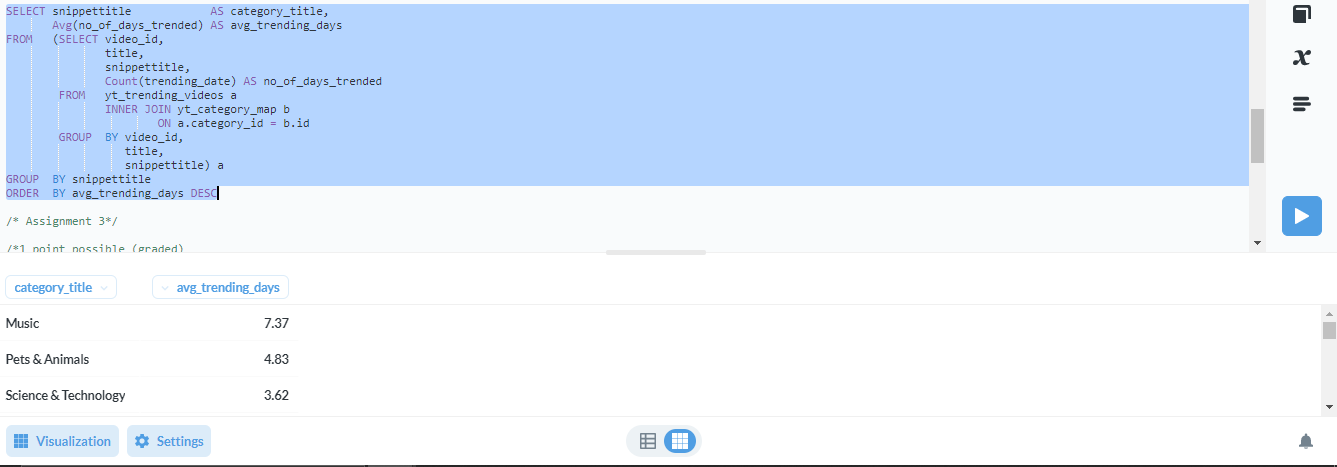
GROUP BY video\_id,

title,

snippettitle) a

GROUP BY snippettitle

ORDER BY avg\_trending\_days DESC

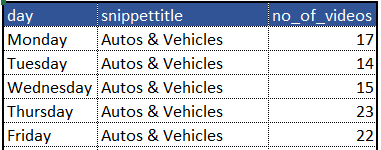


7.37

### Assignment 3

1 point possible (graded)

Create a report for the number of distinct videos trending from each category on day of the week. The report will have weekday, category title, No\_videos\_trended. Sample report:



**Note:** The answer cannot be derived from the report. In the report, we have a number of videos trending for each weekday by category. Hence, if a video has trended on multiple days, it will be counted on each day. The question asks for distinct videos and hence write the query considering that.

**Task:** Save the report with assignment number in the Google work sheet as mentioned previously.

**Question**: How many distinct videos trended from the category ‘Music’ on weekdays (Monday - Friday)?

/\* Assignment 3\*/

/\*1 point possible (graded)

Create a report for the number of distinct videos trending from each category on day of the week. The report will have weekday, category title, No\_videos\_trended.\*/

SELECT \* FROM yt\_category\_map;

SELECT \* FROM yt\_channel\_map;

SELECT \* FROM yt\_trending\_videos;

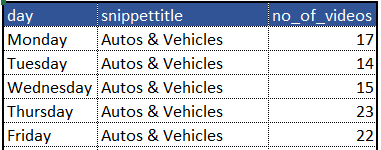
SELECT Weekday(trending\_date) AS day, snippettitle, COUNT(DISTINCT trending\_date) AS No\_videos\_trended

FROM yt\_trending\_videos A

JOIN yt\_category\_map B

ON B.id = A.video\_id

GROUP BY snippettitle, Weekday(trending\_date);



/\*Hint (1 of 1): Use Weekday function\*/

SELECT \* FROM yt\_category\_map WHERE snippettitle = "Music";

/\*Question: How many distinct videos trended from the category ‘Music’ on weekdays (Monday - Friday)?\*/

SELECT snippettitle,

Count(DISTINCT video\_id) AS no\_of\_videos

FROM yt\_trending\_videos a

INNER JOIN yt\_category\_map b

ON a.category\_id = b.id

WHERE Weekday(trending\_date) BETWEEN 0 AND 4

AND snippettitle = 'Music'

GROUP BY snippettitle;



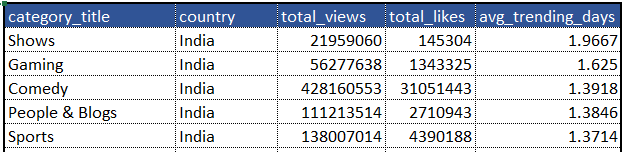
**Brief for Assignment 4 & 5:**

Normally it is believed if the video trends for more number of days, then the video views, likes and comments also is higher. We need to check whether this is true for all the videos.

### Assignment 4

1 point possible (graded)

Create a summary report which contains country, category title, total\_views, total\_likes and avg\_trending\_days. Sample report:



**Task:** Save the report with assignment number in the Google work sheet as mentioned previously.

**Question**: What are the total views for category sports in ‘Canada’?

/\* Assignment 4\*/

/\*1 point possible (graded)

Create a summary report which contains country, category title, total\_views, total\_likes and avg\_trending\_days.\*/

SELECT \* FROM yt\_category\_map;

SELECT \* FROM yt\_channel\_map;

SELECT \* FROM yt\_trending\_videos;

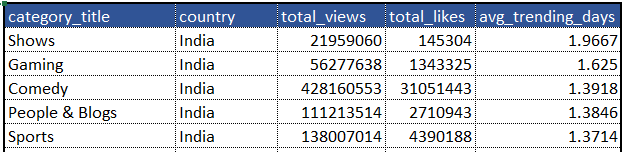
SELECT country, snippettitle AS category\_title, COUNT(views) AS total\_views, COUNT(likes) AS total\_likes, AVG(trending\_date) AS avg\_trending\_days

FROM yt\_trending\_videos A

JOIN yt\_category\_map B

ON A.category\_id = B.id

GROUP BY country, snippettitle



/\*Question: What are the total views for category sports in ‘Canada’?\*/

SELECT country, snippettitle AS category\_title, SUM(views) AS total\_views, SUM(likes) AS total\_likes, AVG(trending\_date) AS avg\_trending\_days

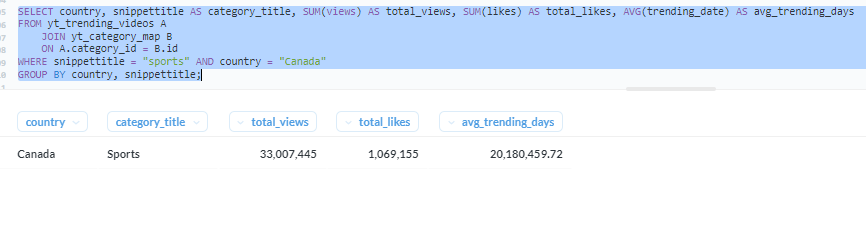
FROM yt\_trending\_videos A

JOIN yt\_category\_map B

ON A.category\_id = B.id

WHERE snippettitle = "sports" AND country = "Canada"

GROUP BY country, snippettitle;



### Assignment 5

1/1 point (graded)

Rank the videos based on views, likes within each country. Which country has the highest number of videos with rank for views and rank of likes both in top 20?

1. Hint (1 of 1):  
   1. Rank the views and likes in the descending order, such that highest views will have rank 1.  
   2. Apply the filter for top 20 in outer query.

/\*Assignment 5\*/

/\*1 point possible (graded)

Rank the videos based on views, likes within each country. Which country has the highest number of videos with rank for views and rank of likes both in top 20?\*/

SELECT \* FROM yt\_category\_map;

SELECT \* FROM yt\_channel\_map;

SELECT \* FROM yt\_trending\_videos;

/\* Rank the videos based on views, likes within each country\*/

SELECT video\_id, country, views, likes,

RANK() OVER(

PARTITION BY country ORDER BY views DESC, likes DESC

) RANK\_VIEOS

FROM yt\_trending\_videos;

/\*Which country has the highest number of videos with rank for views and rank of likes both in top 20?\*/

SELECT country, COUNT(trending\_date) AS No\_videos\_trended

FROM( SELECT \*, RANK()

OVER(

PARTITION BY country ORDER BY views DESC

) AS rank\_views,

RANK()

OVER(

PARTITION BY country ORDER BY likes DESC

) AS rank\_likes

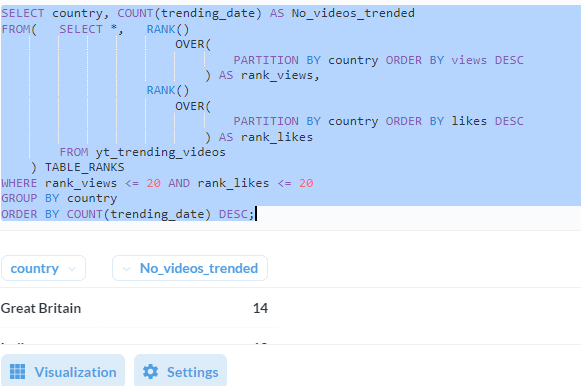
FROM yt\_trending\_videos

) TABLE\_RANKS

WHERE rank\_views <= 20 AND rank\_likes <= 20

GROUP BY country

ORDER BY COUNT(trending\_date) DESC;



**Brief for Assignment 6 & 7:**

We have determined that the Likes to dislikes ratio is a good indicator of popularity. We have to come up with a rating framework to the videos based on the available metrics, which would help us recommend videos better.

### Assignment 6

1 point possible (graded)

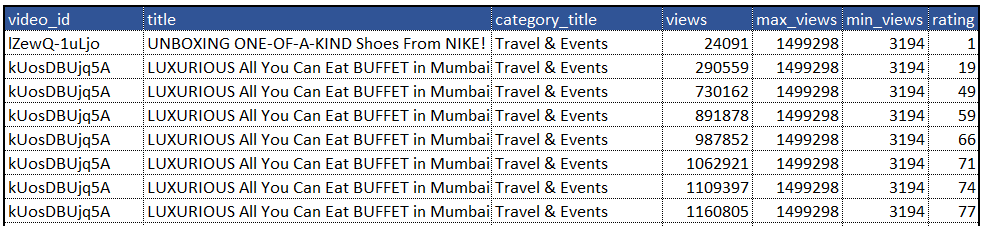
Generate a report at video level with video viewership rating within the category. (Report at video level means the output should be unique at video\_id).

**Formula to assign the rating: ((Views - min(views))\*100 ) / max(views) - min(views)**

where **max(views)** is maximum views in the respective video’s category

and **min(views)** is minimum views in the respective video’s category

Sample report:



**Task:** Save the report with assignment number in the Google work sheet as mentioned previously.

**Question:** What is the average rating of the category Music?

/\*Assignment 6

1 point possible (graded)

Generate a report at video level with video viewership rating within the category. (Report at video level means the output should be unique at video\_id).

Formula to assign the rating: ((Views - min(views))\*100 ) / max(views) - min(views)

where max(views) is maximum views in the respective video's category

and min(views) is minimum views in the respective video's category\*/

SELECT \* FROM yt\_category\_map;

SELECT \* FROM yt\_channel\_map;

SELECT \* FROM yt\_trending\_videos;

SELECT video\_id, title, snippettitle AS category\_title, views, MAX(Views) AS MAX\_VIEWS, MIN(views) AS MIN\_VIEWS, ((Views - min(views))\*100 ) / max(views) - min(views) AS rating

FROM yt\_trending\_videos A

JOIN yt\_category\_map B

ON A.category\_id = B.id

GROUP BY video\_id, title, snippettitle, views;

/\*Question: What is the average rating of the category Music?\*/

SELECT category\_title,

Avg(rating) AS avg\_rating

FROM (SELECT c.\*,

Round(( ( views - min\_views ) \* 100 ) / ( max\_views - min\_views )

, 0) AS

rating

FROM (SELECT DISTINCT video\_id,

title,

snippettitle AS category\_title,

views,

Max(views)

OVER (

partition BY category\_id) AS max\_views,

Min(views)

OVER (

partition BY category\_id) AS min\_views

FROM yt\_trending\_videos a

INNER JOIN yt\_category\_map b

ON a.category\_id = b.id) c) d

GROUP BY category\_title;

### Assignment 7

1/1 point (graded)

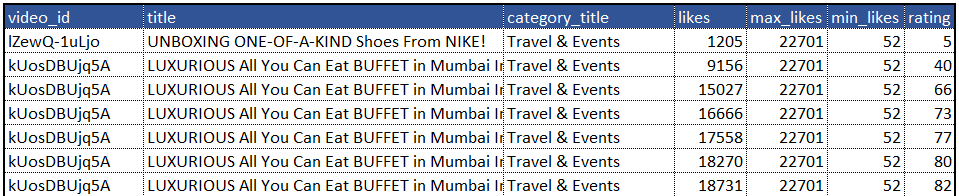
Generate a report at video level with video ratings. (Report at video level means the output should be unique at video\_id).

**Rating formula to assign the rating: ((Likes - min(likes))\*100 ) / max(likes) - min(likes)**

where **max(likes)** is maximum likes in the respective video’s category

and **min(likes)** is minimum likes in the respective video’s category

Sample report:



**Task:** Save the report with assignment number in the Google work sheet as mentioned previously.

**Question:** Which category has the highest average rating based on likes?

SELECT category\_title,

Avg(rating) AS avg\_rating

FROM (SELECT c.\*,

Round(( ( likes - min\_likes ) \* 100 ) / ( max\_likes - min\_likes )

, 0) AS

rating

FROM (SELECT DISTINCT video\_id,

title,

snippettitle AS category\_title,

likes,

Max(likes)

OVER (

partition BY category\_id) AS max\_likes,

Min(likes)

OVER (

partition BY category\_id) AS min\_likes

FROM yt\_trending\_videos a

INNER JOIN yt\_category\_map b

ON a.category\_id = b.id) c) d

GROUP BY category\_title

ORDER BY avg\_rating DESC ;



## **Creative Question**

Bookmark this page

**Creative Question [For the presentation]**

Based on the knowledge acquired about the database while solving the above question, you are expected to do an exploratory data analysis. The analysis can focus on either of the questions below.

* Views are most affected by which of the metrics?
  + 1. Likes
    2. dislikes
    3. comment count
* What is the optimum trending duration?
  1. Trending videos for long duration may see saturation in video views or even decline
  2. Analyze and come up with optimum trending duration for each category
* Detailed channel level analysis, based on frequency of upload and how it impacts trending duration. (One impact could be, channels uploading videos frequently might result in lower trending periods. This is called cannibalization)

**Presentation Task:**

Choose any 2 points for the analysis. You can write SQL queries to get the data at a certain level, export it into Google sheet and analyse it there. Use charts in Google Sheet to justify your analysis and create a PowerPoint presentation.

**Important Deadlines:**

Submit the project report [**here**](https://forms.gle/zaAXPEJfuihmYjkS9) by 09/05/2021

Presentation will be on 16/05/2021.

**Note:** Group names will be shared separately.

**Points for the good presentation:**

1. Not all the analysis that is done needs to be added. Presentation should be crisp, with insights and meaning full analysis
2. It should not exceed 7-8 slides
3. In case of charts, highlight the trend which you want audience to see
4. Balance of visualization and text. The slides should not be just visualization or just text.
5. Don't put a lot of content into one slide, as it affects readability
6. Most importantly, you are presenting insights and findings and not the analysis itself. Hence choose the best chart or couple of charts to showcase the analysis even though you would have analysed more charts.
7. You can keep supporting slides in appendix, if you feel there can be some questions which could be explained well through a slide but cannot be part of main presentation