COS736 – Big Data Engineering

Isaac Dumisani Mtsweni

4126704@myuwc.ac.za

Capstone Project 2025

Question 1B (Inference & Insights):

1. 6 HOURS SEGMENTS

```sql
-- Q1B.1: Count uploads per 6-hr segment
SELECT segment, COUNT(*) AS uploads
FROM (
  SELECT
    CASE
      WHEN TIME(publish_time) >= '00:00:00' AND TIME(publish_time) < '06:00:00' THEN '00:00-06:00'
      WHEN TIME(publish_time) >= '06:00:00' AND TIME(publish_time) < '12:00:00' THEN '06:00-12:00'
      WHEN TIME(publish_time) >= '12:00:00' AND TIME(publish_time) < '18:00:00' THEN '12:00-18:00'
      ELSE '18:00-24:00'
    END AS segment
  FROM all_videos
) x
GROUP BY segment
ORDER BY uploads DESC;
```

segment	uploads
12:00–18:00	17062
18:00–24:00	12715
00:00–06:00	7149
06:00–12:00	4023

    (a) Most uploads: 18:00–24:00
    (b) Fewest uploads: 00:00–06:00

2. TOP 3 CATEGORIES IN YOUTUBE

```sql
-- Q1B.2 (views version): top 3 categories by total views
SELECT category_id, SUM(views) AS total_views
FROM all_videos
GROUP BY category_id
ORDER BY total_views DESC
LIMIT 3;
```
→

category_id	total_views
10	40132892190
24	20604388195
1	7284156721

```sql
-- Q1B.2 (count version): top 3 categories by number of videos
SELECT category_id, COUNT(*) AS videos
FROM all_videos
GROUP BY category_id
ORDER BY videos DESC
LIMIT 3;
```
→

category_id	videos
24	9964
10	6472
26	4146

3. TOP 3 MOST LIKED VIDEOS

```sql
31
32     -- Q1B.3: top 3 most liked videos
33   WITH ranked AS (
34       SELECT
35           v.video_id,
36           v.title,
37           v.channel_title    AS author,
38           v.publish_time,
39           v.likes,
40           ROW_NUMBER() OVER (
41               PARTITION BY v.video_id
42               ORDER BY v.likes DESC, v.publish_time DESC
43           ) AS rn
44       FROM US_VIDEOS v
45   )
46   SELECT
47       video_id,
48       title,
49       author,
50       publish_time,
51       likes
52   FROM ranked
53   WHERE rn = 1
54   ORDER BY likes DESC
55   LIMIT 3;
56
```

	video_id	title	author	publish_time	likes
	Filter	Filter	Filter	Filter	Filter
1	7C2z4GqqS5E	BTS (방탄소년단) 'FAKE LOVE' Official MV	ibighit	2018-05-18 09:00:02	5613827
2	VYOjWnS4cMY	Childish Gambino - This Is America (Official Video)	ChildishGambinoVEVO	2018-05-06 04:00:07	5023450
3	ffxKSjUwKdU	Ariana Grande - No Tears Left To Cry	ArianaGrandeVevo	2018-04-20 04:00:03	3094021

4. 3 VIDEOS HAD THE MOST VIEWER ENGAGEMENT

```sql
57     -- Q1B.4
58
59   WITH ranked AS (
60       SELECT
61           v.video_id,
62           v.title,
63           v.channel_title    AS author,
64           v.publish_time,
65           v.comment_count,
66           ROW_NUMBER() OVER (
67               PARTITION BY v.video_id
68               ORDER BY v.comment_count DESC, v.publish_time DESC
69           ) AS rn
70       FROM US_VIDEOS v
71   )
72   SELECT
73       video_id,
74       title,
75       author,
76       publish_time,
77       comment_countd
78   FROM ranked
79   WHERE rn = 1
80   ORDER BY comment_count DESC
81   LIMIT 3;
82
```

	video_id	title	author	publish_time	comment_count
	Filter	Filter	Filter	Filter	Filter
1	QwZT7T-TXT0	So Sorry.	Logan Paul Vlogs	2018-01-02 16:42:21	1361580
2	7C2z4GqqS5E	BTS (방탄소년단) 'FAKE LOVE' Official MV	ibighit	2018-05-18 09:00:02	1228655
3	FlsCjmMhFmw	YouTube Rewind: The Shape of 2017	#YouTubeRewind	YouTube Spotlight	2017-12-06 17:58:51	827755

5. HOW MANY VIDEOS ARE IN EACH CATEGORY

```
-- Q1B.5

SELECT category_id, COUNT(*) AS video_count
FROM US_VIDEOS
GROUP BY category_id
ORDER BY video_count DESC;


SELECT category_id, COUNT(DISTINCT video_id) AS unique_video_count
FROM US_VIDEOS
GROUP BY category_id
ORDER BY unique_video_count DESC;
```

→

Messages	Result 1	Result 2	
	category_id	video_count	
	Filter	Filter	
1	24	9964	
2	10	6472	
3	26	4146	
4	23	3457	
5	22	3210	
6	25	2487	
7	28	2401	
8	1	2345	
9	17	2174	
10	27	1656	
11	15	920	
12	20	817	
13	19	402	
14	2	384	
15	29	57	
16	43	57	

→

Messages	Result 1	Result 2	
	category_id	unique_video_count	
	Filter	Filter	
1	24	1622	
2	10	801	
3	26	595	
4	23	547	
5	25	507	
6	22	498	
7	17	452	
8	28	381	
9	1	321	
10	27	251	
11	15	141	
12	20	103	
13	2	71	
14	19	60	
15	29	14	
16	43	4	