Most Important Emphasis

1. Examine the distribution of course types across categories to uncover trends and insights, enabling the client to strategically determine which course types to launch in specific categories for maximum impact and alignment with learner demand, also count the number of courses by category and sub-category.

2. Calculate the average number of views for each category, sub-category, and language to provide insights into viewer engagement patterns and inform strategic content development.

3. Identify the most commonly taught skills in today's educational landscape based on the data given based on category to ensure course offerings remain relevant and aligned with current job market demands.

4. What is the distribution of various Languages in which a particular course is created?

5. Determine the language preferences for each category based on viewer preferences, so that clients can optimise course accessibility and better align content with audience demand. Clients only want to analyse this data for the top 5 categories based on user preferences.

6. Investigate the relationship between the availability of subtitles and the number of views for courses to determine how subtitle options may impact viewer engagement and accessibility.

7. Identify the top three instructors for each category and subcategory based on ratings to highlight educators who consistently deliver high-quality content and effectively engage learners so that they can be approached by your client to make content for them and make this visual as static.

8. Examine the relationship between course duration and the number of views to understand how the length of a course may influence viewer engagement and preferences for each category and sub-category, if course duration has a month (in each month only 60 hours of content) and for flexible schedules make the timing as 200 hours.

9. In the context of recorded lectures, we need to investigate whether the variety of skills offered within each category and subcategory has a measurable impact on viewership

10. Analyze the correlation between course ratings and the number of viewers across different categories and subcategories to identify whether highly-rated courses attract more engagement and to help optimize content quality.

11. Determine the impact of the number of instructors per course on the course rating and viewership to assess whether having multiple instructors contributes to higher engagement and better learner satisfaction.

12. Examine the distribution of program types (e.g., specialization, professional certificate) across categories and subcategories to identify the most popular program formats and their alignment with user preferences.

13. Assess the demand for courses across different platforms (e.g., Coursera, Udemy, edX) by category and subcategory to help clients identify the best platforms for launching new courses.

14. Identify the top emerging skills across different categories based on recent course enrollments to ensure course offerings stay updated with industry trends and workforce demands.

SQL

-- 1. Examine the distribution of course types across categories and count courses by category and subcategory
```
SELECT `Category`, `Sub-Category`, `Course Type`, COUNT(*) AS course_count
FROM online_courses
GROUP BY `Category`, `Sub-Category`, `Course Type`;
```

-- 2. Calculate the average number of views for each category, subcategory, and language
```
SELECT `Category`, `Sub-Category`, `Language`, AVG(`Views`) AS avg_views
FROM online_courses
GROUP BY `Category`, `Sub-Category`, `Language`;
```

-- 3. Identify the most commonly taught skills in each category
```
SELECT `Category`, `Skills include`, COUNT(*) AS skill_count
FROM online_courses
GROUP BY `Category`, `Skills include`
ORDER BY skill_count DESC;
```

```
-- 4. Distribution of languages in which a particular course is created
SELECT `Course Title`, `Language`, COUNT(*) AS language_count
FROM online_courses
GROUP BY `Course Title`, `Language`;

-- 5. Determine the language preferences for the top 5 categories based on user preferences
WITH TopCategories AS (
   SELECT `Category`, SUM(`Views`) AS total_views
   FROM online_courses
   GROUP BY `Category`
   ORDER BY total_views DESC
   LIMIT 5
)
SELECT c.`Category`, c.`Language`, COUNT(*) AS course_count
FROM online_courses c
JOIN TopCategories tc ON c.`Category` = tc.`Category`
GROUP BY c.`Category`, c.`Language`;

-- 6. Investigate the relationship between subtitles availability and number of views
SELECT `Has Subtitles`, AVG(`Views`) AS avg_views
FROM online_courses
GROUP BY `Has Subtitles`;

-- 7. Identify the top three instructors for each category and subcategory based on ratings
WITH RankedInstructors AS (
   SELECT `Category`, `Sub-Category`, `Created by` AS instructor, AVG(`Rating`) AS
avg_rating,
        RANK() OVER (PARTITION BY `Category`, `Sub-Category` ORDER BY
AVG(`Rating`) DESC) AS rnk
   FROM online_courses
   GROUP BY `Category`, `Sub-Category`, `Created by`
)
SELECT `Category`, `Sub-Category`, instructor, avg_rating
FROM RankedInstructors
WHERE rnk <= 3;

-- 8. Relationship between course duration and views
SELECT `Category`, `Sub-Category`, `Content Duration`, AVG(`Views`) AS avg_views
FROM online_courses
GROUP BY `Category`, `Sub-Category`, `Content Duration`;

-- Adjust course duration for month and flexible schedules
UPDATE online_courses
SET `Content Duration` = CASE
```

```
    WHEN `Content Duration` LIKE '%month%' THEN '60 hours'
    WHEN `Content Duration` LIKE '%flexible%' THEN '200 hours'
    ELSE `Content Duration`
END;

-- 9. Investigate the impact of skill variety on viewership
SELECT `Category`, `Sub-Category`, COUNT(DISTINCT `Skills include`) AS skill_variety,
AVG(`Views`) AS avg_views
FROM online_courses
GROUP BY `Category`, `Sub-Category`;

-- 10. Analyze correlation between course ratings and number of viewers
SELECT `Category`, `Sub-Category`, `Rating`, AVG(`Views`) AS avg_views
FROM online_courses
GROUP BY `Category`, `Sub-Category`, `Rating`;

-- 11. Determine the impact of the number of instructors per course on ratings and viewership
SELECT `Course Title`, COUNT(DISTINCT `Created by`) AS instructor_count,
AVG(`Rating`) AS avg_rating, AVG(`Views`) AS avg_views
FROM online_courses
GROUP BY `Course Title`;

-- 12. Distribution of program types across categories and subcategories
SELECT `Category`, `Sub-Category`, `Program`, COUNT(*) AS program_count
FROM online_courses
GROUP BY `Category`, `Sub-Category`, `Program`;

-- 13. Assess demand for courses across different platforms by category and subcategory
SELECT `Category`, `Sub-Category`, `Platform`, COUNT(*) AS course_count
FROM online_courses
GROUP BY `Category`, `Sub-Category`, `Platform`;

-- 14. Identify the top emerging skills based on recent enrollments
SELECT `Category`, `Skills include`, SUM(`Enrollments`) AS total_enrollments
FROM online_courses
GROUP BY `Category`, `Skills include`
ORDER BY total_enrollments DESC;
```