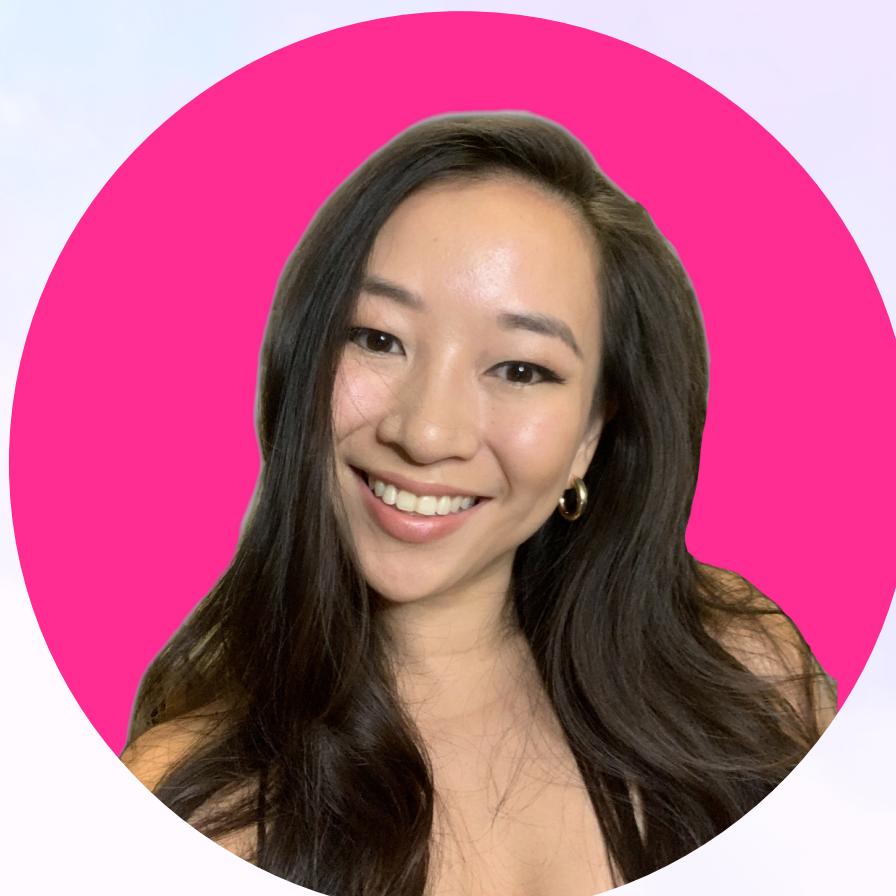


Beginner to Advanced

SQL Interview

Questions

10 questions & answers
with practice datasets



Dawn Choo

**Applying for Data Analyst jobs?
Applying for Data Scientist jobs?
Applying for Data Engineering jobs?**

You need to be a master at SQL.



**Here are 10 SQL interview
questions to practice with
from Beginner to Advanced levels**

Together with answers
& datasets

1. Netflix movies & shows

Dataset

<https://www.kaggle.com/datasets/rahulvyasm/netflix-movies-and-tv-shows/data>

You have this dataset, **netflix_titles**, that contains comprehensive information about movies and TV shows available on Netflix.

The columns in your dataset are as follows:

- **show_id**: A unique identifier for each title
- **type**: The category of the title
- **title**: The name of the movie or TV show
- **director**: The director(s) of the movie or TV show
- **cast**: The list of main actors or actresses in the title
- **country**: The countries where the movie or show was produced
- **date_added**: The date the title was added to Netflix
- **release_year**: The year the movie or TV show was released
- **rating**: The age rating of the title
- **duration**: The duration of the title
- **listed_in**: The genres the title falls under
- **description**: A brief summary of the title

1. Netflix movies & shows

Questions

Beginner

1. How many movies and TV shows are there in the dataset? Display the count for each type.
2. What percentage of content doesn't have a country associated with it?

Intermediate

1. Find the top 3 directors with the most content on Netflix. Display the director's name, the count of their titles, and the year of their most recent content.
2. For each year from 2015 to 2021, calculate the percentage of movies vs TV shows added to Netflix

Advanced

1. Calculate the average month-over-month growth rate of content added to Netflix for each genre. What are the top 5 fastest growing genres?

1. Netflix movies & shows

Answers

Beginner Question 1:

How many movies and TV shows are there in the dataset? Display the count for each type.

```
...  
SELECT type, COUNT(*) as count  
FROM netflix_titles  
GROUP BY type;
```

Beginner Question 2:

What percentage of content doesn't have a country associated with it?

```
...  
SELECT  
    COUNT(CASE WHEN country IS NULL THEN 1 END)  
    / COUNT(*) AS percentage_without_country  
FROM  
    netflix_titles;
```

1. Netflix movies & shows

Answers

Intermediate Question 1:

Find the top 3 directors with the most content on Netflix. Display the director's name, the count of their titles, and the year of their most recent content.

```
...  
WITH director_stats AS (  
    SELECT  
        director,  
        COUNT(*) AS title_count,  
        MAX(release_year) AS most_recent_year  
    FROM netflix_titles  
    WHERE director != '' AND director IS NOT NULL  
    GROUP BY director  
)  
SELECT  
    director,  
    title_count,  
    most_recent_year  
FROM director_stats  
ORDER BY title_count DESC  
LIMIT 3;
```

1. Netflix movies & shows

Answers

Intermediate Question 2:

For each year from 2015 to 2021, calculate the percentage of movies vs TV shows added to Netflix.

```
...  
WITH yearly_counts AS (  
    SELECT  
        EXTRACT(YEAR FROM DATE(date_added)) AS year,  
        type,  
        COUNT(*) AS count  
    FROM netflix_titles  
    WHERE date_added BETWEEN '2015-01-01' AND '2021-12-31'  
    GROUP BY 1,2  
)  
SELECT  
    year,  
    SUM(CASE WHEN type = 'Movie' THEN count ELSE 0 END)  
        / SUM(count) AS movie_percentage,  
    SUM(CASE WHEN type = 'TV Show' THEN count ELSE 0 END)  
        / SUM(count) AS tv_show_percentage  
FROM yearly_counts  
GROUP BY year  
ORDER BY year;
```

1. Netflix movies & shows

Answers

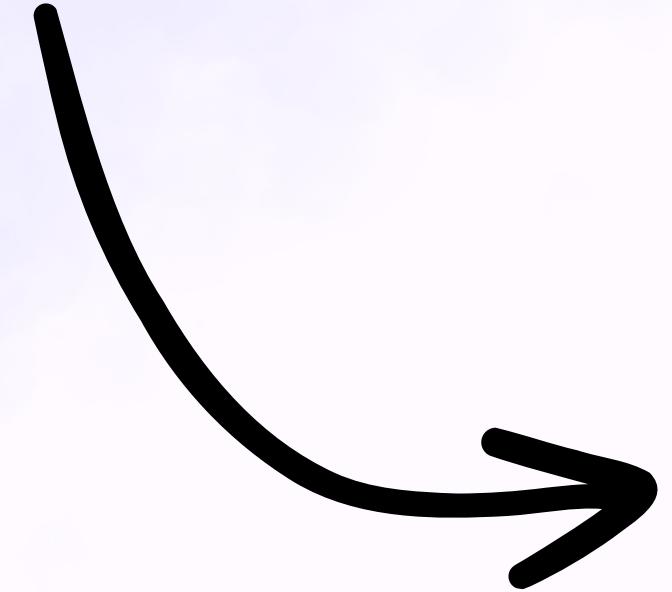
Advanced Question 1:

Calculate the average month-over-month growth rate of content added to Netflix for each genre. What are the top 5 fastest growing genres?

```
WITH genre_months AS (
    SELECT
        DATE_TRUNC('month', DATE(date_added)) AS month,
        UNNEST(STRING_TO_ARRAY(listed_in, ', ')) AS genre,
        COUNT(*) AS monthly_count
    FROM netflix_titles
    WHERE date_added IS NOT NULL
    GROUP BY DATE_TRUNC('month', DATE(date_added)), genre
),
growth_rates AS (
    SELECT
        genre,
        month,
        monthly_count,
        LAG(monthly_count) OVER (PARTITION BY genre ORDER BY month)
            AS prev_month_count,
        (monthly_count - prev_month_count)::float /
            NULLIF(prev_month_count, 0) AS growth_rate
    FROM genre_months
),
avg_growth_rates AS (
    SELECT
        genre,
        AVG(growth_rate) AS avg_growth_rate
    FROM growth_rates
    WHERE growth_rate IS NOT NULL
    GROUP BY genre
)
SELECT
    genre,
    avg_growth_rate
FROM avg_growth_rates
ORDER BY avg_growth_rate DESC
LIMIT 5;
```

**Ready for the next
set of questions?**

Let's do it



2. Employee Information Dataset

[https://www.kaggle.com/datasets/tawfikelmetwally/
employee-dataset](https://www.kaggle.com/datasets/tawfikelmetwally/employee-dataset)

You have this dataset called **employee_information** that contains data about employees in a company.

The columns in your dataset are as follows:

- **Education**: Highest level of education and field of study
- **JoiningYear**: Year the employee started at the company
- **City**: Employee's work location
- **PaymentTier**: Salary category
- **Age**: Current age of the employee
- **Gender**: Employee's gender identity
- **EverBenched**: If employee was ever without assigned work
- **ExperienceInCurrentDomain**: Years in current field
- **LeaveOrNot**: Whether employee has taken leave

2. Employee Information Questions

Beginner

1. How many employees are in each Payment Tier?
2. Calculate the average age of employees for each gender.

Intermediate

1. Find the top 3 cities with the highest percentage of employees who have been benched at least once.
2. For each Education level, calculate the average experience in the current domain for employees who joined before 2015 and those who joined in 2015 or later.

Advanced

1. Calculate an Employee's likelihood to leave based on their Payment Tier, Experience in Current Domain, and the average leave rate of their city.

2. Employee Information

Answers

Beginner Question 1:

How many employees are in each Payment Tier?

```
...  
SELECT PaymentTier, COUNT(*) AS EmployeeCount  
FROM EmployeeInformation  
GROUP BY PaymentTier  
ORDER BY PaymentTier;
```

Beginner Question 2:

Calculate the average age of employees for each gender.

```
...  
SELECT Gender, AVG(Age) AS AverageAge  
FROM EmployeeInformation  
GROUP BY Gender;
```

2. Employee Information

Answers

Intermediate Question 1:

Find the top 3 cities with the highest percentage of employees who have been benched at least once.

```
SELECT City,
       COUNT(*) AS TotalEmployees,
       SUM(CASE WHEN EverBenched = 'Yes' THEN 1 ELSE 0 END)
             AS BenchedEmployees,
       BenchedEmployees / COUNT(*) AS BenchedPercentage
FROM EmployeeInformation
GROUP BY City
ORDER BY BenchedPercentage DESC
LIMIT 3;
```

Intermediate Question 2:

For each Education level, calculate the average experience in the current domain for employees who joined before 2015 vs. in 2015 or later.

```
SELECT Education,
       AVG(CASE WHEN JoiningYear < 2015 THEN ExperienceInCurrentDomain END)
             AS AvgExperienceBefore2015,
       AVG(CASE WHEN JoiningYear >= 2015 THEN ExperienceInCurrentDomain END)
             AS AvgExperience2015OrLater
FROM EmployeeInformation
GROUP BY Education;
```

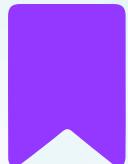
2. Employee Information Answers

Advanced Question 1:

Calculate an Employee's likelihood to leave based on their Payment Tier, Experience in Current Domain, and the average leave rate of their city.

```
WITH CityLeaveRates AS (
    SELECT City,
        AVG(CASE WHEN LeaveOrNot = 'Yes' THEN 1 ELSE 0 END)
        AS CityLeaveRate
    FROM EmployeeInformation
    GROUP BY City
)
SELECT
    ei.PaymentTier,
    CASE
        WHEN ei.ExperienceInCurrentDomain <= 3 THEN 'Low'
        WHEN ei.ExperienceInCurrentDomain <=7 THEN 'Medium'
        ELSE 'High'
    END AS ExperienceLevel,
    ei.City,
    clr.CityLeaveRate,
    COUNT(*) AS TotalEmployees,
    SUM(CASE WHEN ei.LeaveOrNot = 'Yes' THEN 1 ELSE 0 END)
        AS LeavingEmployees,
    LeavingEmployees / COUNT(*) AS GroupLeaveProbability
FROM EmployeeInformation ei
JOIN CityLeaveRates clr ON ei.City = clr.City
GROUP BY
    1,2,3,4
ORDER BY 1,2,3;
```

Found this useful?

 Save it

 Follow me

 Share with a friend



Dawn Choo