SQL

Interview questions

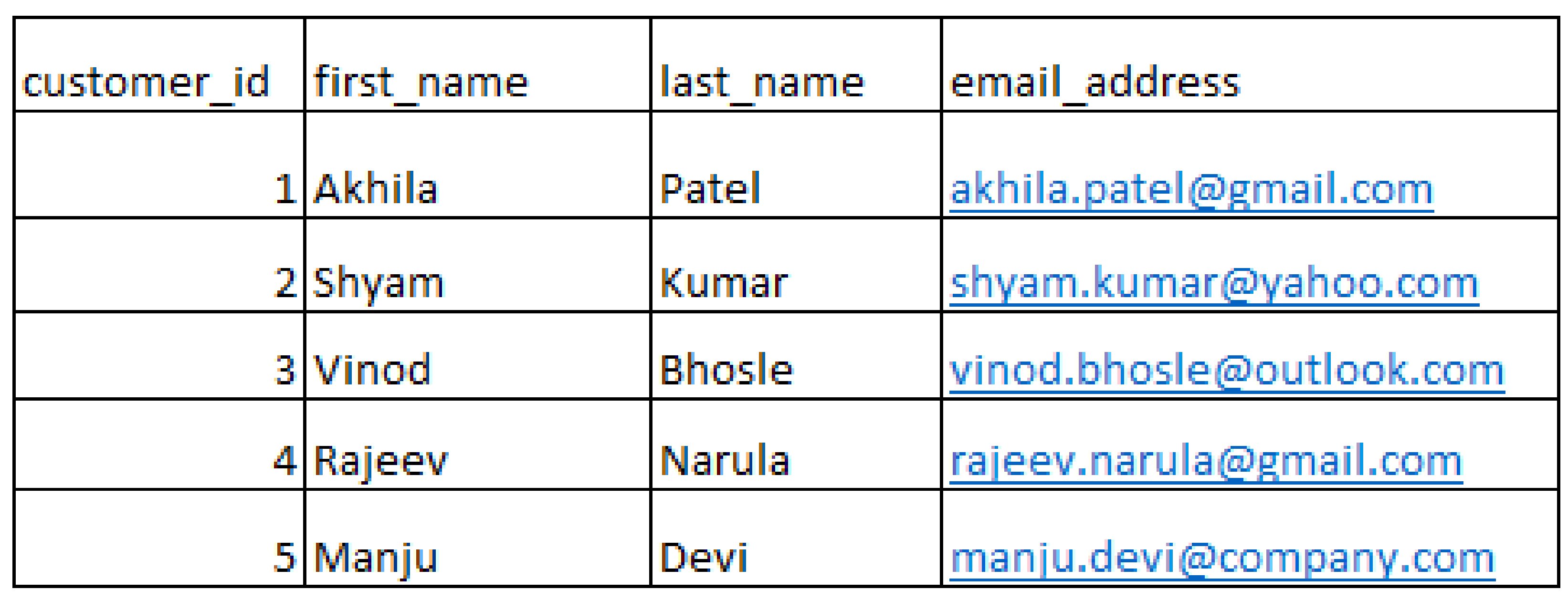
for Data Analysts

Part II

linkedin.com/in/ileonjose

1. Find Customers with Specific Email Providers

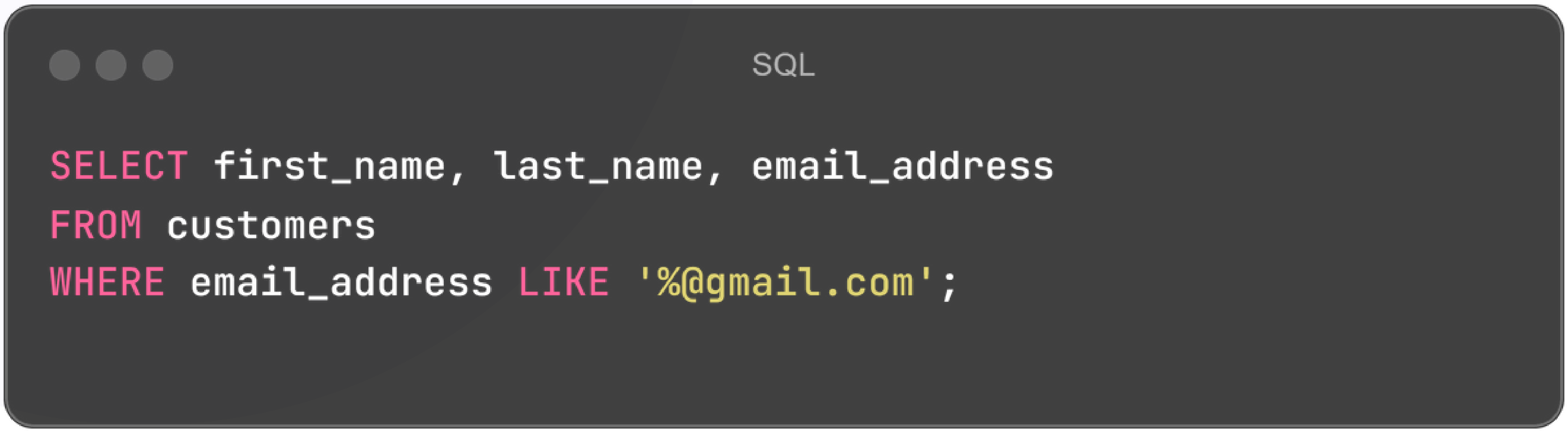
- **Problem Statement:** List all customers whose email addresses use a specific email provider, such as Gmail. The customer data is stored in a table with customer IDs, names, and email addresses.



customer_id	first_name	last_name	email_address
1	Akhila	Patel	akhila.patel@gmail.com
2	Shyam	Kumar	shyam.kumar@yahoo.com
3	Vinod	Bhosle	vinod.bhosle@outlook.com
4	Rajeev	Narula	rajeev.narula@gmail.com
5	Manju	Devi	manju.devi@company.com

How to Solve:

- Use the LIKE operator to filter email addresses by provider.
- Use % as a wildcard to match any email address ending with the specified provider.



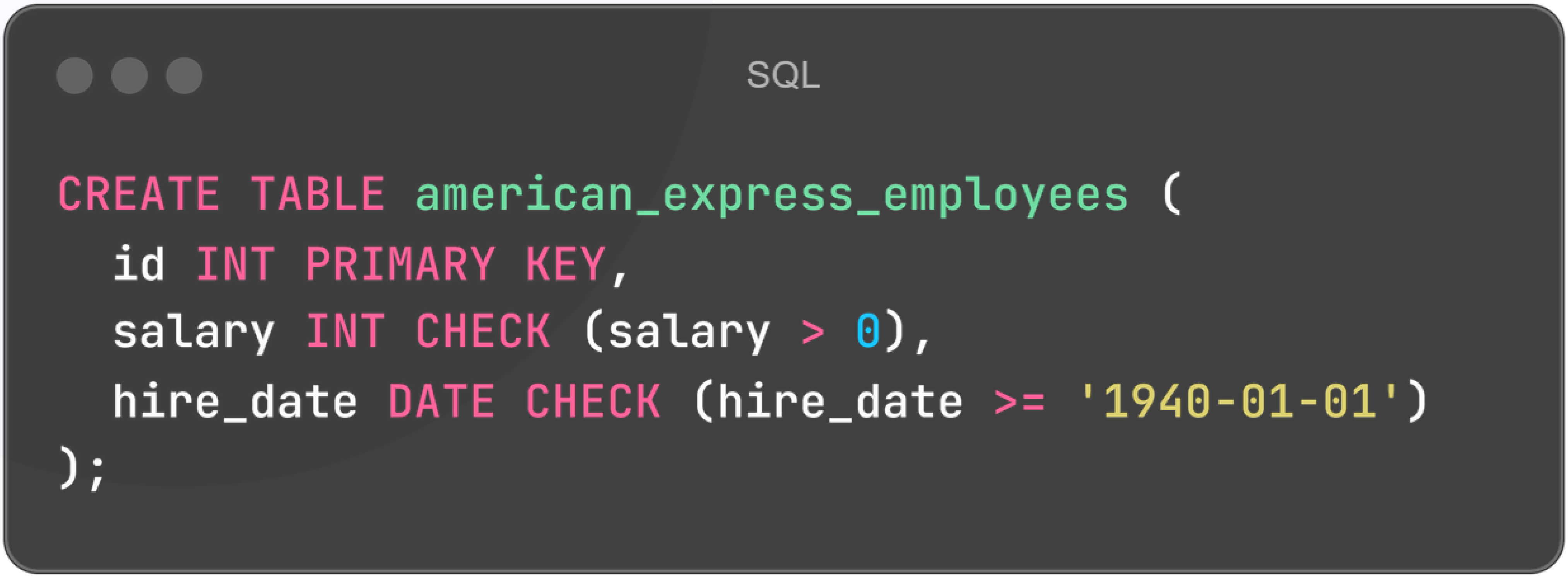
```sql
SELECT first_name, last_name, email_address
FROM customers
WHERE email_address LIKE '%@gmail.com';
```

2. How Does the CHECK Constraint Function?

Problem Statement: Explain the CHECK constraint in SQL and provide an example of how it can be used to ensure data integrity.

How to Solve:

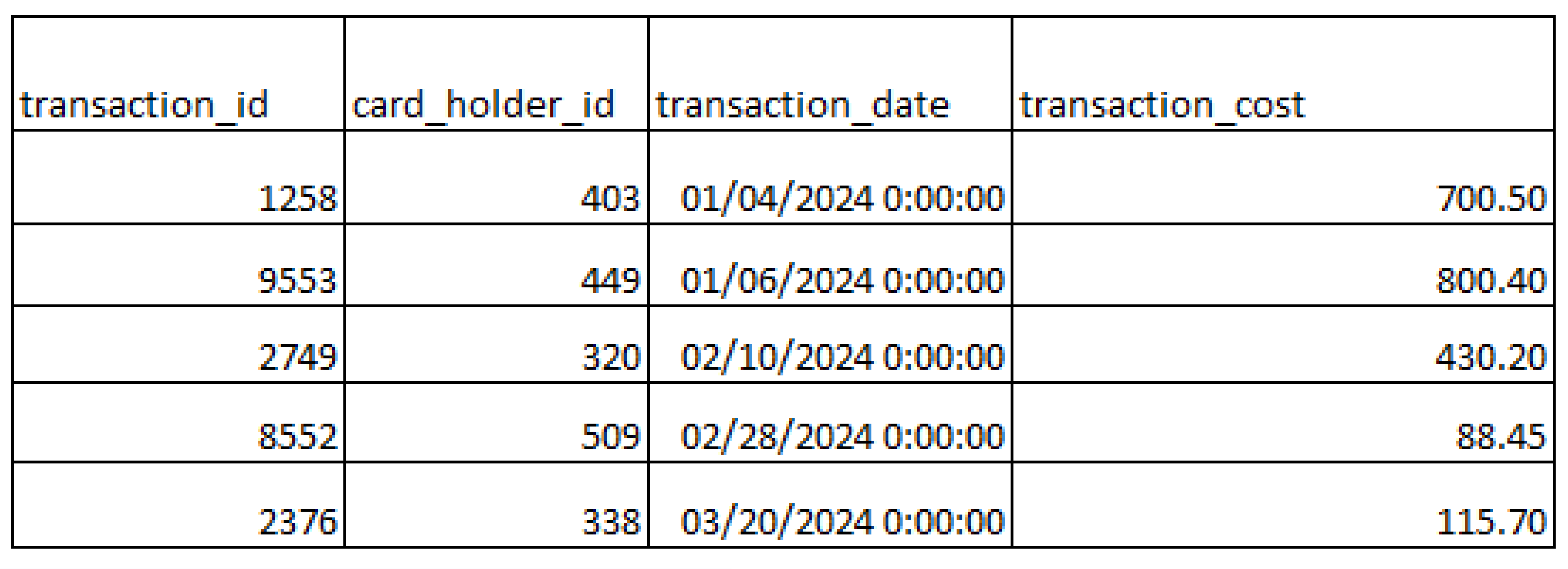
- Define the CHECK constraint as a rule applied to column values.
- Provide examples to show its usage in maintaining data quality, such as ensuring positive values or date ranges.



```sql
CREATE TABLE american_express_employees (
  id INT PRIMARY KEY,
  salary INT CHECK (salary > 0),
  hire_date DATE CHECK (hire_date >= '1940-01-01')
);
```

3. Calculate Average Card Usage Per Month

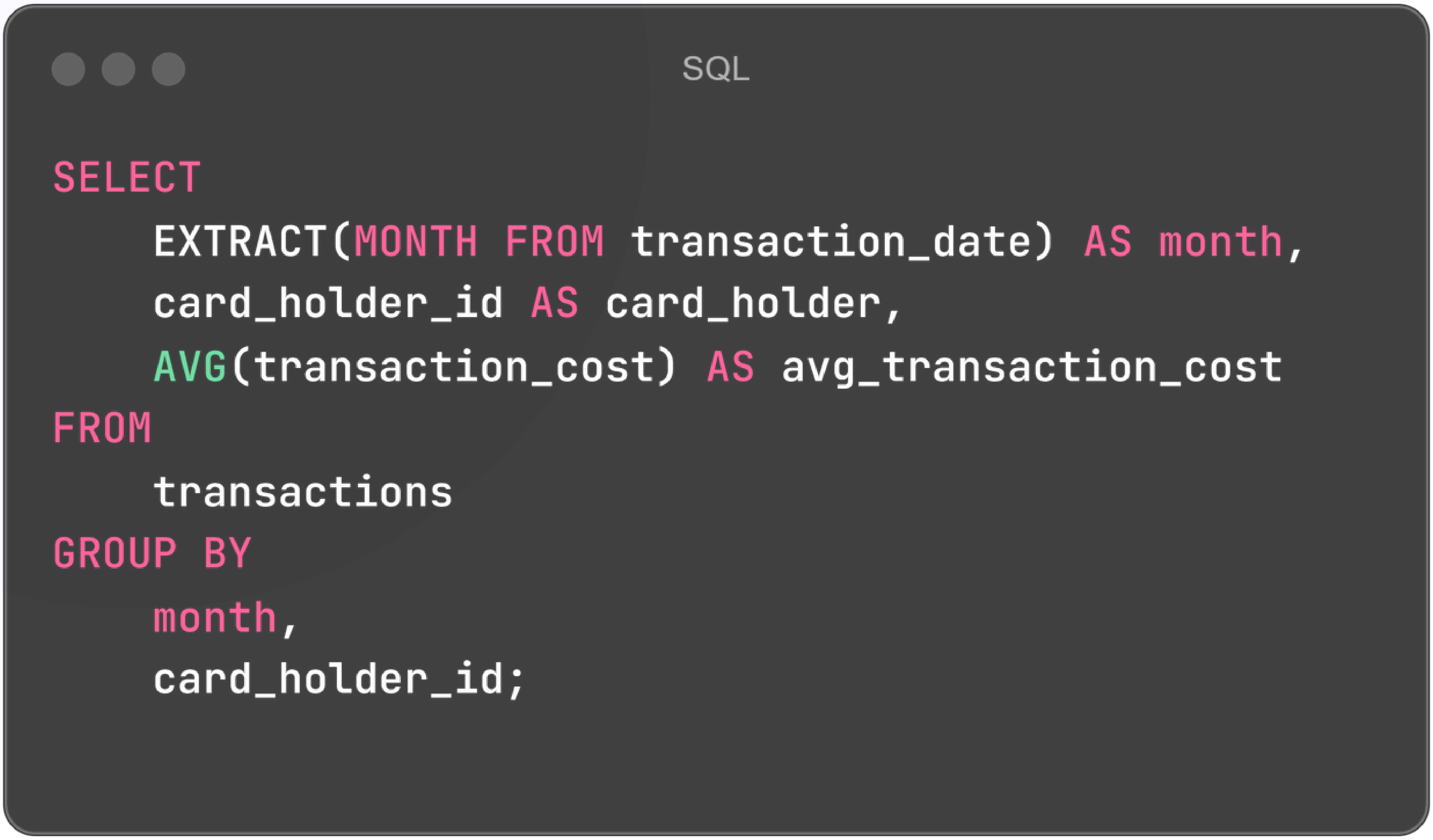
Problem Statement: Find the average transaction cost per cardholder for each month. The transactions are recorded in a table with transaction IDs, cardholder IDs, transaction dates, and transaction costs.



transaction_id	card_holder_id	transaction_date	transaction_cost
1258	403	01/04/2024 0:00:00	700.50
9553	449	01/06/2024 0:00:00	800.40
2749	320	02/10/2024 0:00:00	430.20
8552	509	02/28/2024 0:00:00	88.45
2376	338	03/20/2024 0:00:00	115.70

How to Solve:

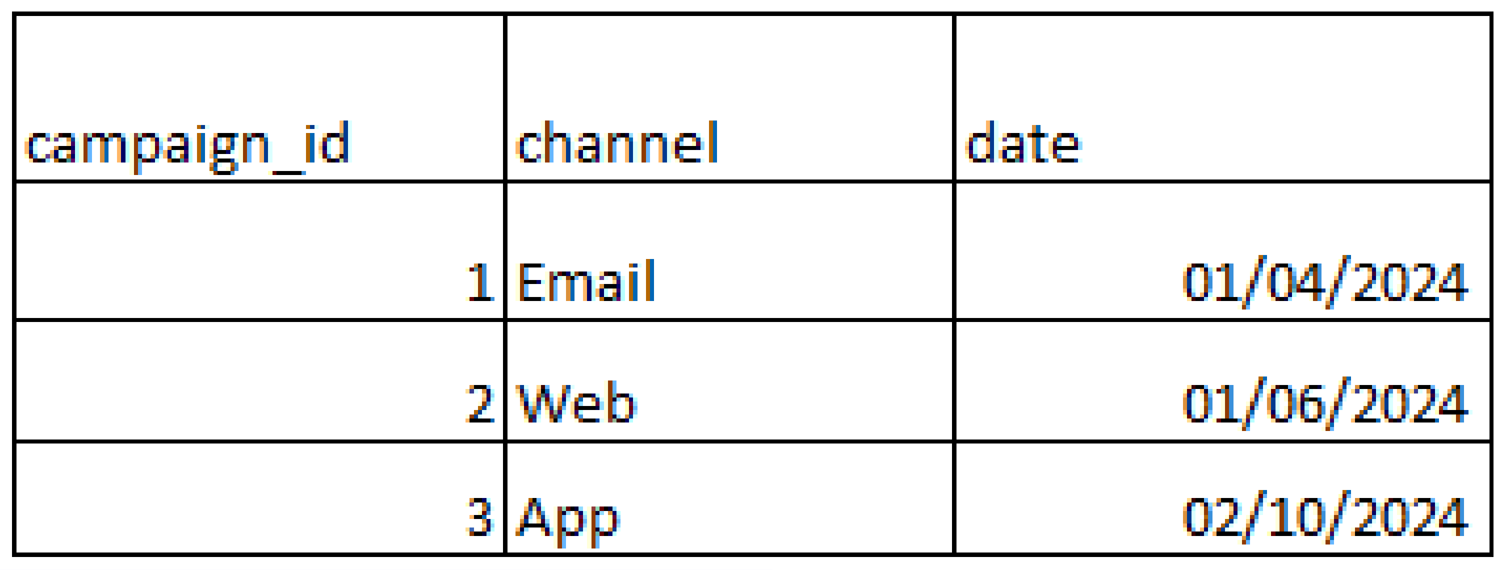
- Extract month from *transaction_date*.
- Group by month and *card_holder_id*.
- Calculate the average transaction cost for each cardholder each month.



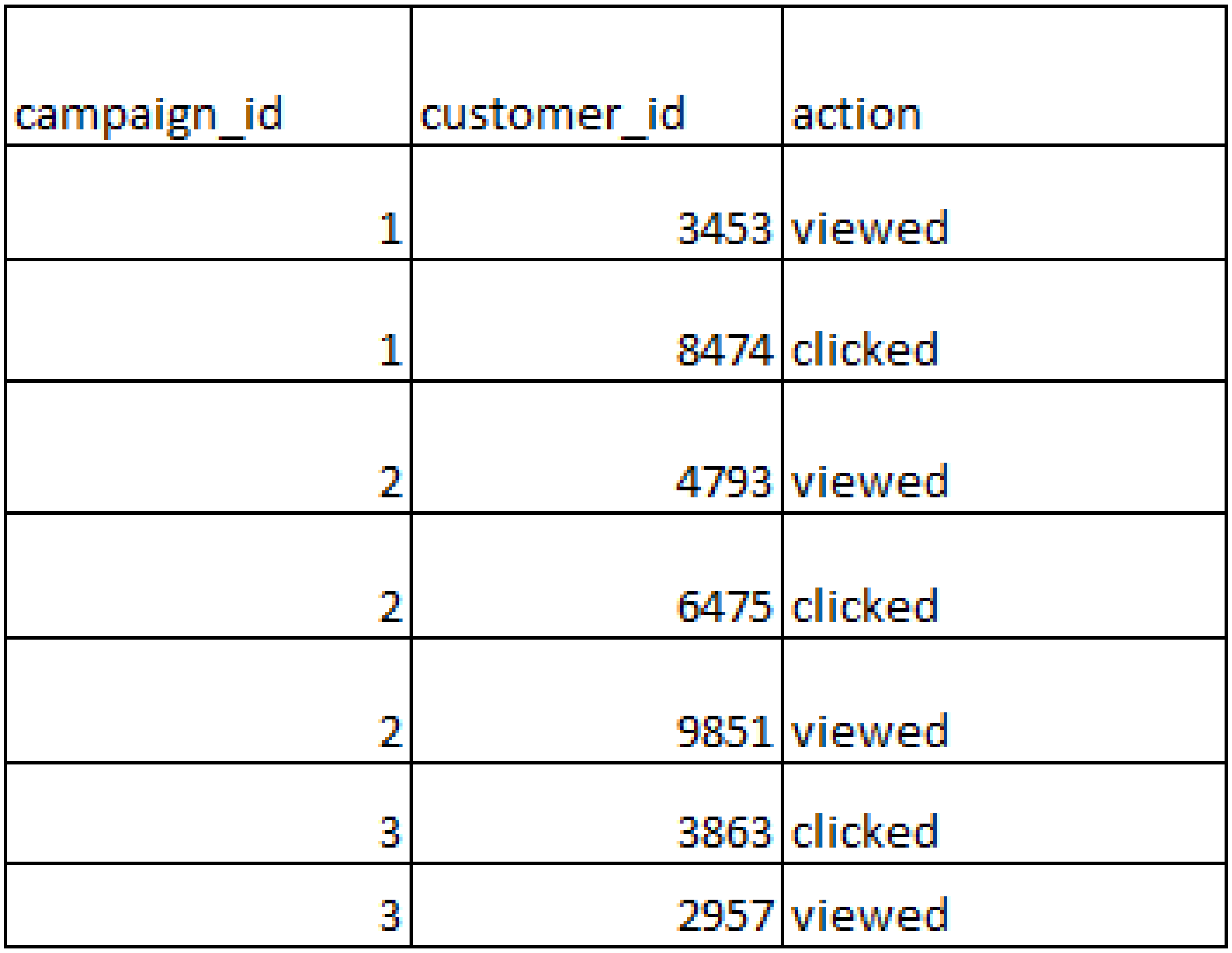
```sql
SELECT
    EXTRACT(MONTH FROM transaction_date) AS month,
    card_holder_id AS card_holder,
    AVG(transaction_cost) AS avg_transaction_cost
FROM
    transactions
GROUP BY
    month,
    card_holder_id;
```

4. Calculating Click-Through-Rate for Marketing Campaigns

Problem Statement: Calculate the click-through rate (CTR) for each marketing campaign. CTR is the ratio of the number of clicks to the number of views, expressed as a percentage. You have two tables: campaigns and clicks.



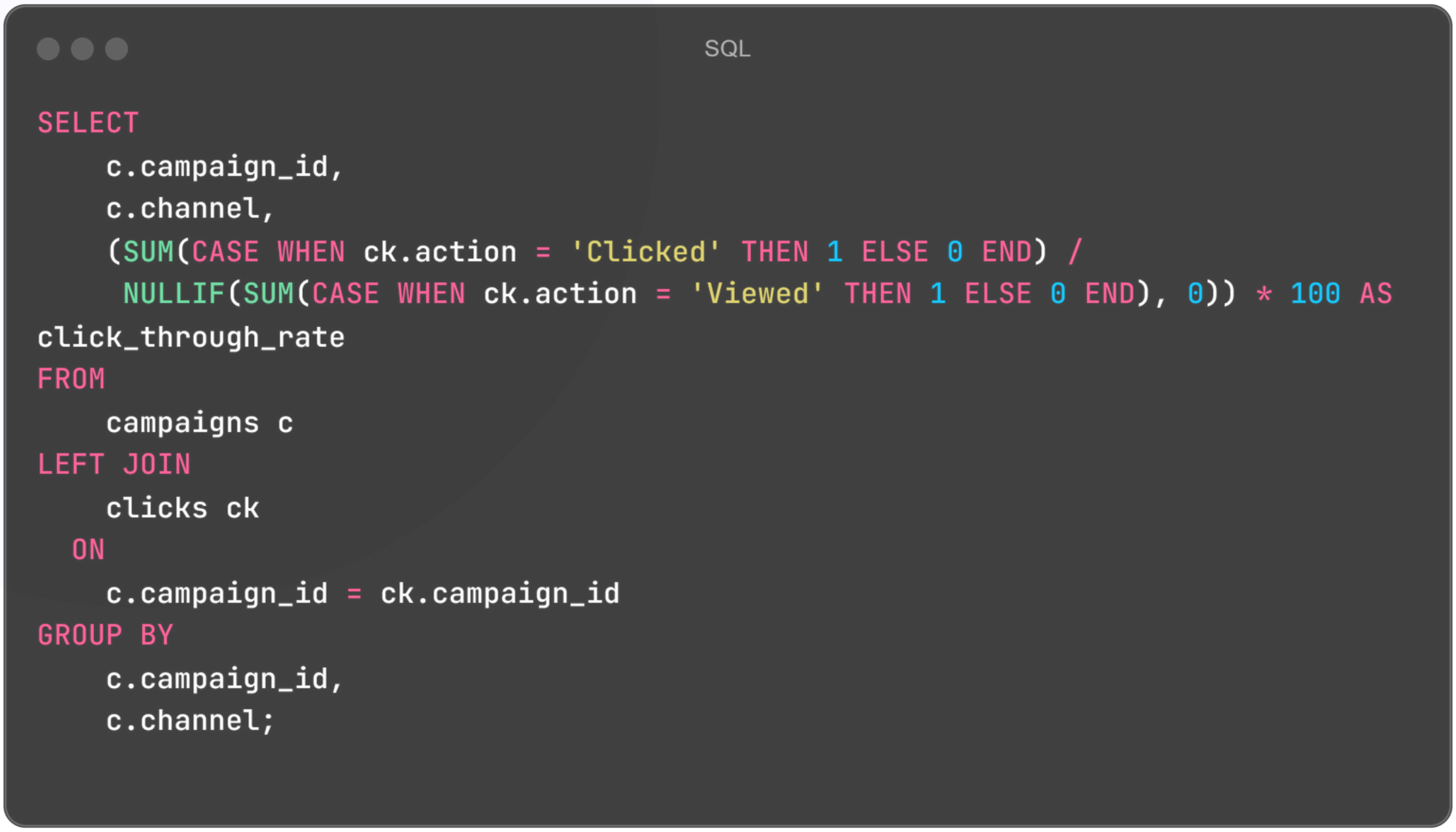
campaign_id	channel	date
1	Email	01/04/2024
2	Web	01/06/2024
3	App	02/10/2024



campaign_id	customer_id	action
1	3453	viewed
1	8474	clicked
2	4793	viewed
2	6475	clicked
2	9851	viewed
3	3863	clicked
3	2957	viewed

How to Solve:

- Join the *campaigns* and *clicks* tables on *campaign_id*.
- Count 'Clicked' and 'Viewed' actions for each campaign.
- Calculate CTR as (Clicked / Viewed) * 100.



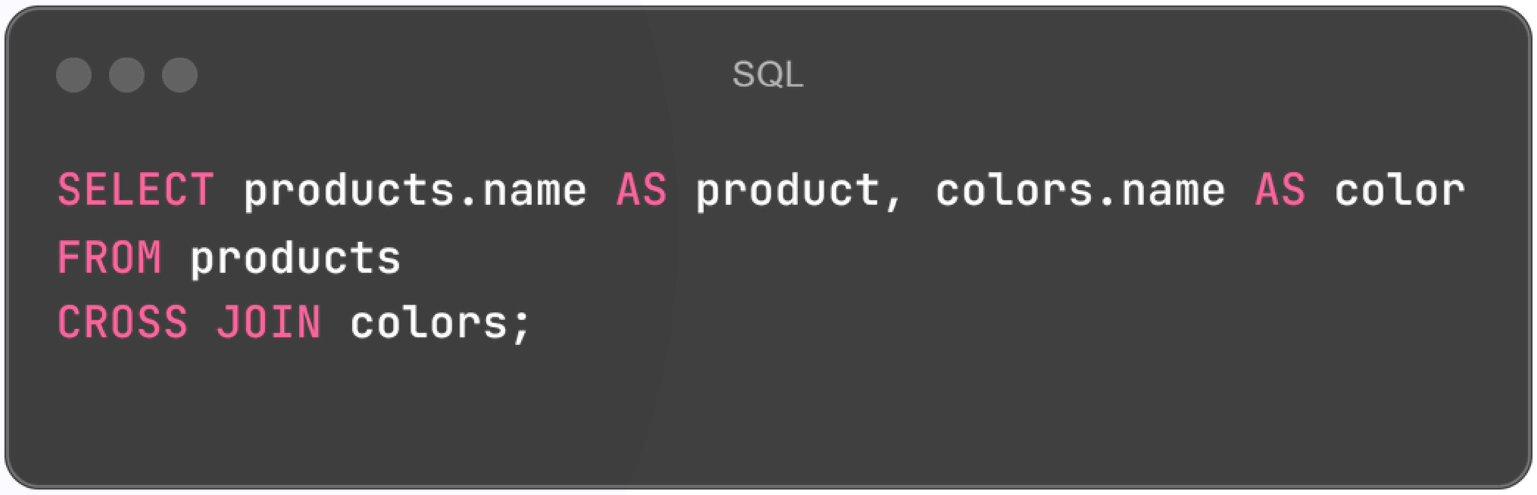
```sql
SELECT
    c.campaign_id,
    c.channel,
    (SUM(CASE WHEN ck.action = 'Clicked' THEN 1 ELSE 0 END) /
     NULLIF(SUM(CASE WHEN ck.action = 'Viewed' THEN 1 ELSE 0 END), 0)) * 100 AS
click_through_rate
FROM
    campaigns c
LEFT JOIN
    clicks ck
  ON
    c.campaign_id = ck.campaign_id
GROUP BY
    c.campaign_id,
    c.channel;
```

5. Distinction Between Cross Join and Natural Join

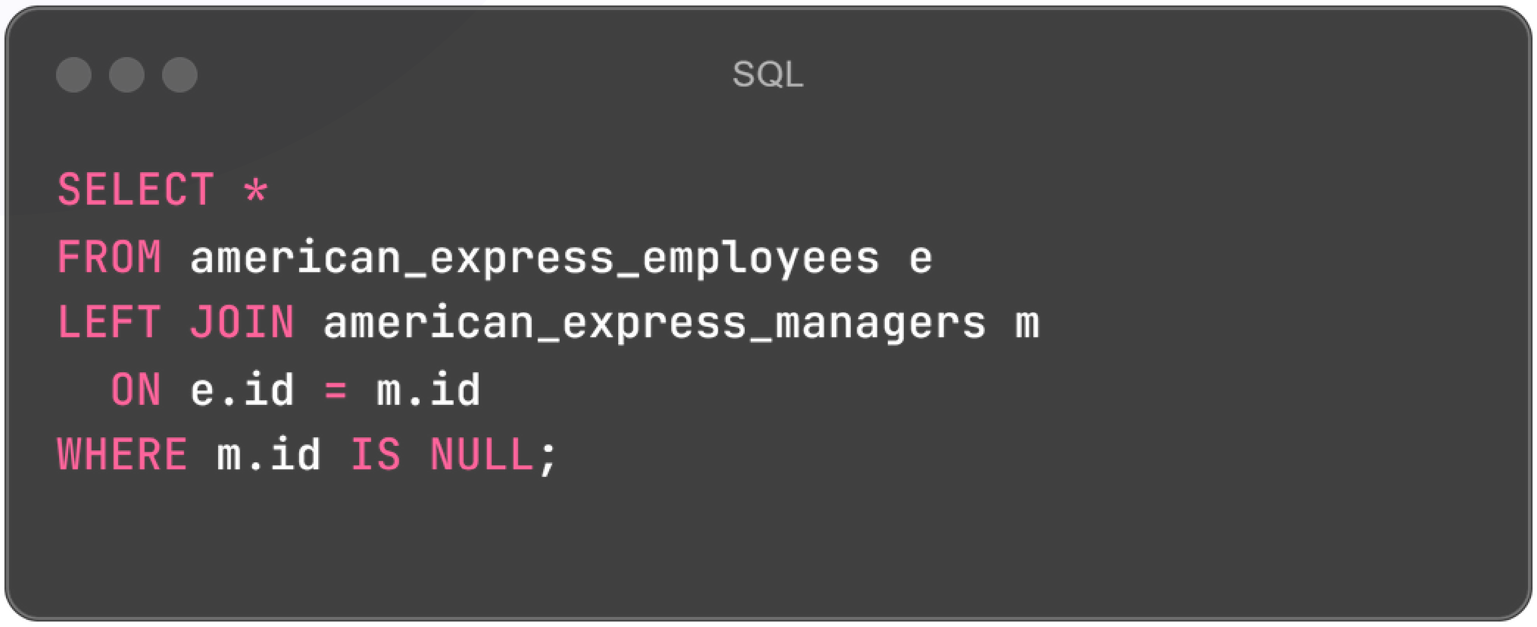
Problem Statement: Explain the difference between a cross join and a natural join. Provide examples for both types of joins.

How to Solve:

- Define cross join (Cartesian product) and natural join (based on common columns).
- Provide example queries for each type of join to illustrate the differences.



```sql
SELECT products.name AS product, colors.name AS color
FROM products
CROSS JOIN colors;
```



```sql
SELECT *
FROM american_express_employees e
LEFT JOIN american_express_managers m
  ON e.id = m.id
WHERE m.id IS NULL;
```

Found this helpful? Repost!

linkedin.com/in/ileonjose