Given a table of candidates and their skills, you're tasked with finding the candidates best suited for an open Data Science job. You want to find candidates who are proficient in Python, Tableau, and PostgreSQL.

Write a query to list the candidates who possess all of the required skills for the job. Sort the output by candidate ID in ascending order.

`candidates` **Table:**

Column Name	Type
candidate_id	integer
skill	varchar

`candidates` **Example Input:**

candidate_id	skill
123	Python
123	Tableau
123	PostgreSQL
234	R
234	PowerBI
234	SQL Server
345	Python
345	Tableau

Solution

```sql
SELECT
  candidate_id
FROM
  candidates
WHERE
  skill IN ('Python', 'Tableau', 'PostgreSQL')
GROUP BY
  candidate_id
HAVING
  COUNT(DISTINCT skill) = 3
ORDER BY
  candidate_id ASC;
```

Assume you are given the table below on Uber transactions made by users. Write a query to obtain the third transaction of every user. Output the user id, spend and transaction date.

`transactions` Table:

Column Name	Type
user_id	integer
spend	decimal
transaction_date	timestamp

`transactions` Example Input:

user_id	spend	transaction_date
111	100.50	01/08/2022 12:00:00
111	55.00	01/10/2022 12:00:00
121	36.00	01/18/2022 12:00:00
145	24.99	01/26/2022 12:00:00
111	89.60	02/05/2022 12:00:00

Solution:

```sql
WITH ranked_transactions AS (
    SELECT user_id, spend, transaction_date,
           ROW_NUMBER() OVER (PARTITION BY user_id ORDER BY transaction_date) AS transaction_rank
    FROM transactions
)
SELECT user_id, spend, transaction_date
FROM ranked_transactions
WHERE transaction_rank = 3;
```

CVS Health is trying to better understand its pharmacy sales, and how well different products are selling. Each drug can only be produced by one manufacturer.

Write a query to find the total sales of drugs for each manufacturer. Round your answer to the closest million, and report your results in descending order of total sales.

Because this data is being directly fed into a dashboard which is being seen by business stakeholders, format your result like this: "$36 million".

`pharmacy_sales` **Table:**

Column Name	Type
product_id	integer
units_sold	integer
total_sales	decimal
cogs	decimal
manufacturer	varchar
drug	varchar

`pharmacy_sales` **Example Input:**

product_id	units_sold	total_sales	cogs	manufacturer	drug
94	132362	2041758.41	1373721.70	Biogen	UP and UP
9	37410	293452.54	208876.01	Eli Lilly	Zyprexa
50	90484	2521023.73	2742445.9	Eli Lilly	Dermasorb
61	77023	500101.61	419174.97	Biogen	Varicose Relief
136	144814	1084258.00	1006447.73	Biogen	Burkhart

Solution

```
SELECT
  manufacturer,
  CONCAT ('$', ROUND(SUM(total_sales)/ 1000000), ' million') AS Total_Sales
FROM pharmacy_sales
GROUP BY manufacturer
```

Assume you are given the table below that shows job postings for all companies on the LinkedIn platform. Write a query to get the number of companies that have posted duplicate job listings.

`job_listings` Table:

Column Name	Type
job_id	integer
company_id	integer
title	string
description	string

`job_listings` Example Input:

job_id	company_id	title	description
248	827	Business Analyst	Business analyst evaluates past and current business data with the primary goal of improving decision-making processes within organizations.
149	845	Business Analyst	Business analyst evaluates past and current business data with the primary goal of improving decision-making processes within organizations.
945	345	Data Analyst	Data analyst reviews data to identify key insights into a business's customers and ways the data can be used to solve problems.

Solution

```sql
WITH duplicate_jobs AS (
    SELECT company_id, title, description,
           COUNT(*) OVER (PARTITION BY company_id, title, description) AS listing_count
    FROM job_listings
)
SELECT COUNT(DISTINCT company_id)
FROM duplicate_jobs
WHERE listing_count > 1;
```

Given the reviews table, write a query to get the average stars for each product every month.

The output should include the month in numerical value, product id, and average star rating rounded to two decimal places. Sort the output based on month followed by the product id.

`reviews` Table:

Column Name	Type
review_id	integer
user_id	integer
submit_date	datetime
product_id	integer
stars	integer (1-5)

`reviews` Example Input:

review_id	user_id	submit_date	product_id	stars
6171	123	06/08/2022 00:00:00	50001	4
7802	265	06/10/2022 00:00:00	69852	4
5293	362	06/18/2022 00:00:00	50001	3
6352	192	07/26/2022 00:00:00	69852	3
4517	981	07/05/2022 00:00:00	69852	2

Solution

```sql
SELECT
  EXTRACT(MONTH FROM submit_date) AS mnth,
  product_id,
  ROUND(AVG(stars), 2) AS avg_stars
FROM reviews
GROUP BY EXTRACT(MONTH FROM submit_date), product_id
ORDER BY mnth, product_id;
```