Practical Exam Sample: Pet Supplies

PetMind is a retailer of products for pets. They are based in the United States.

PetMind sells products that are a mix of luxury items and everyday items. Luxury items include toys. Everyday items include food.

The company wants to increase sales by selling more products for some animals repeatedly.

They have been testing this approach for the last year.

They now want a report on how repeat purchases impact sales.

Data

The data is available in the table `pet_supplies`.

The dataset contains the sales records in the stores last year.

Column Name	Criteria
product_id	Nominal. The unique identifier of the product. Missing values are not possible due to the database structure.
category	Nominal. The category of the product, one of 6 values (Housing, Food, Toys, Equipment, Medicine, Accessory). Missing values should be replaced with "Unknown".
animal	Nominal. The type of animal the product is for. One of Dog, Cat, Fish, Bird. Missing values should be replaced with "Unknown".
size	Ordinal. The size of animal the product is for. Small, Medium, Large. Missing values should be replaced with "Unknown".
price	Continuous. The price the product is sold at. Can be any positive value, round to 2 decimal places. Missing values should be replaced with the overall median price.
sales	Continuous. The value of all sales of the product in the last year. This can be any positive value, rounded to 2 decimal places. Missing values should be replaced with the overall median sales.
rating	Discrete. Customer rating of the product from 1 to 10. Missing values should be replaced with 0.
repeat_purchase	Nominal. Whether customers repeatedly buy the product (1) or not (0). Missing values should be removed.

Task 1

From taking a quick look at the data, you are pretty certain it isn't quite as it should be. You need to make sure all of the data is clean before you start your analysis. The table below shows what the data should look like.

Write a query to return a table that matches the description provided.

Do not update the original table.

Column Name	Criteria
product_id	Nominal. The unique identifier of the product. Missing values are not possible due to the database structure.
category	Nominal. The category of the product, one of 6 values (Housing, Food, Toys, Equipment, Medicine, Accessory). Missing values should be replaced with "Unknown".
animal	Nominal. The type of animal the product is for. One of Dog, Cat, Fish, Bird. Missing values should be replaced with "Unknown".
size	Ordinal. The size of animal the product is for. Small, Medium, Large. Missing values should be replaced with "Unknown".
price	Continuous. The price the product is sold at. Can be any positive value, round to 2 decimal places. Missing values should be replaced with 0.
sales	Continuous. The value of all sales of the product in the last year. This can be any positive value, rounded to 2 decimal places. Missing values should be replaced with the overall median sales.
rating	Discrete. Customer rating of the product from 1 to 10. Missing values should be replaced with 0.
repeat_purchase	Nominal. Whether customers repeatedly buy the product (1) or not (0). Missing values should be removed.

Certification Pet Supplies DB DataFrame as `clean_data`

```
SELECT product_id,
    CASE WHEN category = '-' THEN 'Unknown' ELSE category END AS category,
    animal,
    CASE WHEN price = 'unlisted' then 0 else price::numeric end as price,
    sales,
    CASE WHEN rating IS NULL THEN 0 ELSE rating END AS rating,
    INITCAP(size) as size
FROM pet_supplies
```

	product_id	category	animal	price	sales
0	1	Food	Bird	51.1	
1	2	Housing	Bird	35.98	
2	3	Food	Dog	31.23	
3	4	Medicine	Cat	24.95	
4	5	Housing	Cat	26.18	
5	6	Housing	Dog	30.77	
6	7	Housing	Dog	31.04	
7	8	Toys	Cat	28.9	
8	9	Equipment	Fish	17.82	
9	10	Medicine	Dog	24.93	
10	11	Food	Dog	40.87	
11	12	Medicine	Bird	34.96	
12	13	Food	Dog	31.07	
13	14	Food	Dog	40.8	
14	15	Accessory	Bird	33.13	

1,500 rows

Task 2

You want to show whether sales are higher for repeat purchases for different animals. You also want to give a range for the sales.

Write a query to return the `animal`, `repeat_purchase` indicator and the `avg_sales`, along with the `min_sales` and `max_sales`. All values should be rounded to whole numbers.

You should use the original `pet_supplies` data for this task.

Certification Pet Supplies DB DataFrame as animal_sales

```
SELECT animal,
    repeat_purchase,
    ROUND(AVG(sales)) as avg_sales,
    ROUND(MIN(sales)) as min_sales,
    ROUND(MAX(sales)) as max_sales
FROM pet_supplies
GROUP BY animal, repeat_purchase
ORDER BY avg_sales DESC;
```

	animal	repeat_purchase	avg_sales	min_sales
0	Bird	1	1408	
1	Bird	0	1380	
2	Dog	0	1084	
3	Dog	1	1038	
4	Cat	0	1035	
5	Cat	1	998	
6	Fish	0	705	
7	Fish	1	693	

8 rows

Task 3

The management team want to focus on efforts in the next year on the most popular pets - cats and dogs - for products that are bought repeatedly.

Write a query to return the `product_id`, `sales` and `rating` for the relevant products.

You should use the original `pet_supplies` data for this task.

Certification Pet Supplies DB DataFrame as p

```
SELECT product_id, sales::numeric, rating
FROM public.pet_supplies
WHERE animal IN ('Dog', 'Cat')
    AND repeat_purchase = 1
ORDER BY sales::numeric DESC;
```

	product_id	sales	
0	518	179	
1	280	179	
2	728	179	
3	20	179	
4	946	178	
5	863	172	
6	1383	172	
7	272	147	
8	561	146	
9	285	146	
10	752	146	
11	332	14	
12	152	14	
13	1061	146	
14	88	146	

552 rows