San José State University
Department of Applied Data Science

DATA 225
Database Systems for Analytics

Fall 2023
Section 21
Instructor: Ron Mak

Assignment #5

Assigned:	Monday, September 25
Due:	Monday, October 2 at 5:30 pm
	Individual assignment, 100 points max

SQL practice #2

This is a continuation of individual Assignment #3b. It uses the ZAGI Retail Company Sales Department database that you created for that previous assignment. Problems #7 and #8 are a bit more challenging than the others.

In a Jupyter notebook, connect to the **`zagi_sales`** database and write the Python-embedded SQL code to solve each of the following problems. Display each result in a dataframe.

1. [10 points] Display the region ID, region name, and number of stores in the region for all regions.
2. [10 points] For each product category, display the category ID, category name, and average price of a product in the category.
3. [10 points] For each product category, display the category ID and the total number of items purchased in the category.
4. [10 points] Display the TID and the total number of items (of all products) sold within the transaction for all sales transactions whose total number of items (of all products) sold within the transaction is greater than five.
5. [10 points] Display the product ID and ProductName of the cheapest product.
6. [10 points] Display the product ID for the product that has been sold the most (i.e., that has been sold in the highest quantity).

7. [20 points] Rewrite the following query using a join instead of the nested query. Your rewritten query should produce the same results.

For each product that has more than three items sold within all sales transactions, retrieve the product ID, product name, and product price.

```
SELECT product_id, product_name, product_price
FROM  product
WHERE product_id IN (SELECT product_id
                     FROM sold_via
                     GROUP BY product_id
                     HAVING SUM(no_of_items) > 3)
```

8. [20 points] Rewrite the following query using a join instead of the nested query. Your rewritten query should produce the same results.

For each product whose items were sold in more than one sales transaction, retrieve the product id, product name and product price.

```
SELECT product_id, product_name, product_price
FROM product
WHERE product_id IN (SELECT product_id
                     FROM sold_via
                     GROUP BY product_id
                     HAVING COUNT(*) > 1)
```

What to submit

Submit your notebook including all the output cells.

TIP: Before submitting, do **Run ➔ Restart Kernel and Run All Cells** and then save**.**

Rubric

Criteria	Max points
• Problem 1	• 10
• Problem 2	• 10
• Problem 3	• 10
• Problem 4	• 10
• Problem 5	• 10
• Problem 6	• 10
• Problem 7	• 20
• Problem 8	• 20