Amrita Vishwa Vidyapeetham

Department of Artificial Intelligence

21AIE304 Big Data and DataBase Management Systems

Practice Problems

Date: 24-08-2023

Instructions

a. Please compile the practice problems into a Word document and save it in your OneDrive folder.
b. Create a section titled 'Practice Problem 1.' For each question, provide the associated code in text format immediately below the question.
c. Kindly refrain from including any screenshots.

A) Refer to product.csv and perform the following

1. Count the total number of transactions in the dataset.
2. Calculate the total revenue generated from all transactions.
3. Find the product with the highest price.
4. List the distinct products that were sold in the "Electronics" category.
5. Calculate the average price of products in each category.
6. Calculate the total quantity sold for each product.
7. Calculate the revenue generated for each month.
8. List the top 5 customers who spent the most.
9. Find the products purchased by a specific customer.
10. List the quantity of products sold in each category.
11. Calculate the average quantity of products sold per transaction.
12. List products that were sold at least twice along with their total quantity sold.
13. Find customers who made purchases in both the "Electronics" and "Clothing" categories.
14. List products sold on a specific transaction date along with their quantity sold.
15. List the top N categories by total revenue generated.
16. Find customers who have made more than one transaction along with the number of transactions.
17. List products sold between a specific date range along with the quantity sold.
18. Calculate the total revenue generated by each customer.
19. Calculate the percentage of revenue contributed by each category to the total revenue.
20. Calculate the total quantity sold and total revenue for products in the low, medium, and high price ranges.
21. Count the number of products in each category from the products table.

22. Calculate the total sales for each product category from the sales table.
23. Retrieve products from the products table whose names contain the word 'Laptop'.
24. Retrieve orders from the orders table placed between January 1, 2022, and December 31, 2022.

B) Refer to employee2.csv and do the following

1. Calculate the average salary of employees in each department with a salary greater than $40,000 from the employees table.
2. Retrieve employees from the employees table who have a salary greater than the average salary.
3. Retrieve employees from the employees table with a salary greater than $60,000 and job title is 'Manager'.
4. Retrieve employees from the employees table whose last name starts with 'S'.
5. Update the salary of an employee with employee_id 101 to $55,000 in the employees table.
6. Retrieve employees from the employees table who are in the 'Sales' department and have a salary between $40,000 and $50,000.
7. Retrieve employees from the employees table who have not been assigned to any department (department is NULL).
8. Retrieve employees' full names and a calculated column for their annual bonus (10% of salary) from the employees table.
9. Retrieve distinct job titles from the employees table.
10. Retrieve employees from the employees table who are in the 'Sales' or 'Marketing' departments.
11. Retrieve employees from the employees table who are in the 'Sales' and 'Marketing' departments.
12. Retrieve employees from the employees table who are not in the 'Sales' department.
13. Retrieve employees from the employees table, ordered first by department in ascending order, and then by salary in descending order.
14. Retrieve employees' full names and a column indicating whether their salary is above $60,000 in the employees table.

15. You want to return values in multiple columns as one column. For example, you would like to produce this result set from a query against the EMP table:

CLARK WORKS IN IT

KING WORKS in HR

16. Get the random records from but limit the size to 5

17. Return employee names and departments from the table mployee and sort by the last two characters in the name field.
18. Display the Full Name of the employee whose salary is maximum.
19. Select an attribute of your choice and make a constraint in table if a value is missing by default, the attribute should be filled with a default value and not NULL.
20. Get the products with invalid product_id

Hint:

DATE datatype could be used for date

REGEXP for regular expressions

LIKE for substring matching

Also refer to SUBSTRING