BA 데이터애널리틱스 [BA-DA]
Fall 2020
Homework Assignment #2: Querying data from a relational database

Suggested Answers

Scenario

Increasing sales is the most fundamental problem that every business organization faces throughout its life, and it needs to address the problem properly to survive. In a preliminary effort to deal with this problem, you start investigating popular products and active customers, and have the following tables prepared from the sample database used in class. Given this simple scenario and the tables provided below, answer the following questions.

Table 1: Top 10 Products by Quantity of Product Ordered

productName	productCode	productVendor	SumOfQuantityOrdered
1992 Ferrari 360 Spider red	S18_3232	Unimax Art Galleries	1808
1937 Lincoln Berline	S18_1342	Motor City Art Classics	1111
American Airlines: MD-11S	S700_4002	Second Gear Diecast	1085
1941 Chevrolet Special Deluxe Cabriolet	S18_3856	Exoto Designs	1076
1930 Buick Marquette Phaeton	S50_1341	Studio M Art Models	1074
1940s Ford truck	S18_4600	Motor City Art Classics	1061
1969 Harley Davidson Ultimate Chopper	S10_1678	Min Lin Diecast	1057
1957 Chevy Pickup	S12_4473	Exoto Designs	1056
1964 Mercedes Tour Bus	S18_2319	Unimax Art Galleries	1053
1956 Porsche 356A Coupe	S24_3856	Classic Metal Creations	1052

Note: SumOfQuantityOrdered is the total quantity of each product ordered.

Table 2: Top 10 Customers by Total Amount of All Orders

customerNumber	customerName	city	country	Amount
141	Euro+ Shopping Channel	Madrid	Spain	820689.5
124	Mini Gifts Distributors Ltd.	San Rafael	USA	591827.3
114	Australian Collectors, Co.	Melbourne	Australia	180585.1
151	Muscle Machine Inc	NYC	USA	177914
119	La Rochelle Gifts	Nantes	France	158573.1
148	Dragon Souveniers, Ltd.	Singapore	Singapore	156251
323	Down Under Souveniers, Inc	Auckland	New Zealand	154622.1
131	Land of Toys Inc.	NYC	USA	149085.2
187	AV Stores, Co.	Manchester	UK	148410.1
450	The Sharp Gifts Warehouse	San Jose	USA	143536.3

Note: Amount is the total dollar amount of all orders that each customer has made.

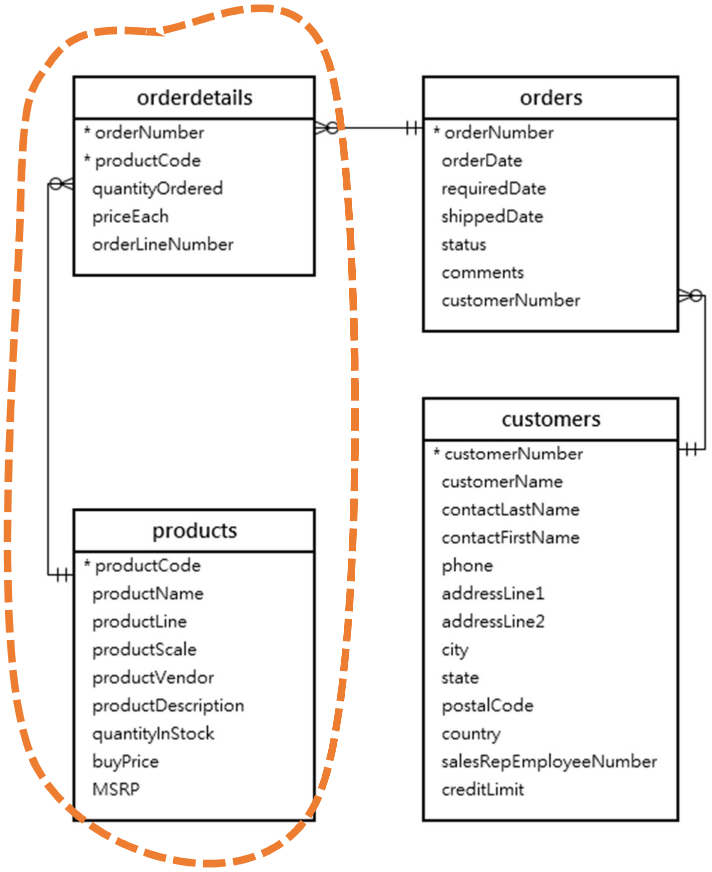
1. Prepare sql statements to produce the two tables, respectively.

-- HW#2 Table 1: Top 10 Products by Quantity of Product Ordered (tables in the orange box)

```
SELECT productName, p.productCode, productVendor, SUM(quantityOrdered) as
SumOfQuantityOrdered

FROM products p

INNER JOIN orderdetails od ON p.productCode = od.productCode

GROUP BY p.productCode

ORDER BY SumOfQuantityOrdered desc

LIMIT 10;
```

Explanations for the select statement:

1) [FROM clause] To retrieve the columns in the resulting table (Table 1) from the database, we need to join two tables, 'products' and 'orderdetails' because the relevant data exists in the two tables as follows:
   - Two columns, productName and productVendor, are in the 'products' table.
   - The column needed to calculate the sum of quantity ordered for each product, quantityOrdered, is in the 'orderdetails' table.
   - productCode is the common value column (PK in the 'products' table and FK in 'orderdatails' table), which is used to join the two tables.



2) [GROUP BY clause] Each product is highly likely to be ordered by several orders, and therefore, there will be multiple rows for each product. Thus, to calculate the sum of quantity ordered for each product, we need to group the rows in the inner join of the two tables by productCode, which is unique to each product

3) [SUM function in SELECT clause] Now, sum up the quantityOrdered for each group of products.

4) [ORDER BY & LIMIT] To present only the top 10 most ordered products, we order the rows by the sum of quantity order in the descending order, and present only the first 10 rows in the resulting table.

-- HW#2 Table2: Top 10 Customers by Total Amount of All Orders (tables in the red box)

SELECT c.customerNumber, customerName, city, country, SUM(quantityOrdered * priceEach) AS Amount

FROM customers c

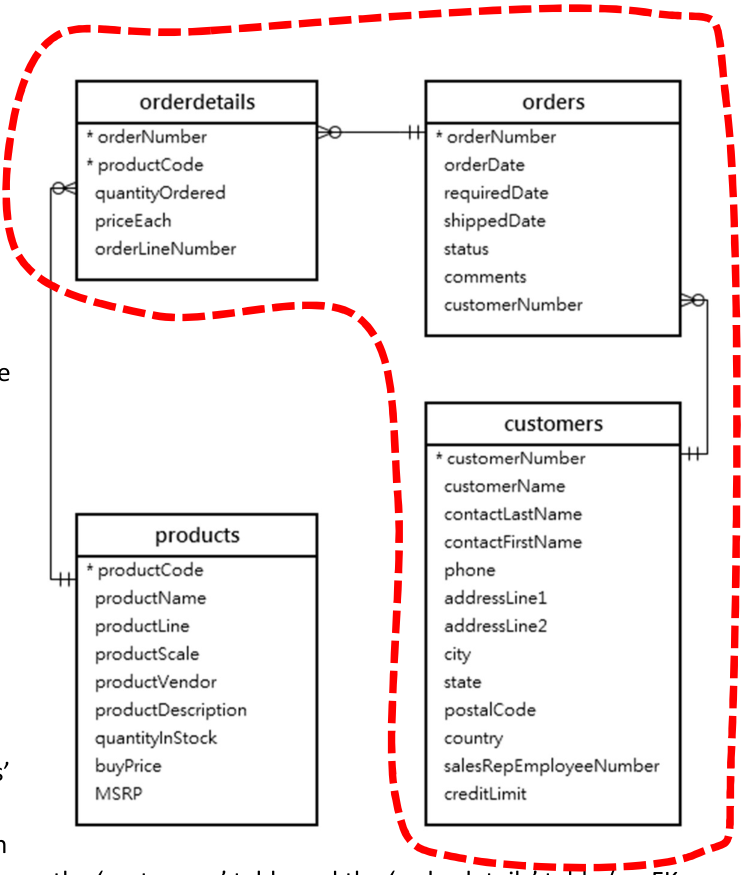
INNER JOIN orders o ON c.customerNumber = o.customerNumber

INNER JOIN orderdetails od ON o.orderNumber = od.orderNumber

GROUP BY c.customerNumber

ORDER BY Amount desc

LIMIT 10;



Explanations for the select statement:

1) [FROM clause] To retrieve the columns in the resulting table (Table 2) from the database, we need to join three tables, 'customers,' 'orders' and 'orderdetails' because the relevant data exists in the three tables as follows:
   - Four columns, customerName, city, state and country, are in the 'customers' table.
   - Two columns needed to calculate the total amount of all orders made by each customer, namely quantityOrdered and priceEach, are in the 'orderdetails' table.
   - customerNumber is the common value column between 'customers' table and the 'orders' table, and orderNumber is the common value column between the 'orders' table and the 'orderdetails' table. We need to include the 'orders' table in the inner join because there is no direct relationship between the 'customers' table and the 'orderdetails' table (no FK linking the two tables). The 'customers' table and the 'orderdetails' table can be joined together through the 'orders' table in between.

2) [GROUP BY clause] Each customer is highly likely to make multiple orders and each order may include multiple products, and therefore, there will be multiple rows for each customer. Thus, to calculate the total amount of all orders made by each customer, we need to group the rows in the inner join of the three tables by customerNumber, which is unique to each customer.

3) [SUM function in SELECT clause] Now, to calculate the total amount of all orders made by each customer, we first calculate the amount of ordered of each product by multiplying quantityOrdered by priceEach, and then, sum up the amount of ordered or each product for each group of customers.

4) [ORDER BY & LIMIT] To present only the top 10 customers who have made most amount of orders, we order the rows by the sum of the amount of all orders in the descending order, and present only the first 10 rows in the resulting table.

2. **What do you learn about your products and customers from the tables? Also, briefly explain your thoughts on limitations of the tables, and what you would like to retrieve more from the database to better understand business, including products, customers, etc. and to improve your sales**.

The two resulting tables produced by the two select statements provide some useful information about popular products and active customers. We have found out which products are the top 10 most popular products in terms of the number of quantities of products ordered and which customers are the top 10 most active customers in terms of the total amount of sales made by each customers.

Creating the two tables is certainly a good start for better understanding of products and customers, having some implications for plans to increase sales and profits. However, the information from the two tables are quite preliminary and limited. We need more granular understanding and insights about products and customers to come up with plans to promote product sales and eventually increase profits.

For example, regarding the products, we may want to learn about the profitability of each product. Products may vary in their profit margins and promoting sales of products with larger margins will contribute more to increasing profits. We may also want to examine whether some products are more popular across different regions, helping to enhance customized product portfolio in different regions.

Regarding customers, we may would like to know which customers are generating more profits, not just simply making higher sales amount. We may want to run reward programs to promote even more sales from profitable customers. Also, we may want to know which sales representatives are associated with active customers, so that we can reward employee performance, enhancing employee's motivation to makes more sales.

These are just a few example scenarios for getting more understanding and insights about products, customers and even employees. You may be able to come up with other ideas based on your own interests in different business disciplines.