Sixth hands-on exercise (questions)

Task 1: In the car retailer database, what is the name of the customer who generated the most profit for the company? ***Assume*** all the orders were accepted and shipped with no cancellation!

- In the table orderdetails of the car retailer database, we can find the information of productCode, the *selling* price [priceEach] as well as the quantity of the product sold [quantityOrdered].
- In the table products, we can find the information of productCode, the price [buyPrice] that the company used to *purchase* the product.
- Therefore, by comparing the purchasing price and selling price as well as considering the quantity of the sold products, you will be able to find the profit of each sales order.
- After above calculations, we can connect the data of profit generated by each sales order to the IDs of the customers (customerNumber) who made the purchases.
- By summing the profits provided by each customer (customerNumber), we can find the most profitable customer.

Please don't check the answer unless you have tried to solve the question by yourself for some time (e.g. 30 minutes) but failed, because after seeing the MySQL code of answers, you will find the question is actually easy to solve.

Even after you read the answer, please write the code by yourself again without 'copying' the code.

If you are able to solve the question by yourself without checking the answer, you are almost good enough to 'graduate' from the course!

- Think about how to do the task 1, if the cancellation of purchase needs to be taken into account? (Answers are not provided.)

Task 2: Based on two tables below, please test different join functions (e.g. left join, inner join and right join) on the condition of column of A_ID = B_ID. I am pretty sure that you are good enough now to develop the tables below by yourself.

A_ID	AWord
1	A_1
2	A_2
3	A_3
4	A_4
5	A_5
6	A_6

B_ID	BWord
1	B_1
2	B_2
3	B_3
7	B_7
8	B_8
9	B_9

- Check the course slide for the answers

Task 3: In TripAdvisor dataset, which location (locality) receives the largest amount of hotel reviews and which one receives the lowest amount of reviews?

- the calculation needs to be based on two tables [tripadvisor_review_sample_without_reviewtext.sql] and [tripadivsor_hotel_sample.sql]
- in [tripadvisor_review_sample_without_reviewtext] table, you can find the reviews for each hotel.
- in [tripadivsor_hotel_sample] table, you can find the locality for each hotel

Task 4: In TripAdvisor dataset, let's say that the review samples collected from location ('locality' of hotel) with less than 700 reviews are not worth to be included for further analysis, and we want to drop them. So please create a new table [name: new_data] that has a structure similar to "tripadvisor_review_sample_without_reviewtext" table, but only include the reviews from the locations ['locality'] that received over 700 reviews.

Task 5: Based on the table "tripadvisor_review_sample_without_reviewtext", please create a new table [name: new_data2] including 1000 rows of reviews for hotels in "New York city" and 500 rows of reviews for hotels in "Chicago".

Advance tasks [The below tasks are beyond the requirement for a MySQL beginner. However, I would like to provide some MySQL examples of how it could be used, so that you can think about it. The exam questions will not be that difficult.]

Task 6: In the "tripadvisor_data_for_handson_assignment_ONLY" dataset, please calculate the **number** of the reviews that hotels received, and only consider those hotels that received reviews having at least one review of an overall rating of 5, 3 and 1, respectively. If all the reviews given to a hotel XXX ONLY having an overall rating of either 4 or 5, it will be excluded, because it does not have any reviews offering an overall rating of 3 or 1. Please obtain following result in one command.

tripadvisor_data_for_handson_assignment_ONLY (4×20)

hotel_id	Num_of_5_star_rating	Num_of_3_star_rating	Num_of_1_star_rating
75,737	4	1	1
81,192	8	5	1
93,437	6	1	1
93,450	4	4	2
93,517	2	1	2
93,520	11	4	1

Task 7: Based on the task 6, think about how to do the task, if a hotel that did not receive any rating of 5, 3 or 1 will still be included in the result. By learning the ifnull (see. https://www.jquery-az.com/mysql-ifnull/) function by yourself, you will figure out the solution.

Task 8: In the "tripadvisor_data_for_handson_assignment_ONLY" dataset, please calculate the **proportions** of reviews giving an overall rating of 5 and 3 and 1 respectively, and only consider those hotels that received reviews having at least one review of an overall rating of 5, 3 and 1, respectively. Excluding those hotels with less than 10 reviews from the analysis. Please obtain following result via one command.

tripadvisor_data_for_handson_assignment_ONLY (5×17)

hotel_id	Num_of_total_review	ratio_of_5_star_rating	ratio_of_3_star_rating	ratio_of_1_star_rating
81,192	20	0.4000	0.2500	0.0500
93,437	15	0.4000	0.0667	0.0667
93,450	18	0.2222	0.2222	0.1111
93,517	11	0.1818	0.0909	0.1818
93,520	20	0.5500	0.2000	0.0500
93,545	13	0.0769	0.2308	0.0769

Task 9: Please finish all the tasks from exercise 4 and 5 again by yourself **without looking at answers and in a very short period of time**!