**Sixth hands-on exercise (answers)**

**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!

**alter** **table** orderdetails **add** **column** customerNumber **int**;

**update** orderdetails **join** orders **on** orders.orderNumber = orderdetails.orderNumber **set** orderdetails.customerNumber = orders.customerNumber;

**alter** **table** orderdetails **add** **column** cost **decimal**(10,2);

**UPDATE** orderdetails

**JOIN** products **ON** products.productCode = orderdetails.productCode **SET** orderdetails.cost = products.buyPrice;

**select** customerNumber, **round**(**sum**((priceEach-cost)\*quantityOrdered),2) **as** profit **from** orderdetails **group** **by** customerNumber

**order** **by** profit **desc**;

A simple code to solve the above problem by joining three tables.

**select** tb.customerNumber, c.customerName,c.contactLastName, c.contactFirstName, **sum**((o.priceEach - p.buyPrice)\*o.quantityOrdered) **AS** profit

**from** orderdetails o

**join** orders tb **on** o.orderNumber = tb.orderNumber

**join** products p **on** o.productCode = p.productCode

**JOIN** customers c **ON** tb.customerNumber = c.customerNumber

**group** **by** tb.customerNumber **ORDER** **BY** profit **desc**

From the table customers, you can the name of customerNumber of 141 (Freyre, Diego)

**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

**select** b.locality, **count**(a.id) **as** fre **from** tripadvisor\_review\_sample\_without\_reviewtext a **join** tripadivsor\_hotel\_sample b **on** a.hotel\_id = b.hotel\_id **group** **by** b.locality **order** **by** fre **desc**

**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.

**create** **table** temp **as** (**select** b.locality, **count**(a.id) **as** fre **from** tripadvisor\_review\_sample\_without\_reviewtext a **join** tripadivsor\_hotel\_sample b **on** a.hotel\_id = b.hotel\_id **group** **by** b.locality **having** fre > 200);

**create** **table** new\_data **as** (**select** \* **from** tripadvisor\_review\_sample\_without\_reviewtext **where** hotel\_id **in** (**select** hotel\_id **from** tripadivsor\_hotel\_sample **where** locality **in** (**select** locality **from** temp)));

**drop** **table** temp;

**- What you would do, if it took a long time for your computer to run the query?!**

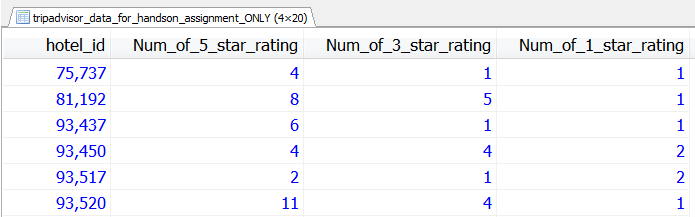
**Task 5:** Based on the table “tripadvisor\_review\_sample\_without\_reviewtext”, please create a new table [name: new\_data2] including 1000 rows of reviews for hotels in “New York city” and 500 rows of reviews for hotels in “Chicago”.

**create** **table** new\_data2 (**select** \* **from** tripadvisor\_review\_sample\_without\_reviewtext **where** hotel\_id **in** (**select** hotel\_id **from** tripadivsor\_hotel\_sample **where** locality = 'New York city') **limit** 1000)

**union**

(**select** \* **from** tripadvisor\_review\_sample\_without\_reviewtext **where** hotel\_id **in** (**select** hotel\_id **from** tripadivsor\_hotel\_sample **where** locality = 'Chicago') **limit** 500)

**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.



**select** tb1.\*, tb2.Num\_of\_3\_star\_rating, tb3.Num\_of\_1\_star\_rating **from** (**select** hotel\_id, **count**(\*) **as** Num\_of\_5\_star\_rating **from** tripadvisor\_data\_for\_handson\_assignment\_ONLY **where** overall\_rating = 5 **group** **by** hotel\_id ) tb1

**join**

(**select** hotel\_id, **count**(\*) **as** Num\_of\_3\_star\_rating **from** tripadvisor\_data\_for\_handson\_assignment\_ONLY **where** overall\_rating = 3 **group** **by** hotel\_id ) tb2

**on** tb1.hotel\_id = tb2.hotel\_id

**join**

(**select** hotel\_id, **count**(\*) **as** Num\_of\_1\_star\_rating **from** tripadvisor\_data\_for\_handson\_assignment\_ONLY **where** overall\_rating = 1 **group** **by** hotel\_id ) tb3

**on** tb3.hotel\_id = tb2.hotel\_id

**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.

**select tb0.hotel\_id, ifnull(tb1.Num\_of\_5\_star\_rating,0), ifnull(tb2.Num\_of\_3\_star\_rating,0), ifnull(tb3.Num\_of\_1\_star\_rating,0) from (select hotel\_id, count(\*) as total\_number from tripadvisor\_data\_for\_handson\_assignment\_ONLY group by hotel\_id ) tb0**

**left join**

**(select hotel\_id, count(\*) as Num\_of\_5\_star\_rating from tripadvisor\_data\_for\_handson\_assignment\_ONLY where overall\_rating = 5 group by hotel\_id ) tb1**

**on tb0.hotel\_id = tb1.hotel\_id**

**left join**

**(select hotel\_id, count(\*) as Num\_of\_3\_star\_rating from tripadvisor\_data\_for\_handson\_assignment\_ONLY where overall\_rating = 3 group by hotel\_id ) tb2**

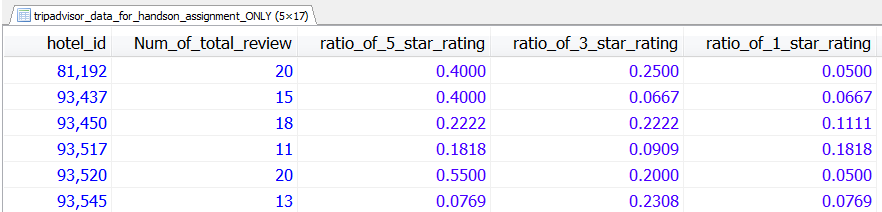
**on tb0.hotel\_id = tb2.hotel\_id**

**left join**

**(select hotel\_id, count(\*) as Num\_of\_1\_star\_rating from tripadvisor\_data\_for\_handson\_assignment\_ONLY where overall\_rating = 1 group by hotel\_id ) tb3**

**on tb0.hotel\_id = tb3.hotel\_id**

**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.



**select** tb1.hotel\_id, tb4.Num\_of\_total\_review,

tb1.Num\_of\_5\_star\_rating/tb4.Num\_of\_total\_review **as** ratio\_of\_5\_star\_rating,

tb2.Num\_of\_3\_star\_rating/tb4.Num\_of\_total\_review **as** ratio\_of\_3\_star\_rating,

tb3.Num\_of\_1\_star\_rating/tb4.Num\_of\_total\_review **as** ratio\_of\_1\_star\_rating **from**

(**select** hotel\_id, **count**(\*) **as** Num\_of\_5\_star\_rating **from** tripadvisor\_data\_for\_handson\_assignment\_ONLY **where** overall\_rating = 5 **group** **by** hotel\_id ) tb1

**join**

(**select** hotel\_id, **count**(\*) **as** Num\_of\_total\_review **from** tripadvisor\_data\_for\_handson\_assignment\_ONLY **group** **by** hotel\_id ) tb4

**on** tb1.hotel\_id = tb4.hotel\_id

**join**

(**select** hotel\_id, **count**(\*) **as** Num\_of\_3\_star\_rating **from** tripadvisor\_data\_for\_handson\_assignment\_ONLY **where** overall\_rating = 3 **group** **by** hotel\_id ) tb2

**on** tb1.hotel\_id = tb2.hotel\_id

**join**

(**select** hotel\_id, **count**(\*) **as** Num\_of\_1\_star\_rating **from** tripadvisor\_data\_for\_handson\_assignment\_ONLY **where** overall\_rating = 1 **group** **by** hotel\_id ) tb3

**on** tb3.hotel\_id = tb2.hotel\_id **where** tb4.Num\_of\_total\_review >= 10