

WEEK 5 CHALLENGE QUESTIONS

In this challenge, practice creating subqueries and cleaning data.

Instructions: Try to answer these questions first and then compare your answers with the provided solutions. You may also submit your answers to avacas@datascienceupskill.com to get personal feedback from me. Good luck!

1. What is the average amount spent by a subset of customers *who paid more than the average amount spent by all the customers in the database?*

Hint: Divide this query into three parts:

Step 1: Compute the average amount paid by all customers.

Step 2: Select or show all the amount spent by customers who paid more than the average amount spent by all customers (computed in step 1)

Step 3: Compute the average amount spent by a subset of customers obtained from step 2.

Note: The average amount should be computed for all payment transactions regardless of the number of transactions or amount paid per customer.

2. Show the film's title, its description and another attribute that you need to create called **string_film**. Under this attribute, if the title starts with 'A', extract or show the first word in the title. Otherwise, write 'NA' (which stands for 'not applicable').

3. In the email attribute, extract the 'sakilacustomer.org' and call this new attribute web_address.

4. The company has a holiday special. The special promo is to charge only the integer part of the amount (or to remove the decimal part). For example, if the amount is \$7.99, the customer will be charged \$7 only.

Come up with this new column that shows this new price (call it holiday_promo).

*Hint: You may refer to the amount attribute in payment table. Now since the amount is **numeric**, you need to **first convert it to string** using a function called CAST and specify the number of characters --- for ### format it would be 4 characters). It looks like the one below and you can put this entire thing inside another function.*

CAST(amount AS character varying(4))