Explore the Magist tables

You have now learnt enough SQL to explore the Magist tables. There is not a single approach for data exploration: your approach will depend on the dataset and the business challenge you’re facing. Whenever you are not very familiar with either of these, the best approach is to ask yourself a few basic questions, starting with “big picture” questions and then zooming in to any aspect that peaks your interest. Let’s try it out with these questions:

1. **How many orders are there in the dataset?** The orders table contains a row for each order, so this should be easy to find out!
2. **Are orders actually delivered?** Look at columns in the orders table: one of them is called order\_status. Most orders seem to be delivered, but some aren’t. Find out how many orders are delivered and how many are cancelled, unavailable or in any other status by grouping and aggregating this column.
3. **Is Magist having user growth?**A platform losing users left and right isn’t going to be very useful to us. It would be a good idea to check for the number of orders grouped by year and month. Tip: you can use the functions YEAR() and MONTH() to separate the year and the month of the order\_purchase\_timestamp.
4. **How many products are there in the products table?** (Make sure that there are no duplicate products.)
5. **Which are the categories with most products?**Since this is an external database and has been partially anonymised, we do not have the names of the products. But we do know which categories products belong to. This is the closest we can get to knowing what sellers are offering in the Magist marketplace. By counting the rows in the products table and grouping them by categories, we will know how many products are offered in each category. This is not the same as how many products are actually sold by category. To acquire this insight we will have to combine multiple tables together: we’ll do this in the next lesson.
6. **How many of those products were present in actual transactions?** The products table is a “reference” of all the available products. Have all these products been involved in orders? Check out the order\_items table to find out!
7. **What’s the price for the most expensive and cheapest products?**Sometimes, having a basing range of prices is informative. Looking for the maximum and minimum values is also a good way to detect extreme outliers).
8. **What are the highest and lowest payment values?**Some orders contain multiple products. What’s the highest someone has paid for an order? Look at the order\_payments table and try to find it out.

Probably, these questions do not reveal anything extraordinary to you, but it is critical to get a sense of what the data holds in a broad sense. Ideally, you can come up with more questions of your own, or simply tweak your queries and absorb more information. It’s also a good idea to start taking notes as early as now to contextualise your presentation.

When you feel you have the big picture of the database, move to the next lesson for more specific business questions.