

Case ETL

Introduction

If you are reading this, you are in the second round of the interview for the Data Engineer position at Data Wizards. We are excited to assess your skills in problem-solving, data modeling and data transformation!

Context

Our valued customer, WizardsCo, operates in multiple product sectors and has provided us with datasets from their sales, customers, and product installations. They have questions about their installations and revenue trends and it's up to you to help them find the answers.

Your mission is to build a structured fact table to facilitate easy querying of the data.

Dataset

You are provided with several datasets detailing the company's customers, product installations, and products. They are described as follows:

1. **country.csv:** Contains information about the countries including their id, name, and region.
2. **customer.csv:** Contains information about the customers, including their id, name, email, country_id, and whether they are a premium customer.
3. **installation.csv:** Contains information about the product installations, including their unique id, name, description, product_id, customer_id, and date of installation.
4. **product.csv:** Contains information about the products, including their id, reference, name, category_id, and price.
5. **product_category.csv:** Contains information about the product categories, including their id and name.

Deliverables

We expect the following for this take-home:

- Create a new table called *fact_installation* with relevant keys and measures that will help in answering the following questions:
 - What is the number of installations that the company is doing every month?
 - Which product category brings us more revenues?
 - Which region of the world is our best market?
- Write SQL queries to answer the above questions based on the *fact_installation* table.
- Document the process, explaining how you created the fact table and how your queries work.

You are free to use whatever approach or tools you deem appropriate, as long as you showcase your ETL and SQL skills.

When you are done, you can a repository or zip file with your code with us and we will review it before the second interview.

You can also prepare a brief presentation to illustrate your approach, results, and insights from the datasets. Should you have any question, do not hesitate to reach out.

Best of luck!