

Before you start

What We're Assessing

This case study evaluates your ability to:

- Work with real-world data and ambiguity
- Deliver quality analytics solutions under time constraints
- Use modern tools (including AI) effectively
- Think critically about data quality and business logic

****Important**:** We value your reasoning and problem-solving approach over "perfect" answers. There are often multiple valid solutions - what matters is that you can justify your choices.

AI Usage Policy

AI tools (ChatGPT, Claude, Copilot, etc.) are encouraged - use them as you would in your daily work. We evaluate your effectiveness and critical thinking.

This test is designed to be completed in 1h to 1h30 with AI assistance. Take your time to ensure quality.

Tips for Success

On Assumptions

- Make assumptions when facing ambiguity - this reflects real-world work
- Explain your reasoning clearly in your submission
- We're interested in *how* you handle uncertainty, not just final answers

On Scope

- You may not need to use every table mentioned
- Focus on answering the core questions well
- Quality over comprehensiveness

Data & Resources

Download Link: [link](#)

The folder contains:

- fact_order_line.csv - Order line level data
- fact_shop.csv - Shop/seller information
- table_documentation.txt - Schema details

Tools: Use SQL, Python, or any tool you prefer for analysis and visualization

External Resources: You're free to use any datasets or information that enriches your work

Submission Instructions

Time: The completion of the case study should take 1h to 1h30.

Output Format: Free format - use whatever structure works best for you (Jupyter notebook, PDF, etc.)

When you're done, email your submission to: marine.auffredou@mirakl.com

- *Subject line:* Analytics Engineer Case Study - [Your Name]
 - *Include:*
 - Your analysis document
 - Your SQL / code used to answer the questions
 - Any assumptions or notes
-

Part 1: Business Analysis [30 - 45 min]

Context

Gross Merchandise Value (GMV) is a critical metric for marketplaces, representing the total value of goods sold. For example, if a marketplace sells 1,000 items at €100 each in a given period, the GMV would be €100,000.

The Customer Success team needs a dashboard to monitor marketplace performance in 2024.

You have access to `fact_order_line` and `fact_shop` tables and the YAML file documenting each table and their columns.

Questions

1. Using the available data, what do you think would be a good definition of GMV and why? Explicitly state what you include/exclude and why?
 2. Based on your definition : identify the top 5 best shops contributing the most to the 2024 GMV for each marketplace. *(Please join your results and your SQL / code in your answer)*
 3. Calculate how many shops were active each month in 2024 and show the month-over-month change *(Please join your results and your SQL / code in your answer)*
 4. Using the available data, suggest 3 additional KPIs that would be valuable for the Customer Success dashboard. For each KPI, briefly state:
 - KPI name and definition
 - Business value (what decision does it inform?)
-

Part 2: Data Modeling [20 - 30 min]

Context

Market+ is a marketplace platform connecting third-party sellers with customers.

They work with sellers ranging from small local shops to large distributors, offering products across many categories and geographic regions leading to a large amount of orders

Order Process Flow

1. Product Listing: Sellers publish products associated with one or more categories. (e.g. "Electronics," "Home & Decor," "Fashion," etc.).
2. Orders: Customers order products, creating order items for each purchased product.
3. Payment: Customer can choose from different payment methods; Every payment information and their related status are recorded
4. Shipping: Order handed to a carrier (or more than one if the order contains items from different sellers) for delivery
5. Commission: Marketplace charges commission per transaction on the total order amount or specific products
6. Reviews: Customers can review products/sellers after delivery

Source Tables Available

Table	Description	Key columns
vendors	Seller information	vendor_id, name, country, signup_date, status
products	Catalog of all products referenced on the marketplace	product_id, vendor_id, name, category_id, price
categories	Product categories (hierarchical)	category_id, name, parent_category_id
customers	Registered buyers	customer_id, name, email, country, signup_date

orders	Order headers	order_id, customer_id, order_date, total_amount, status, carrier_id, commission
order_items	Order items details	order_item_id, order_id, product_id, quantity, unit_price, vendor_id, commission_rate
payments	Payment transactions	payment_id, order_id, payment_method, amount, payment_date, status
carriers	Delivery services	carrier_id, name, service_area, avg_rating
reviews	Customer feedback	review_id, order_id, product_id, customer_id, rating, comment, review_date

Business Rules

- A seller can offer multiple products
- Orders can have ****multiple carriers**** (items from different sellers)
- Order total amount includes the sum of the prices of the ordered products plus any shipping fees and taxes.
- Categories are hierarchical (parent-child relationships)

Task: Design a STAR Schema

You need to model the data to support analytics on Market+'s sales and performance by implementing a star schema.

Your model should be able to provide answers to the following business questions:

1. What is the revenue generated by category and by seller for a given month?
2. What is the order volume and total billed amount by region and/or customer country?
3. How do the marketplace's collected commissions evolve over time?
4. Which payment methods are most used, and what impact do they have on approval or rejection rates?

Guideline questions

- Which fact table(s) do you need? What grain?
- Which dimensions should you extract from existing tables?
- Which metrics (measures) would you store in the fact table?

Deliverable

⚠ You do not need to calculate the answers to the questions above - just design a model that would enable the analysis.

1. Create a STAR schema diagram using dbdiagram.io
 2. Written Summary to address these points
 1. Your choice, trade-offs or assumptions
 2. Potential Modeling Challenges
-