

Candidate Assignment - Data Analyst (Take-Home Assignment)

Time expected: ~4–5 hours

Deadline: Please return within 5 days

Hi,

Thank you again for your interest in joining **Yaballe** as our first Data Analyst.

Since this role involves building the data function from scratch and working independently with multiple stakeholders, we've designed a short take-home assignment that simulates the type of work you would do here.

The task is intentionally scoped so you can complete it within **no more than 5 hours**.

Overview

You are the first data hire at a SaaS company providing automation tools for dropshipping sellers, operating in the US, UK, DE, FR and IL. We help our customers operate and scale their dropshipping activity and sell products from Amazon on eBay.

The CEO has asked for the company's first KPI dashboard to better understand Yaballe product usage, user behavior, and overall business performance.

Your assignment includes:

1. Requirements & Data Modeling
2. Dashboard Design & Build

You may use any tool you like (Tableau, Power BI, Looker Studio, Excel, etc.)

Provided Data

You will receive two CSV files:

1. **customers**
`customer_id, signup_date, country, segment, connection_type, number_of_stores`
2. **subscription_events**
`customer_id, event_ts, event_type, plan_from, plan_to, price_usd, status`

And one JSON file:

- **product_events (JSON array of objects):**
 - `event_id: unique event identifier`
 - `event_name: event type (e.g. user_signed_up, store_connected, first_listing_created, trial_started)`
 - `ts: event timestamp (UTC)`
 - `data: nested object containing event-specific attributes such as user_id, country, segment, store info, listing info, plan, etc.`
`data.user_id aligns to Customers.customer_id.`

PART 1: Requirements & Data Modeling (*Estimated: 1-1.5 hours*)

Please submit a short document (max ~1.5 pages) covering:

A. Requirements Gathering

- What are the most important questions a Product Manager or CEO would want to answer about:
 - User behavior
 - Product usage
 - Business performance
- How would you define a “successful user” in this product? Clearly explain your definition, why you chose it, and what its limitations are.
- What early signals might indicate whether a user is likely to succeed or churn?
- Which KPIs would you prioritize in the first version of company-level analytics? How would you define them?
- What assumptions are you making?

- What questions would you ask stakeholders if this were a real project?
- What data is missing that you would ideally want to answer these questions better?
- Where relevant, explain how Events data informed your answers, and where it was insufficient.

If a question cannot be reliably answered with the available data, explain why and what data you would request next.

B. Data Model Proposal

Design a simple analytics data model using the provided datasets.

Please include:

- A proposed schema (a diagram or structured table is fine)
- Identification of fact and dimension tables
- The grain of each fact table (e.g. one row per event, one row per customer)
- Expected relationships between tables
- How you would handle different types of data (e.g. structured tables vs raw event data)
- A brief explanation of why you structured the model this way and any trade-offs you made
- SQL-style table definitions or transformation notes

This does not need to be complex, clarity and reasoning are what matter.

C. Data Quality & Trust

- Provide **5–10 concrete data quality checks** you would prioritize implementing across these datasets
- For each check, explain:
 - What could go wrong
 - The business impact if the issue is not detected
- Which issues would you prioritize fixing first, and why?

PART 2: Dashboard (Minimal Version)

(*Estimated: 2.5–3 hours*)

Using any tool you prefer, please create a simple dashboard that helps answer: How is the business performing overall, and what signals would leadership monitor to understand performance?

Focus on user behavior, product usage, and subscription outcomes (conversion, retention, churn).

Your dashboard should include:

- **3–5 core KPIs** (defined by you)
- At least 2 breakdowns (e.g. by country, user segment, lifecycle stage, connection_type, number_of_stores)
- **3–5 charts** total - keep it simple and focused
- Clear metric definitions and readable formatting

You may submit:

- A PDF export
- A link to a public Tableau / Power BI / Looker Studio dashboard
- Screenshots

What to Submit

Please send:

- Requirements & data model document (PDF or doc)
- Dashboard export or link
- Any additional assumptions or notes

Evaluation Criteria

We will evaluate your submission across:

- Product and business understanding
- Data modeling clarity and correctness
- Data quality thinking and ownership
- Analytical reasoning and prioritization
- Communication with non-technical stakeholders
- Metric clarity and prioritization

We do not expect perfection. Focus on showing how you think and how you make decisions.

Thank You!

We appreciate your time and effort. If anything is unclear, feel free to ask clarifying questions before you begin.

Looking forward to reviewing your work!