Emergence Data Analyst Take-Home Assessment

SaaS Growth & GTM Analytics (MySQL- Python - Power BI)

---

1. Introduction

Thank you for your interest in the **Data Analyst** role.

This take-home assessment is designed to evaluate your ability to:

- Work with messy, real-world data
- Use SQL effectively for analytical problem-solving
- Apply SaaS and GTM metric knowledge
- Communicate insights clearly through documentation and dashboards
- Independently use documentation and tools to solve problems

This is **not a trick test** and does **not require perfect answers**.
We care more about **your reasoning, assumptions, and approach** than exact numbers.

---

2. Business Context

You are a **Data Analyst at a B2B SaaS company**.

The leadership team wants to better understand:

- Revenue performance
- Customer churn
- Funnel efficiency
- Acquisition channel effectiveness

You are provided with **raw CSV datasets** that intentionally contain:

- Missing values
- Duplicate records
- Inconsistent events
- Edge cases

Your goal is to transform this data into **clear, actionable insights** that leadership can use to make decisions.

---

3. Data Provided

You will receive **three CSV files**:

`customers.csv`

Customer profile information.

`subscriptions.csv`

Subscription lifecycle and revenue data.

`events.csv`

User funnel events and acquisition source data.

⚠️ **Important:**
The data is intentionally imperfect.
You are expected to identify data issues, make reasonable assumptions, and clearly document those assumptions.

---

4. Required Tools (Mandatory)

You **must** use the following tools to complete this assessment.

4.1 MySQL (Mandatory)

Used for:

- Creating tables
- Data cleaning logic
- Analytical queries
- Funnel analysis
- SaaS metric calculations

Expected usage includes:

- Joins
- CTEs (`WITH`)
- Window functions
- Date-based aggregations

4.2 Python (Mandatory)

Used for:

- Loading CSV files
- Initial data exploration
- Validation and sanity checks

Required libraries:

- `pandas`
- `numpy`
- MySQL connector (`mysql-connector-python` or `SQLAlchemy`)

4.3 BI / Dashboard Tool (Choose One)

You may use **any one** of the following:

- Metabase
- Tableau
- Power BI
- Looker / Looker Studio

Used to visualize:

- Revenue trends
- Funnel conversion
- Churn overview
- Segment or acquisition source breakdown

---

5. Tasks

Task 1: Data Loading & Cleaning

- Load all CSV files into MySQL
- Identify data quality issues
- Handle:
    - Duplicate records
    - Missing values
    - Inconsistent or conflicting events
- Clearly document assumptions in SQL comments and/or README

Task 2: Core SaaS Metrics (MySQL Required)

Calculate and explain:

- Monthly MRR
- ARR
- Customer (logo) churn rate
- Revenue churn rate
- Average Revenue per Customer (ARPC)

Task 3: Funnel Analysis

Build the following funnel:

`Signup → Trial → Activated → Paid → Churned`

Analyze:

- Conversion rates
- Drop-off points
- Funnel performance by acquisition source or customer segment

---

Task 4: Dashboard

Create **one dashboard** showing:

- MRR trend over time
- Funnel conversion
- Churn overview
- Source or segment breakdown

Focus on **clarity and insight**, not visual polish.

Task 5: Insights & Recommendations

Provide written insights covering:

- Key growth bottlenecks
- Strongest and weakest acquisition channels
- What you would investigate next
- **1–2 actionable recommendations** for leadership

---

6. Expected Output & Submission Format (Very Important)

6.1 GitHub Repository (Public)

You must submit a **public GitHub repository** with the following structure:

```
data-analyst-assessment/
│
├── data/
│   ├── customers.csv
│   ├── subscriptions.csv
│   └── events.csv
│
├── sql/
│   ├── 01_table_creation.sql
│   ├── 02_data_cleaning.sql
│   ├── 03_core_metrics.sql
│   ├── 04_funnel_analysis.sql
│   └── 05_optional_analysis.sql
│
├── python/
│   └── data_validation.ipynb (or .py)
│
├── dashboard/
│   ├── dashboard_screenshots.png
│   └── dashboard_link.txt   (if applicable)
│
└── README.md
```

6.2 Live Dashboard Access (Preferred)

If your BI tool supports public or read-only access, please include a **live dashboard link**.

Accepted options:

- Public dashboard link
- Read-only or viewer access link
- Embedded dashboard URL

If a live link is **not possible** due to tool or license limitations, please include **dashboard screenshots** instead.

In either case:

- Ensure the dashboard can be understood without additional explanation
- Briefly explain each chart in your README

⚠️ Providing a live dashboard link is **preferred but not mandatory**.
You will **not be penalized** if screenshots are provided instead.

6.3 README.md (Mandatory)

Your README **must** include:

- Overview of the analysis
- Tools used
- Data issues identified
- Metric definitions
- Key insights
- Dashboard explanation
- Assumptions and limitations
- Instructions to reproduce results

⚠️ README quality is **heavily weighted** in evaluation.

---

7. Evaluation Criteria

We evaluate submissions based on:

Area	What We Look For
SQL	Correctness, clarity, structure
Analytics	Proper SaaS metric usage
Reasoning	Clear assumptions and trade-offs
BI	Insight-driven visuals
Documentation	Clear and reproducible explanations

---

8. What We Are NOT Looking For

- Machine learning models
- Over-engineered pipelines
- Perfect UI/UX design
- Guessing without explanation

---

9. Final Notes

- There is **no single correct answer**
- Focus on **clarity, reasoning, and business impact**
- Treat this as a **real-world business problem**

We look forward to reviewing your work.
Good luck!