**💼 Data Engineering Challenge – dbt Case Study (ca. 2–4 Stunden)**

**🧩 Ziel der Aufgabe**

Wir möchten dich im Rahmen des Bewerbungsprozesses als Data Engineer dazu einladen, ein kleines dbt-Projekt umzusetzen. Die Aufgabe ist so gestaltet, dass sie einen realistischen Einblick in deine Arbeitsweise gibt – vom Datenverständnis über die Modellierung bis hin zur technischen Umsetzung und Dokumentation.

Die Bearbeitung sollte insgesamt **nicht mehr als 2–4 Stunden** deiner Zeit in Anspruch nehmen.

**📦 Aufgabe: dbt Mini-Data-Warehouse**

Du erhältst zwei CSV-Dateien mit simulierten Geschäftsdaten:

1. leads.csv – Informationen zu generierten Leads
2. campaigns.csv – Metadaten zu Marketingkampagnen

**Deine Aufgabe:**

1. **Erstelle ein lauffähiges dbt-Projekt**, das die beiden Datensätze verarbeitet.
2. **Modelliere mindestens 3 dbt-Modelle** nach dem folgenden Schema:
   * stg\_\*: Bereinigen und vereinheitlichen der Rohdaten
   * int\_\*: Anreichern / Zusammenführen
   * fct\_\* oder dim\_\*: Finales Datenmodell (z. B. für ein Reporting)
3. **Nutze dbt-Funktionalitäten**, wie z. B.:
   * Tests (z. B. Not Null, Uniqueness)
   * Dokumentation (description: in YAML oder Markdown)
   * (Optional) Jinja-Makros zur Wiederverwendung
4. **Beantworte eine Beispiel-Analysefrage** anhand deiner Modelle, z. B.:
   * „Welche Kampagnen hatten die höchste Conversion Rate pro Kanal?“
   * „Wie viele Leads wurden pro Tag im Zeitverlauf generiert?“
5. *(Optional, Bonus)* Richte ein GitHub-Repository ein:
   * Lade dein Projekt dort hoch (gern öffentlich)
   * Integriere einen einfachen CI-Prozess mit **GitHub Actions**, der:
     + dbt installiert
     + ein dbt build durchführt
     + optional: Tests ausführt

**🗃️ Datenbeschreibung**

**leads.csv (Beispiel)**

| **lead\_id** | **campaign\_id** | **created\_at** | **lead\_type** |
| --- | --- | --- | --- |
| 1 | 123 | 2024-05-10 14:32:00 | inbound\_form |
| 2 | 124 | 2024-05-11 09:50:00 | callback\_form |

**campaigns.csv (Beispiel)**

| **campaign\_id** | **channel** | **campaign\_name** | **start\_date** | **end\_date** | **budget** |
| --- | --- | --- | --- | --- | --- |
| 123 | Google | Brand-Kampagne Mai | 2024-05-01 | 2024-05-31 | 10000 |
| 124 | Facebook | Retargeting Q2 | 2024-04-01 | 2024-06-30 | 5000 |

**📌 Anforderungen & Hinweise**

* Du darfst Python, dbt Core oder dbt Cloud verwenden – bitte beschreibe in der README, wie dein Projekt ausgeführt wird (z. B. dbt run + dbt docs serve)
* Lokale Ausführung mit SQLite oder DuckDB reicht aus
* Das Projekt muss lauffähig und dokumentiert sein
* Du kannst frei wählen, wie umfangreich du die Modelle baust – lieber klar & durchdacht als überkomplex

**📤 Abgabe**

Bitte reiche dein Ergebnis folgendermaßen ein:

* **GitHub-Repo-Link** (empfohlen) oder ZIP-Datei
* Kurze **README.md** mit:
  + Installationsanleitung
  + Beschreibung deines Modells & deiner Analysefrage(n)
  + Optional: Ideen zur Weiterentwicklung

Deadline & Abgabetermin sprechen wir individuell mit dir ab.

**💬 Präsentation im Gespräch**

In einem ca. 20-minütigen Slot im Interview möchten wir dich bitten:

* Kurz dein Projekt & deine Modelle zu erklären
* Deine technische Herangehensweise zu erläutern
* Offene Fragen zu beantworten (z. B. „Wie skalierbar wäre dein Ansatz?“)