CANDIDATE ASSESSMENT

Senior Data Analyst Case

Take-home assignment for data analyst candidates

Submission: Create a GitHub repository on your own account and send us the link
Tools: Use whatever tools you're comfortable with (Python, R, SQL, Excel, etc.)

About This Case

At Techleap, we analyze the Dutch tech ecosystem. We work with datasets from sources like Dealroom, Crunchbase, and government registries to produce flagship publications like the State of Dutch Tech Report and Academic Venture Monitor.

This case mirrors the actual work you'll do: transforming messy startup data into clear insights, and designing data infrastructure that scales. We value clear thinking over perfect code.

What We're Looking For

Analytical Thinking

Can you extract meaningful insights from noisy data?

Communication

Can you tell a clear story to non-technical stakeholders?

Technical Design

Can you design scalable, maintainable data pipelines?

Tool Judgment

Can you choose the right tools and justify your choices?

On Tools: We care about your reasoning and results. AI assistants are not required — strong coding skills with traditional tools are equally valued. If you do use AI, tell us what for and why.

The Dataset

You'll work with the **Startup Investments (Crunchbase)** dataset from Kaggle. This public dataset closely resembles the ecosystem data we analyze at Techleap.

Download: https://www.kaggle.com/datasets/arindam235/startup-investments-crunchbase

Dataset Overview

- **54,000+ companies** from 1902-2014
- **39 columns** covering company info, funding rounds, and investor details
- **750+ market categories**

Key Fields

- name, category, market, status
- country, state, region, city
- funding_total_usd, funding_rounds
- seed, venture, angel, grant, private_equity

- Data quality issues typical of real-world datasets

1 Data Analysis Report

Scenario

Techleap's director asks: "We're meeting with the Ministry of Economic Affairs next week. They're interested in how we can better support the startup ecosystem. Can you explore this dataset and put together a briefing with your findings?"

Your Task

Create a concise report (1-2 pages, PDF or HTML) that:

1. **Tells a story:** What patterns or trends do you find meaningful? Some angles to consider: time between funding rounds, funding patterns by sector or geography, investor concentration, survival rates — but follow what the data tells you.
2. **Considers implications:** What might your findings mean for policymakers? For founders? Be honest about what the data can and cannot tell us — external factors (regulations, interest rates, market conditions) matter but aren't in the dataset.
3. **Acknowledges limitations:** What questions couldn't you answer? What additional data sources would help? Who else would you want to talk to?

Requirements

- Include 2-3 visualizations that support your narrative
- Write for a non-technical audience
- Include a "What's Missing" section: What analysis would you still like to do? What data would you need? What expertise would complement yours?

DELIVERABLES

- Report (PDF, HTML, or Markdown) - max 2 pages
- Analysis code (Python notebook or scripts)
- Brief note on tools used

2 Data Engineering Design

Scenario

Techleap receives daily data exports from Dealroom (35-40GB across 4 NDJSON files). We want to make this data queryable for analysts. How would you approach this?

The Data (Simplified)

Each record looks roughly like this:

```
{
  "company_id": "abc123",
  "name": "TechStartup B.V.",
  "industries": ["logistics", "artificial-intelligence"],
  "founding_date": "2018-03-15",
  "employee_count": 85,
  "total_funding_eur": 12500000,
  "funding_rounds": [
    {"date": "2019-06-01", "type": "seed", "amount_eur": 500000},
    {"date": "2023-09-01", "type": "series-b", "amount_eur": 8000000}
  ],
  "investors": [{"name": "Peak Capital", "type": "vc", "lead": true}],
  "headquarters": {"city": "Amsterdam", "country": "Netherlands"},
  "updated_at": "2024-01-15T08:30:00Z"
}
```

Your Task

Sketch out how you'd structure this data for analysis. We're looking for your thinking, not a production-ready design.

Consider

- How would you model this for SQL-based analysis?
- How would you handle the nested arrays (funding_rounds, investors)?
- What if we want to track how companies change over time?
- How would you integrate additional data sources (e.g., patent filings, scientific publications, news articles)?

DELIVERABLES

- A simple diagram or description of your proposed tables
- Brief notes on your approach, trade-offs, and how you'd extend it

Submission Guidelines

Create a public GitHub repository (or private with access granted to us) and send us the link. We'll clone your repo to review your work.

Repository Structure

```
your-name-techleap-case/
├── README.md              # Overview, how to run your code, and tool choices
├── task-1-analysis/
│   ├── report.pdf         # Your 1-2 page briefing
│   └── analysis.ipynb     # Or .py files
├── task-2-design/
│   ├── design-doc.md      # Written design document
│   └── data-model.png     # Or .drawio, .mermaid, etc.
└── tools.md               # What tools you used and why (optional if covered in README)
```

On Tool Choices

Tell us what you used and why. We're interested in your reasoning, not specific tools.

Evaluation Criteria

Task 1: Analysis (50%)

Insight quality, clarity of communication, honest about limitations

Task 2: Design (30%)

Practical data model, clear trade-off reasoning

Tool Choices (20%)

Justified tool selection, pragmatic approach

Bonus Points

Clean code, creative insights, thoughts on unstructured data

Questions?

If anything is unclear, email us. We'd rather answer questions upfront than have you guess wrong. Asking good clarifying questions is a strength, not a weakness.

Final Note: We don't expect perfection. We want to see how you think, communicate, and approach problems. A simple solution with clear reasoning beats a complex solution you can't explain. Good luck!

Techleap | Supporting the growth of Dutch startups and scaleups
techleap.nl