

Pioneers Internship - assignment

Context

Pioneers is exploring an automated way to **discover potential founders** who could be a strong fit for our programs. We care about candidates with deep domain expertise, repeat-founder signals, and/or strong technical backgrounds. Your task is to build an **AI-powered sourcing agent** that finds and ranks such people, and (in a separate step) a **simple website** to browse the results.

1) Goal

TASK 1: Build an agent that:

- Takes as input a list of selection criteria for founders
- Searches across one or more people-data sources (we prefer LinkedIn but use any source you deem relevant)
- Returns a list of candidates with justification and contact links
 - Bonus points if it is ranked
- Exports results to a CSV, database or an Airtable

TASK 2: Build a website to view the results in whichever format you chose (like a founder CRM).

- Bonus points if you can login and sign-up on the website.

OBS: we know that this is a big assignment but the goal is for you to show us what you are capable of. Kill it!

2) Data Sources (choose any)

We suggest LinkedIn (<https://harvest-api.com/> or via another provider).

- Harvest has free credits, you can just recreate an account if you run out.

Feel free to use any other relevant sources (e.g., company sites, GitHub, Google Scholar, Crunchbase, academic portals).

3) Inputs → Criteria

Your agent should accept structured criteria such as:

- **Experience depth:** e.g., ≥10 years in *Supply Chain* at C-level or key role.
- **Academic:** e.g., PhD in Computer Science or adjacent field.
- **Founder signal:** repeat founder, prior exit, or startup leadership.
- **Technical signal:** clear technical expertise (e.g., CTO, notable OSS, patents).

Example criteria payload:

- If Tech profile the founder has already worked in a start-up
 - The founder has already launched a company or tech start-up
 - If tech profile, they build products from scratch
 - The founder is a sector expert with 5 years of experience in X industry
-

4) Required Output (CSV, database or airtable)

For each candidate, produce the following info:

- **name**
- **profile_type** (business | technical)
- **summary** (2–4 lines, your synthesis)
- **contacts** (LinkedIn URL and/or email or other contact path)
- **source_links** (list of URLs used as evidence)
- **match_justification** (why they fit the criteria)
- **tier** (A/B/C; A = best fit)