Kickoff meeting questions

**🔹 Data & Database**

* What’s the exact schema of the Postgres DB? (tables, columns, keys)
* Where/how do I get access? (credentials, VPN, etc.)
* How often does the scraper update the data? (daily, weekly?)
* Do I need to **write enriched skills back** into the same table, or a separate table?

**🔹 Skills Extraction (NLP / LLM)**

* Should we **stick to local LLM (Ollama)** for now, or is external API (OpenAI, Anthropic, etc.) already being considered?
* Is there an existing **taxonomy or categories of skills** they want us to use (e.g. “hard vs soft”, or “match to reference architecture components”)?

**🔹 Reporting / Dashboards**

* Power BI
* What’s the **priority metric** for Sales (UC1)? (e.g. “Top companies hiring for Python”, “New Data Engineer jobs in Belgium last 7 days”)

**🔹 Project Setup & Workflow**

* Should we **version-control** the scripts (Git)?
* Any coding standards or guidelines to follow internally?

Workflow proposal

Tables:

* + job\_postings (job\_posting\_ID, title, function, ..., description, ...)
  + skills(skill\_ID, name, group, subgroup)
  + job\_postings\_skill\_mapping(job\_posting\_ID, skill\_ID, conf\_score, context, sentence, extracted\_at, ...)

Capture skills level/context options:

* + Rating system via appendable dict of words (“expert”: 9-10, “proficient”: 7-8, “basic”: 3-4)
  + NLP embedding based (in job\_postings\_skill\_mapping OR vectorDB for faster queries)
  + Tiered skill levels (beginner, intermediate, advanced, expert -> better for reporting?)
  + Hybrid approach (best?): tiered levels but extra category vector embedding as backup for semantic queries
  + NEW SKILLS: not append on taxonomy (OR use ‘new’ flag?) , new category ‘other’ for each (sub)group? OR vector embedding!
  + Vector embedding would also handle duplicates efficiently (backup?)

Active postings flag:

* + Compare DB (for search word) with scrapings to check for postings that are still active and which ones have been taken
  + Draw line where (history)?