Context: Our Directory Workflow (Read before answering)

We are building a community-verified directory (Practo-for-Animals style) for categories like vets, shelters/NGOs, feeders, boarding, trainers, groomers, pet shops, etc. Many of these are unorganised/local businesses, so information will often be incomplete or inconsistent.

How data comes in (Primary Layer)

Listings will be sourced through:

- Intern research (Google Maps, Instagram pages, Justdial, Facebook groups, websites, etc.)
- Community submissions (WhatsApp, forms, DMs, referrals)
- Existing Pawzz network inputs (rescuers, feeders, partners)
- You scraping the data yourself

Q1) Tools + System Design for Community-Verified Directory (Long Answer)

We are building a community-verified directory of unorganised/local businesses (vets, shelters/NGOs, feeders, boarding, trainers, groomers, etc.). Data will come from interns + public sources (google, social media scraping) + community submissions, so it will be messy and inconsistent.

Question:
If you were the lead Data Analyst for this directory, **how would you design the entire data collection + cleaning + verification workflow** so that listings are accurate, scalable, and auditable?
In your answer, explain **which tools you would use and why** (Google Sheets/Excel, SQL, Python, Power BI, Airtable/CRM, forms, automations, etc.) and how you would structure:

- required fields & proof standards per category
- data pipeline stages (submitted → needs fix → ready → verified → published)
- dashboards/KPIs to track throughput, quality, and backlog
- how you would handle unstructured inputs (WhatsApp messages, screenshots, inconsistent addresses/phone formats)

(Answer in a detailed plan — assume you are building it for a real team and must reduce founder dependency.)

Q2) Red Flags + Verification + Duplicate Prevention (Long Answer)

Because listings are from unorganised/local sources, we'll face fake/incorrect data, duplicates, and inconsistent details.

Question:
What are the top red flags you would use to detect unreliable listings, and what would your verification + duplicate-prevention system look like end-to-end?
In your answer, cover:

- your verification checklist for at least one category (e.g., vet clinic or NGO)
- how you'd handle conflicts (two different phone numbers/addresses for same business)
- how you'd design an audit trail (who verified, when, what proof, what changed)
- escalation rules: when to reject, when to re-check, and when to publish with warnings