Rapid API Data Pipeline

Data Sources
- TechCrunch RSS Feed
- RapidAPI SEO Keyword Endpoint

⬇

Ingestion
- Python scripts (for API + RSS extraction)
- GitHub Actions (daily scheduled pipeline)

⬇

Processing
- Extract API/Web data
Rename text → keyword
- Drop null or empty keywords
- Add metadata:
source = 'RapidAPI'
Scraped_at = UTC timestamp

⬇

Storage
- PostgreSQL RDS Instance
- Schema: raw
Tables:
- api_keywords
- techcrunch_articles

⬇

Users
- Looker Studio Dashboards
SQL Notebooks (Jupyter)
- Final Presentation Assets (PDF/Slides)