**Plano completo de ambiente, estudos e empregabilidade — Samuel**

Objetivo: montar um ambiente local (com integrações gratuitas de nuvem quando úteis), aprender as ferramentas essenciais de Data Science/Engineering e automatizar sua prospecção de vagas, usando RPA e workflows para acelerar seus resultados.

# Estrutura de pastas (criar antes de instalar)

D:\Micro Samuel Atual\Documentos\  
└─ dados\_lab\  
 ├─ \_bin\ (scripts .bat / utilidades)  
 ├─ \_logs\ (logs de start/backup)  
 ├─ \_notes\ (anotações, .md, .docx)  
 ├─ docker\_stacks\  
 │ ├─ n8n-project\ (n8n + Postgres + Jupyter)  
 │ ├─ Mage-Projetos\ (pipelines Mage)  
 │ └─ monitoring\ (Portainer/Dozzle – opcional)  
 ├─ notebooks\ (Jupyter / estudo)  
 ├─ datasets\ (dados públicos/baixados)  
 ├─ projects\  
 │ ├─ 01-emprego-automation\ (projeto piloto 1)  
 │ └─ 02-analytics-portfolio\ (projeto piloto 2)  
 ├─ bi\_tools\  
 │ ├─ powerbi\  
 │ ├─ tableau\_public\  
 │ ├─ qlik\_sense\_desktop\  
 │ └─ metabase\ (self-hosted, via Docker)  
 ├─ rpa\  
 │ ├─ tagui\ (RPA #1)  
 │ └─ openrpa\ (RPA #2 – opcional)  
 ├─ ml\_stack\  
 │ ├─ python\_envs\ (ambientes virtuais)  
 │ ├─ mlflow\ (tracking – opcional)  
 │ └─ models\ (modelos exportados)  
 └─ docs\  
 └─ plano\ (este documento e versões)

# Onda 0 — Pré‑requisitos

* Windows 10/11 atualizado; Docker Desktop com WSL2; Git; PowerShell 7 (opcional).
* Contas: GitHub, Kaggle, Google (Drive/Colab), Hugging Face (para modelos), LinkedIn.

# Onda 1 — Núcleo do ambiente (Docker)

## 1. n8n + Postgres + Jupyter (stack já operacional)

* Pastas: D\...\docker\_stacks\n8n-project e Mage-Projetos.
* Scripts prontos: start\_n8n.bat (sobe serviços e gera dashboard HTML) e backup\_n8n.bat (para e faz push no Git com log).
* Jupyter em http://localhost:8888; n8n em http://localhost:5678; Postgres na 5432.

## 2. Mage (ETL de código + UI)

* Container mage-etl ligado à mesma rede Docker da stack.
* Pipelines Python para ingestão/transformação com versionamento no Git.

## 3. Observabilidade (opcional)

* Portainer (UI de Docker) e Dozzle (logs) para facilitar troubleshooting.

# Onda 2 — BI & Analytics

* Power BI Desktop (grátis) — foco no mercado brasileiro; exportar amostras para .pbix no repositório.
* Tableau Public (grátis) — publicar visuais no perfil público.
* Qlik Sense Desktop (avaliar disponibilidade) — aprender a lógica associativa.
* Metabase (open‑source via Docker) — dashboards self‑hosted ligando no Postgres.

# Onda 3 — RPA & Job Automation

* TagUI (RPA open‑source) — scripts simples para automação de navegação e captura.
* OpenRPA (open‑source) — alternativa mais robusta com orquestração (Node‑RED/OpenFlow).
* Integração com n8n: orquestrar rotinas de prospecção, deduplicação e envio de mensagens.

# Onda 4 — Data Science / IA (local + nuvem gratuita)

* Ambientes Python isolados (venv/conda) com pacotes: numpy, pandas, polars, scikit‑learn, nltk, spacy, jupyterlab, matplotlib/plotly, requests/httpx, sqlalchemy/psycopg2, mlflow (opcional).
* Hugging Face: baixar modelos graciosos para embeddings/classificação; uso local sempre que possível.
* Colab/Kaggle Notebooks (grátis) para treinos pontuais e publicação de notebooks.
* MLflow (opcional) para tracking de experimentos.

# Ferramentas adicionais (empregabilidade)

* VS Code e/ou Cursor (IA pair‑programmer) — produtividade e integração com Git.
* DBeaver (cliente SQL universal) — dominar SQL e modelagem.
* DBT (opcional, após base sólida) — padronizar transformações em SQL com testes e docs.
* LinkedIn Helper/Phantombuster (trial) — evitar abuso; preferir n8n + RPA próprios.

# Piloto 1 — Prospecção de emprego automatizada

1. Pipeline de leads: buscar vagas (APIs, sites), normalizar, classificar, gravar no Postgres.
2. Ranqueamento: heurísticas + ML leve (ex.: regressão logística) gerando score de aderência.
3. Ações: n8n dispara RPA (TagUI) para preparar mensagens personalizadas e abrir links.
4. Logs e evidências: tudo no Postgres + Metabase para acompanhar funil e taxas.

# Piloto 2 — Portfolio de Analytics

* Dados públicos (Kaggle/IBGE): montar painéis comparáveis em Power BI, Tableau, Qlik e Metabase.
* Publicar notebooks no Kaggle/Colab e artigos no LinkedIn com links para o GitHub.

# Trilha de estudo (4–6 semanas intensivas)

## Semana 1 — Fundamentos & SQL

* Git/GitHub, Docker, Postgres: CRUD e JOINs; prática no DBeaver; primeiros dashboards no Metabase.

## Semana 2 — Python para Dados

* Pandas/Polars, Jupyter, requests; leitura/escrita Postgres; tratamento de dados reais.

## Semana 3 — ETL & Orquestração

* Mage (pipelines), n8n (triggers, filas), padrões de logs e retries.

## Semana 4 — BI & Storytelling

* Power BI, Tableau, Qlik: mesmo dataset em 3 ferramentas; publicação e narrativa.

## Semana 5 — RPA & Automação

* TagUI/OpenRPA: automação segura; integração com n8n; limites anti‑spam.

## Semana 6 — ML prático

* scikit‑learn básico, métricas, validação; opcional: MLflow; deploy simples do modelo para scoring.

# Boas práticas & rotinas

* Backups diários (git push + logs), versionamento sem dados sensíveis; .env em local seguro.
* Kanban pessoal (Projects do GitHub) e diários de bordo em docs\plano.
* Checklist antes de entrevistas: portfolio atualizado, repositórios limpos, README com prints e links.

Pronto. Este documento resume o que instalar, como organizar e o que praticar para maximizar sua empregabilidade em dados/IA — mantendo tudo funcional no seu ambiente local e com trilhas gratuitas em nuvem quando fizer sentido.