

Arfan Shaikh

Data Engineer

github.com/Arfan10 | candidates.intervues.club/u/arfan-shaikh

SUMMARY

Data Engineer candidate with background in software development internships and academic qualifications in MSc and BSc. Experienced in Python-based backends, data extraction, Streamlit dashboards, and mobile app development using SQLite. Skilled in API integrations, data processing, and cloud database technologies.

EXPERIENCE

Data Engineer Intern

BirthdayMessaging

Jan 2025 – Present

- Extracted and processed project and admin data from Gitea to construct an interactive Streamlit dashboard.
- Diagnosed and resolved production issues in a FastAPI-based email verification service.
- Engineered backend features using Python and Flask while providing support for a React, TypeScript, and Vite Chrome extension.

Frontend Developer Intern (Flutter)

REYANK Technology

Nov 2024 – Mar 2025

- Developed a Flutter mobile application for a wholesale cloth seller to track inventory levels using local on-device SQLite storage.
- Integrated client application with backend REST APIs to enable real-time synchronization of stock data.

EDUCATION

University of Hertfordshire, Hatfield · MSc · 2027

KM Aggarwal College, Mumbai · BSc · 2025

SKILLS

Languages and Frameworks: JavaScript, Jupyter Notebook, Python, TypeScript, FastAPI, Flask, Flutter, React, Streamlit

Databases: AWS RDS, PostgreSQL, SQLite

Tools: Airflow, API, AWS, AWS IAM, AWS S3, dbt, Docker, Gitea, Matplotlib, Microsoft Excel, NumPy, Pandas, Plotly, Power BI, Vite

PROJECTS

UK JOB MARKET TIGHTNESS INDEX (JMTI)

Built staging/fact dbt models (stg_jobs, fct_jobs, stg_job_skills, fct_job_skills) on PostgreSQL and implemented skills keyword extraction, processing and validating ~1,500 daily records. Constructed staging and fact dbt models on PostgreSQL to process and validate approximately 1,500 daily job listings. Implemented skills keyword extraction mechanisms using Airflow, AWS S3, RDS, and Docker.

PostgreSQL, dbt, Power BI, Airflow, Docker, AWS, S3, RDS, IAM, WSL2, psycopg2, psycopg3, Adzuna API

github.com/Arfan10

Coffee Sales Dashboard

Cross-referenced three linked tables (orders, customers, products) using XLOOKUP/INDEX-MATCH to build a unified sales view.

Cross-referenced multiple relational datasets utilizing advanced Excel formulas to establish a unified sales view. Designed dynamic pivot tables and integrated slicers to visualize store sales performance metrics.

Microsoft Excel

github.com/Arfan10/Excel-coffee-dashboard

Indian Startup Funding EDA & Dashboard

Cleaned a messy, real-world funding dataset in Python/Pandas and built an interactive exploration dashboard with Streamlit and Plotly.

Cleaned and restructured real-world venture funding datasets using Python and Pandas. Developed an interactive web visualization platform utilizing Streamlit and Plotly for exploratory data analysis.

Python, Pandas, Streamlit, Plotly

github.com/Arfan10/Startup-funding-

Additional projects: Airline Passenger & Pricing Analysis