NINAD N PATIL

United Kingdom | +44 7471566316 | pninad009@gmail.com | LinkedIn |GitHub

PROFESSIONAL SUMMARY

Experienced Data Analyst with nearly 4 years of experience who focused on business outcomes with SQL/Python, Power BI/DAX, Excel and Azure. At DRG, built ADF pipelines that cut daily refresh ~50% and automated ~16 staff-hours/week. Delivered 15+ executive dashboards enabling self-serve analytics and faster decisions across Ops & Finance. Databricks (DLT/Delta) personal projects demonstrate end-to-end data ops tied to clear KPIs. Translate ambiguous asks into SQL/Python analyses and clear KPIs/OKRs, pairing A/B testing with data quality & governance to drive confident, SLA-aligned decisions.

SKILLS

- **Analytics & BI:** Power BI, DAX, Power Query (M), KPI design, data visualization, dashboarding, row-level security (RLS), data storytelling, self-serve BI, Excel (Pivot Tables, VBA Automation)
- **Programming & Data:** SQL (CTEs, window functions, performance tuning), Python (Pandas, NumPy, PySpark, SQLAlchemy, Flask), Excel (advanced formulas, PivotTables, VBA), REST APIs (JSON/XML)
- **Data Modeling & Warehousing:** Star/Snowflake schema, dimensional modeling, SCD1/SCD2, semantic models, ELT/ETL, data marts
- **Cloud Platforms:** Azure Data Factory (ADF), Databricks (Delta/Autoloader, DLT), ADLS Gen2, Azure SQL Database, PostgreSQL, Git/GitHub, CI/CD
- **Statistics & Experimentation:** Hypothesis testing, A/B testing, regression, cohort analysis, segmentation, time-series analysis, sample sizing
- **Quality & Governance:** Data profiling, validation rules, anomaly detection, documentation, lineage, UAT, SLA/OKR alignment
- **Ways of Working:** Stakeholder management, requirements gathering, JIRA, cross-functional collaboration

WORK HISTORY

Data Analyst   **May 2024 – Present**
The DRG, United Kingdom

- Built a data & reporting stack using Azure, Python, SQL, Power Bi and Excel, enabling near-real-time insights for business stakeholders and faster decision-making.
- **Reduced daily refresh ~50% (120 to 60 mins)** across core **ADF** pipelines via SQL optimization, **incremental loads**, and reusable **Python** transforms data ready for **6am** Ops/Finance huddles.
- **Automated ~16 staff-hours/week** by orchestrating **VBA/Python** jobs and SQL scheduling to burst standardized PDF packs to **20+ sites**, cutting manual reporting and formatting defects.
- **Designed and analyzed A/B tests** (hypotheses, sample sizing, lift analysis, decision memos) to guide campaign rollout for Loyalty/Marketing.
- **Integrated REST/JSON and web data** into normalized relational models (**Azure PostgreSQL**); improved query ability and shortened time-to-insight for weekly performance reviews.
- Implemented automated Power BI dashboards for **nearly real-time** reporting and strategic decision-making, increasing quarterly revenue by 15% through enhanced sales insights.
- Automated daily REST API ingestion and web scraping (**BeautifulSoup, Selenium**); normalised **JSON/XML** to relational tables, improving database query performance and usability for analysts.

Data Analyst   **April 2021 - March 2022**
NSEIT, Nashik

- **Streamlined SQL/Python ETL** for operations and examination data, **freeing ~10 hours/week** and improving report turnaround for internal stakeholders.
- Implemented data validation & reconciliation checks (row counts, referential integrity, anomaly flags) to raise reporting reliability and reduce rework.
- Produced and maintained **20+ decision-support reports/dashboards** with defined **KPIs** and **SLAs**, improved stakeholder satisfaction and consistency.
- **Standardized documentation and lineage** with a six-member data team (Git/GitHub), improving auditability and onboarding.

Data Analyst   **March 2020 – March 2021**
Swami Vivekanand Education and Research Centre, **Mumbai**

- Built **Power BI** dashboards and optimized **SQL** models to surface customer analytics, **reducing data processing time ~20%**.
- Developed Power BI dashboards for real-time visualization of key metrics.

- Partnered with Marketing on **KPI definitions** and segmentation; **increased stakeholder engagement ~30%** through actionable visuals.
- Partnered with Admissions/Marketing on **KPI definitions** and segmentation; **increased stakeholder engagement ~30%** through actiona

EDUCATION

- MSc in Advance Computer Science with Data science (2:1) **January 2023 – January 2024**
  University of Strathclyde**,** School of Computing Science
- Bachelor of Computer Applications (2:1) **May 2016 – March 2020**
  Savitribai Phule Pune university

PROJECTS

End-to-End Azure Data Engineering Lakehouse ([GitHub Link](#))

- Medallion (Bronze/Silver/Gold) on **ADLS Gen2**; **ADF** pipelines are parameterized & metadata-driven with incremental loads, backfill controls, and CDC watermarking (JSON).
- **Databricks Autoloader + Structured Streaming** refine Bronze to Silver **Delta**; governance via **Unity Catalog** (metastore, external locations, access connector).
- **Delta Live Tables** curate Gold: **SCD2** dims (sequence_by) & **SCD1** upserted facts; **Expectations** + visual lineage DAG; analytics-ready for **SQL Warehouse/Power BI**.
- Ops & DevEx: Python utilities (cleaning, de-dupe), checkpointing & idempotent runs, empty-file cleanup, **Logic Apps** alerts, secrets via **Managed Identity/Key Vault**, CI/CD with **Databricks Asset Bundles** + GitHub

DataBricks-Declarative-Pipelines ([GitHub Link](#))

- Built an end-to-end **Lakehouse** with **DLT** using a **Medallion** design, unifying **batch + streaming** with automatic DAG/lineage.
- Bronze: **streaming ingestion** with **Expectations** (warn/drop/fail) and **Append Flow** to consolidate multi-region sources; captured run-time **audit metrics** for observability.
- Silver: **Auto-CDC (Type-1)** upserts (keys + sequenceBy) with reusable Python utilities (casting, de-dup, common transforms) to keep curated tables incrementally accurate.
- Gold & Ops: **SCD2 dimensions** + Type-1 fact and a materialized business view (revenue by region×category); **Unity Catalog** governance, parameterized dev/prod configs, SQL seed scripts, and monitored alerts.

Azure Data Factory End-to-End (On-Prem + API + SQL) ([GitHub Link](#))

- Built a hybrid ingestion layer in ADF with **Self-Hosted IR** for on-prem file shares, **HTTP/REST** for GitHub raw JSON, and Azure SQL incremental loads without watermark tables (state kept in **JSON** in the lake) enabling **idempotent re-runs** and drift-tolerant landings
- Orchestrated a Bronze, Silver, Gold flow: raw landings to Bronze, **Mapping Data Flows** to standardize/clean/type-cast into **Silver Delta** with **Alter Row** upserts on business keys, then curated **Gold** views (joins, aggregations, **dense rank, Top N**) optimized for BI.
- Centralized control via a **Parent pipeline** that chains on-prem, API, and SQL paths, parameterizes file lists/batch sizes, and runs safe steps in parallel reducing manual coordination and making backfills/replays predictable
- Shipped with **ARM publish artifacts** and documented runbooks for **repeatable CI/CD**, plus clear guardrails on schema drift vs. validation, partitioning options, and overwrite strategy for Gold snapshots.

Dockerized End-to-End ETL Pipeline with PostgreSQL, ChromaDB & Sentence-Transformers:

- Developed a Python-based ETL pipeline to extract and clean UK legislation Webpages into structured JSON.
- Stored data in PostgreSQL, generated embeddings with sentence-transformers, and indexed for vector search using ChromaDB.
- Built a CLI for semantic querying, retrieving relevant documents from ChromaDB.
- Containerized the complete solution (PostgreSQL, ChromaDB, ETL, CLI) with Docker for streamlined deployment.