Akshat Mishra

New Delhi, India | +91 9599318801 | akshatm.work23@gmail.com | linkedin.com/in/akshat-mishra-9967b9108 | github.com/froxily

Professional Summary

Computer Science and Engineering undergraduate (B.Tech, 2027) with two software internships, seeking graduate data analyst, data engineering and data science roles. Built tested Python and SQL data pipelines and analytics applications on PostgreSQL and DuckDB, spanning ETL, data quality, statistical analysis and dashboards. English C2; Japanese JLPT N3 level.

Education

Bennett University — **Greater Noida, India**
B.Tech in Computer Science and Engineering — *2023 – 2027 (expected)*
Coursework: Data Structures, Algorithms, Object-Oriented Programming, DBMS, Operating Systems, Computer Networks

St Thomas School — **Ghaziabad, Uttar Pradesh, India**
CBSE Board

Experience

Ornate TechnoServices Pvt. Ltd. — **Jan 2026 – Feb 2026**
Software Development Intern — *Greater Noida, India (on-site)*

- Handled data-management, backend and frontend tasks with MySQL, Java/Spring Boot and React for an in-development government-related research portal that stores and searches scientists' profiles and research records.
- Worked in a five-person team using Git/GitHub and daily stand-ups; changes were merged into the team codebase.

Medelan Life Sciences Pvt. Ltd. — **Jun 2025 – Jul 2025**
Web Development Intern — *Ghaziabad, India (hybrid)*

- Built pages for the company website in React, JavaScript, HTML and CSS with one other developer, some independently.
- Designed page layouts in Figma and refined them through regular review feedback; managed changes with Git/GitHub.

Projects

Hoshizora | *Python, SQL, PostgreSQL, Flyway, Spring Boot, React, TypeScript, ECharts, Azure* — Live app GitHub

- Built a bilingual (English/Japanese) space-data analytics platform: a Python ETL pipeline loads five public APIs (launches, satellites, asteroids, space weather, exoplanets) into PostgreSQL and serves 15 analytical marts via a read-only API.
- Modelled the data in raw, core, mart and fact layers across 21 Flyway migrations, covering 8,100+ launches, 71,000 space objects and 52,000 asteroid close approaches, with lineage links from 390,000+ rows back to their source records.
- Made ingestion quota-safe and resumable with per-source rate limits, exponential backoff, watermark-based incremental loads, SHA-256-deduplicated raw storage and idempotent upserts, so repeated runs change nothing.
- Wrote the analytics in SQL (window functions, filtered aggregates, materialized views): launch cadence by nation, rolling 20-launch success rates, booster-reuse trends and log-binned exoplanet size distributions, charted in ECharts.
- Enforced data quality with Pydantic schemas per payload, blocking checks for unmapped codes and a shrink guard on snapshot loads; 541 automated tests (pytest, JUnit, Testcontainers, Vitest, Playwright) run in GitHub Actions CI.

MIERU | *Python, DuckDB, pandas, SciPy, statsmodels, scikit-learn, FastAPI, React, Plotly, Azure* — Live app GitHub

- Built a retail analytics platform that turns CSV/Excel sales exports into dashboards: it detects encodings (incl. Shift_JIS), delimiters and column roles, maps to a canonical schema and applies reversible, audited data-quality policies.
- Implemented business metrics in parameterized DuckDB SQL and pandas: a revenue bridge from gross sales to gross profit, returns, RFM customer segmentation, right-censored acquisition cohorts and Pareto/Gini seller concentration.
- Validated the pipeline on the public Olist marketplace dataset: a generic upload of 112,650 order lines reproduces direct SQL exactly, and seven DuckDB marts power dashboards over 96,470 delivered orders.
- Analysed delivery performance with hypothesis tests (chi-square, Welch t-test, Mann–Whitney) and logistic regression: a missed delivery date is associated with 16.3 times higher odds of a 1–2 star review (95% CI 15.4–17.3, observational).
- Compared naive and moving-average baselines with Ridge and Random Forest forecasts (scikit-learn) using leakage-free rolling-origin backtests, selecting by MAE; 119 backend and 113 frontend tests pass; deployed on Azure Container Apps.

Technical Skills

Programming Languages: Python, SQL, Java, TypeScript, JavaScript
Data Analysis: pandas, NumPy, SciPy, statsmodels, scikit-learn, exploratory data analysis, data cleaning and validation
Statistics and Methods: hypothesis testing, confidence intervals, logistic regression, RFM and cohort analysis, forecasting
Databases and SQL: PostgreSQL, DuckDB, MySQL, window functions, CTEs, materialized views, layered data modeling
Data Engineering: ETL pipelines, incremental and idempotent loads, data-quality checks, data lineage, Flyway migrations
Visualization and Reporting: ECharts, Plotly, interactive dashboards, KPI reporting, CSV/Excel/HTML report exports
Backend and Tools: FastAPI, Spring Boot, REST APIs, React, Docker, GitHub Actions (CI), Azure, Git, Linux, pytest
Competitive Programming: CodeChef 2063, AtCoder 1159, LeetCode 125+ problems
Spoken Languages: English (C2, EF SET 71/100), Japanese (JLPT N3 level)

Certifications

Databases and SQL for Data Science with Python (IBM), *Nov 8, 2024* • Algorithmic Toolbox (UC San Diego), *Mar 18, 2025*
EF SET English Certificate (C2), *Dec 18, 2025* • Claude with the Anthropic API, *Jun 14, 2026*
Getting Started with Accelerated Computing in Modern CUDA C++ (NVIDIA), *Nov 20, 2025*
JavaScript (HackerRank), *Jan 12, 2026*
Goldman Sachs Software Engineering Virtual Experience Program (Forage job simulation), *Aug 13, 2023*