

# Harshwardhan Patil

Data Analyst | AI & Analytics Engineer

9881771444 | harshkvpatil@gmail.com | Bengaluru, India | linkedin.com/in/hpatil10

## SUMMARY

Data Analyst with 5+ years across banking, retail, and supply chain, using SQL, Python, and ETL automation to find what the numbers are actually saying and get it in front of the people deciding. Cut fraud review time 30%, raised check deposit completion 12%, and held four audit cycles at zero reportable findings. Fluent in Tableau, Power BI, star schema modeling, A/B testing, and regression, now extending that work into LLM and agentic analytics.

## EDUCATION & CREDENTIALS

**M.S. Information Systems**, *Pennsylvania State University, USA*

**May 2025**

**B.Tech, Electronics and Communication**, *SRM Institute of Science and Technology, India*

**Jun 2021**

## SKILLS

**AI & LLM Engineering:** LLMs, Generative AI, Multi-Agent Orchestration, Agentic Workflows, RAG, Prompt Engineering, Semantic Search, Vector Databases (FAISS, Chroma, pgvector), LangChain / LangGraph, OpenAI & Anthropic APIs, LLM Evaluation (LLM-as-judge), Function Calling, JSON-Schema Outputs

**AWS & Cloud:** EC2, S3, Lambda, IAM, VPC, Redshift, RDS, CloudWatch, CloudFormation, Auto Scaling, ELB, Route 53, Well-Architected Framework; Azure; Docker, Kubernetes, FastAPI, CI/CD, Git, Databricks

**Data & Analytics:** SQL (SQL Server, PostgreSQL, MySQL), Snowflake, Python (Pandas, NumPy, SciPy), R, Alteryx, Talend, REST APIs, Star-Schema Modeling, Query Optimization, Data Quality & Governance, PII Protection

**Statistics & ML:** A/B Testing, Hypothesis Testing, Regression, Classification, K-Means, Time Series (LSTM, ARIMA), Random Forest, XGBoost, Scikit-learn, TensorFlow, Bootstrap Resampling, Feature Engineering, NLP

**BI & Delivery:** Tableau, Power BI, Advanced Excel, Streamlit, Plotly, KPI Dashboards, Executive Reporting, Stakeholder Management, Agile / SDLC

## PROFESSIONAL EXPERIENCE

**Data Analyst | PNC, USA**

**Jan 2025 – Sep 2026**

- Analyzed 4.2M transactions in SQL Server and Python to identify 11 recurring fraud anomaly patterns and wrote them into a reviewer rulebook that reduced manual review time by 30 percent, or about 15 hours per week.
- Managed regulatory reporting for six filings per quarter and rebuilt the Alteryx ETL workflow with validation checks that caught over 40 data errors before submission, resulting in zero reportable findings across four audit cycles.
- Built a Tableau dashboard on PostgreSQL data that traced check deposit abandonment to the ID verification step and recommended a fix to nine product and operations stakeholders that raised completion by 12 percent, worth about \$50K in monthly revenue.
- Added automated data quality checks across 18 Azure pipelines to catch schema and volume issues before they reached 12 executive dashboards, reducing downstream reporting errors by 40 percent.
- Tested whether mobile deposit users churned at a higher rate than branch users using Snowflake data, and presented the results to seven senior stakeholders, supporting a \$2M product roadmap decision that improved satisfaction scores by 18 points.

**Research Data Analyst | Pennsylvania State University, USA**

**Aug 2023 – May 2025**

- Managed analytics for three faculty research streams on Industry 4.0 and supply chain resilience, reviewing over 200 papers and 14 executive interviews that contributed to two publications.
- Built a Python pipeline using LLM summarization, keyword mining, and thematic clustering across more than 200 papers, reducing literature review time from six weeks to five days.
- Compared LSTM, ARIMA, ANN, SVM, Random Forest, and XGBoost forecasting models across 24 demand series, and documented LSTM as the most accurate at 8.4 percent MAPE, 3.1 points better than the ARIMA baseline.
- Converted over 60 hours of interview transcripts and qualitative notes into coded datasets and evidence matrices, reducing analysis time per study by 45 percent and producing material used in two published papers.
- Grouped challenges raised across 14 executive interviews at Amazon, Nestlé, and Loblaw into five analytics themes that formed the disruption response framework in the Industry 4.0 survivability paper.

**Data Analyst | Tata Consultancy Services, India**

**Jun 2021 – Jun 2023**

- Rebuilt 12 ETL jobs in Alteryx and Python, reducing nightly data processing from six hours to 45 minutes, holding pipeline uptime at 99.5 percent, and removing about 10 hours of manual data preparation each week.
- Segmented over 100K customers into five behavioral groups using K-means in Scikit-learn and supported targeted campaigns per group, tested against a 10K customer holdout, which raised email conversion from 3 percent to 12 percent over four months.
- Migrated 8 TB of legacy warehouse data to AWS S3 and Redshift, reducing average query response time from 45 seconds to six seconds and monthly infrastructure spend by 22 percent.
- Created a Power BI dashboard tracking 15 KPIs that identified four underperforming sales regions, which management used to reallocate \$400K in marketing budget across three client accounts and increase quarterly revenue by 25 percent.
- Tested nine pricing and campaign proposals using A/B tests and regression analysis in R, and recommended dropping the three that showed no measurable improvement, saving an estimated \$150K.

## AI PROJECTS

**Multi-Agent AI Orchestration System** — Python, LangGraph, OpenAI / Anthropic APIs, PostgreSQL, pgvector, FastAPI, Docker

- Built a multi-agent system that compiles free-text scenarios into typed causal graphs via a draft, critique, and repair LLM pipeline, grounded by time-locked retrieval over a 1.3M document corpus in PostgreSQL and pgvector.
- Engineered a deterministic 24-step simulation kernel with seeded runs and record-replay caching for bit-exact reproducibility and isolated each component's contribution through a 12-cell ablation with bootstrap testing.

**Metric Movement Diagnostics Platform** — Python, PostgreSQL, Claude API, Statistical Bootstrap, Streamlit

- Built an AI system that explains why a business metric moved, pairing a PostgreSQL warehouse of 112K+ transactions with a 2,000-iteration bootstrap engine and Claude-generated explanations, validated on a +52.1% revenue spike.
- Designed a three-agent pipeline (Investigator, Analyst, Critic) in which a fact-checking agent verifies every claim against source data, shipped as a Streamlit app with 63 automated tests.

## CERTIFICATIONS & PUBLICATIONS

**Certifications:** AWS Certified Solutions Architect Associate | AWS Certified Generative AI Developer Professional | Databrick Certified Data Engineer Associate | IBM Data Science Professional

**Publications:** Survivability of Supply Chains in the Era of Industry 4.0; LSTM Neural Network for Stock Market Prediction