

Nitish Rohilla

Data Analyst / Data Scientist

Gurugram, Haryana, IN | +91 72063 68408 | nitishrohilla77624@gmail.com

[Portfolio](#) | [GitHub](#)

EXPERIENCE

Business Analyst — PolicyBazaar.com

April 2026 – Present

Dashboarding, analytics, and reporting automation for the DocPrime health-tech platform

- **User & engagement analytics:** Built and maintain a suite of user & engagement dashboards for the DocPrime app in AWS QuickSight, writing and optimizing the underlying SQL in Athena — covering 400,000+ employee/user records with monthly signup trends, and a 10M+ row clickstream dataset tracking page-level activity, feature usage, and user flow across the app to surface which sections drove the most engagement.
- **API monitoring dashboards:** Built API health dashboards tracking daily errors on the ECards API and other core booking APIs (active-lives fetch, endorsements), breaking failures down by insurer–TPA combination to help the tech team pinpoint recurring failure sources faster.
- **Text-to-dashboard AI agent:** Designed and shipped a natural-language query bot (Python, LangChain, GPT-4o mini) that runs semantic search over the schema/table metadata in a vector DB, generates the SQL, executes it via PySpark, and renders the results as interactive web dashboards — cutting manual reporting workload by 80–90%.
- **Reporting automation:** Automated a QuickSight report that had no native distribution option: a Python script parses the emailed report via BeautifulSoup, generates a personalized AI summary per recipient (Gemini), and pushes both to a stakeholder WhatsApp group via PyWhatKit — full run completes in under 1 minute, replacing a manual, ad-hoc distribution step.
- **RM performance dashboard:** Built a Relationship Manager performance dashboard (200K+ records) tracking onboarding completions, pending vs. completed renewals, and a renewal-likelihood model built on 10,000+ customer remarks, alongside per-RM ticket assignment, resolution, and monthly resolution-rate metrics.
- **Call analytics dashboard:** Built a call-logging dashboard tracking inbound/outbound volumes and connectivity % per RM on a monthly basis, giving managers a single view of call activity alongside the performance dashboard above.
- **Dashboard UX:** Introduced interactive tooltips on compressed/aggregated visuals to surface underlying record-level detail on hover — a QuickSight capability not previously used on the team — removing the need to query the database for specifics, and redesigned the visual language across the team's dashboards (chart selection, layout, parameter-driven interactivity) to make them faster to read for non-technical stakeholders.

Freelance Data Scientist — SkillUpTech

March 2025 – March 2026

End-to-end AI/ML solutions across NLP, LLM fine-tuning, and intelligent automation

- **n8n AI agent automation:** Architected a fully autonomous Password Reset Agent in n8n with secure OTP generation, automated email dispatch, and token-based authentication — designed with a deterministic, auditable backend rather than relying on the LLM for control logic.
- **GCP Vertex AI — Gemini fine-tuning & RAG agent:** Fine-tuned a Gemini model on Vertex AI for domain-specific tasks and built a Retrieval-Augmented Generation agent with embedding models and vector search, reducing hallucination rate versus the ungrounded baseline.
- **AWS Cloud AI (SageMaker, Lambda):** Built spaCy model training and deployment pipelines on SageMaker, and serverless inference endpoints on Lambda for cost-efficient, auto-scaling deployment.

EDUCATION

B.Tech, Computer Engineering with Data Science — CGPA 7.4

J.C. Bose University (YMCA), Faridabad, Haryana, India | 2020 – 2024

SKILLS

Languages: Python, SQL, C++, JavaScript

ML / AI: PyTorch, HuggingFace (PEFT, TRL), LoRA/QLoRA, RAG, LangChain, ChromaDB, spaCy, NLP

Cloud & Platforms: AWS (Athena, QuickSight, SageMaker, Lambda), GCP (Vertex AI), Ollama, n8n

Data Tools: PySpark, Pandas, Matplotlib, NLTK, Excel, SQL, DataFrames

Automation: BeautifulSoup, PyWhatKit, Gemini API

PROJECTS

QWEN 2.5 LLM FINE-TUNING WITH QLoRA | [Link](#)

Python, PyTorch, HuggingFace, QLoRA, PEFT, TRL

- Fine-tuned a Qwen 2.5 0.5B instruction model on local hardware using QLoRA (4-bit quantization + LoRA adapters via PEFT/TRL) on the DialogSum conversational summarization dataset. Achieved a +5% gain in BERTScore F1 (0.8381 → 0.8805) after fine-tuning, with corresponding precision and recall improvements, validated via automated BERTScore evaluation.

LOCAL RAG CHATBOT | [Link](#)

Python, Ollama, ChromaDB, RAG, Retrieval Systems

- Built a fully local, retrieval-augmented chatbot serving Gemma 4 (E2B) via Ollama, with a hand-built retrieval pipeline (custom chunking, nomic-embed-text embeddings, ChromaDB vector store) implemented without a RAG framework for full transparency and control. Verified correct retrieval ranking against source documents and confirmed the system declines out-of-scope questions rather than hallucinating. Supports multi-format ingestion (PDF/Markdown/text) and multi-turn conversational memory, running entirely offline.