

SUYASH

Data Scientist • AI Architect

[LinkedIn](#) • [GitHub](#) • justsuyash.com

suyashkumarthakur@gmail.com • +1(469) 428-6430 • Dallas, Texas, USA

Data Scientist & AI Architect with 8+ years delivering \$500M+ in validated business impact. Specializes in developing novel data science solutions—taking concepts from 0 to 1. Rapidly prototypes using emerging tech, aligns cross-functional teams, and productionizes the solution at scale when needed.

WORK EXPERIENCE

Albertsons - **Senior Data Scientist (Principal Platform Architect)**, Pleasanton, CA Feb 2024 – Present

- Architected enterprise experimentation infrastructure across In-Store, Pharmacy, Digital, and Private Brands, scaling testing capacity from 3 to 10+ experiments per week and supporting \$200M+ in annual business decisions.
- Cut experiment cycle times 40%, from 28 to 12 days, by operationalizing CUPED variance reduction across cross-functional business units.
- Designing a Bayesian evaluation layer (Monte Carlo simulation) on top of CUPED to enable live monitoring and early-stopping of high-impact experiments, further reducing cycle time.
- Deployed a text-to-SQL RAG agent (LangChain, Claude) over a 50TB+ enterprise dataset, making embedded PMs self-serve and reducing post-experiment turnaround from days to minutes for partner teams.
- Rebuilt the pre-split randomization algorithm from the ground up for 30x faster, balanced assignment by minimizing Mahalanobis distance, and quantified the imbalance-versus-power tradeoff so teams can see the statistical power lost at any level of covariate imbalance.
- Designed a human-in-the-loop pipeline where junior data scientists review AI-generated queries and feed corrections back into the RAG layer, improving first-pass validation accuracy 30% and accelerating onboarding.

Applied AI Research Lab - **Principal Engineer & Product Architect**, Remote, US Jan 2023 – Present

- Founded applied AI research and development lab to proactively build and stress-test emerging technologies and extract proven architectures to de-risk enterprise deployments.
- ClinicOS: Built an AI-powered healthcare operations SaaS live in 8 rural Indian clinics, with real-time consultation transcription, prescription generation, appointment workflows, and inventory management.
- QuizBeef: Launched an LLM-powered learning platform (700+ users) that stores questions in Open Knowledge Format (OKF) and generates personalized lessons on the fly, using semantic parsing and FAISS vector retrieval for grounded, adaptive question generation.
- mana-health: Architected a multi-agent wellness platform coordinating daily guidance, food analysis, and biomarker interpretation workflows (LangGraph, CrewAI, DSPy), enforcing typed outputs with PydanticAI and running LangSmith/Braintrust tracing and regression evals via a FastAPI and Next.js product surface.

Discover Financial Services - **Senior Data Science Analyst**, Remote, US Aug 2022 – Oct 2023

- Unlocked \$114M in annualized volume by disproving a decade-old employment-verification policy using XGBoost and propensity score matching, then designing an A/B test that drove a permanent policy change.
- Recovered \$1.8M in loan approvals by diagnosing iOS funnel defects in the Automated Loan Approval platform and partnering with Engineering and UX to restore conversion flow.
- Migrated legacy SAS workflows to GCP and BigQuery, reducing infrastructure cost 12% and improving decision speed by 10+ hours per week.

Vodafone - **Data Scientist**, Pune, IN Jun 2016 – Feb 2019

- Generated \$11.2M in incremental revenue for Vodafone Germany's 45M+ user base through customer segmentation and A/B-tested email personalization, increasing click-through rate 28%, from 1.4% to 1.8%.
- Developed retention and pricing models that improved customer lifetime value 5% and informed retention-based product strategy.
- Reduced Priority-2 production incidents 60%+ by deploying Prophet-based anomaly monitoring across core models, saving approximately 20 engineering hours per week.

Kaizen - **Data Science Engineering Intern**, Remote, US Jun 2021 – Dec 2021

- Improved Toyota vehicle-sales forecast accuracy 8.5% using a hybrid Prophet and LSTM model; reduced two-factor-authentication false positives 13% through an AWS SageMaker anomaly-detection pipeline.

LEADERSHIP

President, Data Science Club - University of Texas at Dallas, TX

Jan 2021 - Jun 2022

- Led 30+ officers and 500+ members as President; ran weekly standups and used member analytics to drive learning and career outcomes.
- Delivered workshops on data science, product analytics, and tooling (Git) to 300+ students; doubled membership through social media and Discord campaigns.
- Organized speaker events with Walmart, Amazon, Google, 2K Games, Sam's Club, and Forbes.

WRITING & SPEAKING

- Author, justsuyash.com — technical essays on experimentation, causal inference, and scaling ML systems to production (e.g., CUPED validation protocols, parallelizing experimentation inference for 75% runtime reduction).
- Attended MIT CODE 2025 (Conference on Digital Experimentation); published a reflection on validating experimental intuition and the future of experimentation.

EDUCATION

MS, Business Analytics - University of Texas at Dallas, TX

Jan 2021 - Jun 2022

- Dean's Scholar
- Top Student Mentor for Data Science track subjects. Answered 800+ questions and coached 70+ students.

BS, Computer Science and Engineering - Pune University, India

Jun 2016

SKILLS

GenAI & Agentic Systems: LangChain, RAG (Retrieval-Augmented Generation), Context Engineering, LLMs, Autonomous Agents, dbt, Snowflake, LangGraph, DSPy, CrewAI, PydanticAI, LangSmith, Braintrust, FastAPI

Causal Inference & Experimentation: Causal Inference, CUPED, Bayesian Methods (PyMC), Propensity Score Matching, Synthetic Controls, Heterogeneous Treatment Effects (HTE), Uplift Modeling, A/B Testing, SRM Detection

Data & Infrastructure: Python, SQL, PySpark, BigQuery, Airflow, Spark, GCP (Vertex AI), AWS (S3, SageMaker), Docker, Pandas, NumPy, Scikit-learn, TensorFlow

Visualization & Tools: Tableau, Streamlit, Git, Azure