

# SUYASH

Senior Data Scientist | Product | Experimentation & GenAI Platform Architect

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

[suyashkumarthakur@gmail.com](mailto:suyashkumarthakur@gmail.com) ♦ [+1\(469\) 428-6430](tel:+1(469)428-6430) ♦ Dallas, Texas, USA

Architecting AI-powered causal experimentation systems at the intersection of Causal Inference and GenAI. Driven \$150M+ in validated business impact by scaling testing velocity from 3 to 8+ experiments/week. Specialized in building distributed causal engines and autonomous GenAI agents that bridge the gap between complex data and actionable product strategy

## WORK EXPERIENCE

Albertsons

Feb 2024 – Present

### Senior Data Scientist

Pleasanton, CA

- **Architected End-to-End Causal Engine:** Replaced manual workflows with a proprietary experimentation platform, scaling velocity from 3 to 8+ experiments/week and driving \$150M+ in annual validated decisions.
- **Engineered Advanced Statistical Pipeline:** Integrated Spark pipelines with CUPEd variance reduction (35% error reduction) and Bayesian inference, compressing test cycles from 28 to 12 days.
- **Deployed GenAI Analyst Agent:** Designed a SQL+RAG system using LangChain and Gemini that translates natural language queries into verified SQL; reduced post-experiment analysis time from days to hours.
- **Operationalized Context Engineering:** Developed error-logging and context protocols for LLM agents, reducing hallucinations by 60% and saving stakeholders 8+ hours/week.
- **Automated Results Summarization:** Built a Beta Auto-Summary Agent that autonomously reads experiment outputs and generates stakeholder reports, eliminating the final manual bottleneck in the testing lifecycle.

Applied AI Research Lab

Jan 2023 – Present

### Principal Engineer & Product Builder

Remote, US

- **Established Technical Sandbox:** Founded an applied research lab to validate emerging Agentic and RAG architectures before enterprise deployment.
- **QuizBeef (700+ Users):** Architected an AI quiz app utilizing LLM semantic parsing and FAISS vector retrieval for document comprehension; insights informed Albertsons' GenAI Copilot patterns.
- **mana-health:** Developed a multi-agent health platform orchestrating specialized agents (physio, nutritionist) to process biometric and workout data.
- **Aligned:** Built a dating app integrating astrology compatibility after market analysis revealed an underserved segment in the dating space; demonstrates product-market fit analysis and consumer behavior insights.

Discover Financial Services

Aug 2022 – Oct 2023

### Senior Data Scientist

Chicago, IL

- **Overturned decade-old risk policy via causal inference:** Used XGBoost + Propensity Score Matching to prove employment verification had zero predictive value for high-FICO borrowers ( $R^2 < 0.01$ ), then designed surgical A/B test that unlocked \$114M annualized volume and 300+ incremental customers/month through friction reduction.
- **Enabled \$1.8M in loan approvals:** Identified and resolved critical iOS bugs through funnel analysis on the Automated Loan Approval platform, collaborating with engineering and UI/UX teams to restore conversion flow.
- **Infrastructure Modernization:** Migrated legacy SAS workflows to GCP BigQuery, reducing infrastructure costs by 12% and improving decision speed by 10+ hours/week.

Vodafone

Jun 2016 – Feb 2019

### Assistant Manager, Data Science

Pune, IN

- Increased marketing revenue by \$11.2M and improved CTR by 28% by orchestrating email personalization using logistic regression to optimize content targeting, then A/B testing new campaign strategies.
- Enhanced customer lifetime value by 5% by developing Linear Regression models to optimize subscription pricing based on user data and predicted retention patterns across customer segments.

### Senior Executive, Data Science

- Enabled real-time performance tracking across 10+ models by partnering with product teams to define and instrument 15+ custom metrics, building SQL pipelines and executive dashboards for new feature evaluation.
- Reduced Priority 2 production issues by 60%+ by developing Tableau dashboards with time series forecasting alerts, enabling proactive incident detection and faster resolution.

Kaizen	Jun 2021 – Dec 2021
Data Science Engineering Intern	Atlanta, GA
<ul style="list-style-type: none"> <li>▪ Informed 3+ marketing campaigns by analyzing user clustering patterns to uncover distinct customer journey types, translating data insights into actionable segmentation strategies.</li> <li>▪ Reduced authentication false positives by 13% by building ETL pipelines and deploying an anomaly detection model on AWS SageMaker, adding risk-based flagging to enhance 2FA security.</li> <li>▪ Improved forecasting accuracy by 8.5% by designing a hybrid time-series model integrating Prophet and LSTM for Toyota sales prediction, outperforming baseline statistical models.</li> </ul>	

## EDUCATION

---

The University of Texas at Dallas	Jun 2022
MS, Business Analytics	Dallas, TX
<ul style="list-style-type: none"> <li>□ Dean's Scholar (Top 20 %); President, Data Science Club — led workshops and speaker series.</li> <li>□ Top Student Mentor for Data Science track subjects. Answered 800+ questions and coached 70+ students.</li> </ul>	

Pune University	Jun 2016
BS, Computer Science and Engineering	Pune, IN

## SKILLS

---

**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

**GenAI & Agentic Systems:** LangChain, RAG (Retrieval-Augmented Generation), Context Engineering, LLMs, Prompt Engineering, Autonomous Agents, SQL+RAG

**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

## Certifications

---

[Google Analytics](#)