

## Contact

4694286430 (Home)  
suyashkumarthakur@gmail.com  
[www.linkedin.com/in/justsuyash](https://www.linkedin.com/in/justsuyash)  
(LinkedIn)

## Top Skills

Bayesian statistics  
Google Cloud Platform (GCP)  
A/B Testing

## Certifications

Machine Learning  
DeepLearning.AI Deep Learning Specialization.

## Honors-Awards

Top Performer of the Year  
2017-2018  
Top Mentor

# Suyash .

Senior Data Scientist · Product, Experimentation & AI | \$150M+ Impact | H1B (Approved 2025)  
Dallas, Texas, United States

## Summary

I am a Product Data Scientist focused on building world-class products and driving growth through rigorous experimentation and data-driven decision-making.

My work operates at the intersection of Enterprise Scale and AI Innovation, bridging the gap between traditional causal inference and the emerging field of autonomous context engineering.

**Enterprise Architecture:** At a Fortune 100 retailer, I architected AI experimentation engines and spearheaded the platform development that scaled testing velocity from 3 to 8+ experiments/week. This infrastructure helped drive a cumulative \$150M+ in measurable business impact.

**The AI Research Lab:** Exploring the rapidly evolving landscape of GenAI, I run Bheed Studio, my applied AI research lab. I identify critical market gaps and build 0->1 products from first Principles using autonomous AI architectures in the domains of Health, Social, and EdTech, using both modern orchestration frameworks like LangChain to create RAG pipelines and traditional data science models for compatibility engines, statistical analysis, and metric evaluations.

## What I Learned Building Solo:

**Context Engineering as a Velocity Multiplier:** I learned to ship complex, full-stack products rapidly by designing "Context Systems" that maintain architectural consistency across the LLM. Rapidly iterating on data inputs to observe model responses gave me a deep, practical intuition for how models "think" and adapt.

**Metric Design & Instrumentation:** Building from scratch taught me how to design event schemas that actually matter. I learned to define

"North Star" metrics, implement tracking pipelines from Day 1, and ensure every data point serves a specific business decision.

Product Optimization Loop: I learned how to use early analytics to diagnose user friction and refine the core product loop—turning raw usage logs into a concrete roadmap for retention and growth.

The Goal: I implement these rapid-iteration learnings at an enterprise scale to build and grow 1->N products.

Check out my case studies for market strategy, the unit economics that validate these product concepts, and how I have brought value to organizations I have been a part of.

Website : [justsuyash.com](http://justsuyash.com)

---

## Experience

Albertsons Companies

Senior Data Scientist

February 2024 - Present (2 years)

San Francisco Bay Area

Leading the transformation of Albertsons' A/B testing infrastructure from manual workflows to an AI-driven engine.

-- GenAI Integration in Experimentation: Leading the design and development of an internal GenAI agent integrated with the experimentation platform.

-- Boosted testing velocity from 3 to 8+ experiments per week by re-architecting our experimentation pipeline across ingestion, processing, and visualization layers, eliminating Excel workflows and configurable notebooks that can be run as automated workflows, resulting in very short turnaround time for experiments. This is being used as a baseline for the next-gen engine we are collaborating on.

-- Algorithmic Auditing (The "Presplit" Fix): Uncovered a critical "Presplit Bias" flaw in the randomization engine where control/treatment groups seemed random but yielded a high R2, uncovered the flaw in logic, and implemented a better-suited algorithm that was also 30x faster and very reliable.

**Business Impact:**

- Metric Expansion: Engineered the tracking of 17+ new KPIs (including Account Health & Non-Zero E-commerce metrics) to provide granular visibility into user behavior.
- Automations - Eliminated 8+ Hrs of work by writing code that made Excel work redundant and automated metric ingestion for a 3rd party platform, which significantly reduced turnaround time for in-Store metrics.

Tech Stack: Python, PySpark, GCP (BigQuery, Vertex AI), Gemini API, SQL, Bayesian Statistics.

**Discover Financial Services**

Senior Data Scientist

August 2022 - November 2023 (1 year 4 months)

Chicago, Illinois, United States

Revenue Optimization & Risk Intelligence. Leveraged advanced predictive modeling (XGBoost, Decision Trees) and A/B testing to uncover inefficiencies in the loan origination funnel, directly capturing \$50M+ in realized loan volume.

The "\$114M Friction" Fix (Employment Verification):

The Challenge: The bank assumed "Employment Verification" (EV) reduced risk, but it caused massive drop-offs. My analysis proved EV status had zero correlation with default rates for high-FICO segments.

The Solution: Designed a targeted A/B test to remove this barrier for low-risk applicants.

The Numbers: Converted ~300+ additional customers monthly (Avg Loan: ~\$31k). The initiative captured \$50M+ in realized volume during rollout, with a projected impact of \$114M+ annualized.

Risk Modeling & Process Automation:

High-Risk Segmentation: Built Decision Trees to pinpoint consumer pockets with >4% loss rates. Translated complex model findings into simple rules that Operations teams used to save ~\$2.5M annually in bad debt (preventing ~\$200k/month in losses).

Infrastructure: Modernized the reporting stack by migrating legacy SAS workflows to a modular SQL/Cloud architecture, speeding up executive decision-making by 10+ hours/week.

Kaizen Analytix  
Data Scientist Intern  
June 2021 - November 2021 (6 months)  
Atlanta, Georgia, United States

Advanced Forecasting & Fraud Detection.

Hybrid Architecture (Toyota): Outperformed the production Prophet baseline by 8.5% (RMSE reduction). Architected a hybrid LSTM + Prophet model to capture complex residuals and seasonality in vehicle sales data (see attached architecture).

Fraud Detection System: Built an Anomaly Detection pipeline on AWS SageMaker to flag high-risk transfers. Created a dynamic thresholding system that reduced false positives in 2FA challenges, improving user experience without compromising security.

Vodafone  
2 years 6 months  
Assistant Manager, Data Scientist  
August 2018 - February 2019 (7 months)  
Pune Area, India

Growth Strategy & Predictive Modeling. Spearheaded data-driven marketing initiatives for the German market (45M+ user base), utilizing logistic regression and A/B testing to drive massive revenue uplifts.

\$11.2M Revenue Lift: Orchestrated an email personalization strategy using customer segmentation models. The resulting A/B test improved CTR by 28% (1.4% to 1.8%), directly generating \$11.2M in incremental revenue.

Pricing Optimization: Developed Linear Regression models to predict retention curves across price points. Identified sub-optimal LTV in mid-tier pricing, redirecting marketing spend to high-LTV segments that improved overall Customer Lifetime Value by 5%.

**Strategic Roadmap:** Collaborated with Product leadership to define the roadmap for retention-based pricing, translating complex churn predictions into actionable go-to-market strategies.

#### Senior Executive

January 2017 - July 2018 (1 year 7 months)

Pune Area, India

Operational Intelligence & Reliability Engineering. Focused on stabilizing data infrastructure and democratizing insights for executive teams.

**System Reliability (P2 Reduction):** Built a Time-Series Anomaly Detection system (Prophet) to monitor error rates across 10+ models. Configured real-time alerts that reduced "Priority 2" production incidents by 60%+, saving engineering teams ~20 hours/week.

**Executive Visibility:** Partnered with stakeholders to define 15+ custom performance metrics. Built automated SQL pipelines feeding Tableau dashboards that enabled real-time tracking of model health and business KPIs.

**Automation:** Migrated manual "stuck order" processing to automated Shell/SQL scripts, eliminating 4+ hours of manual toil per week.

#### Engineering Trainee

September 2016 - December 2016 (4 months)

Pune, Maharashtra, India

Data Engineering Foundations. Completed intensive training on the full data stack (Python, SQL, Shell Scripting) and Lean Six Sigma methodologies. Built foundational ETL scripts for internal reporting systems.

---

## Education

### The University of Texas at Dallas

Master of Science - MS, Business Analytics, Data Science, Artificial Intelligence · (2020 - 2022)

### Savitribai Phule Pune University

Bachelor of Engineering - BE, Computer Science