

SUYASH

[LinkedIn](#) | [GitHub](#) | [Tableau](#) | [Website](#) | [Publication](#) | [Statistical Study](#)

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

Senior Data Scientist with 7+ years driving product growth through experimentation and ML at Fortune 500 companies. Built the SAFE platform scaling testing velocity from 3 to 8+ experiments/week and combined product sense, analytical rigor, and ML-driven optimization to deliver \$150M+ in measurable business impact.

WORK EXPERIENCE

Safeway Inc.

Feb 2024 – Present

Data Scientist IV

Pleasanton, CA

- Boosted testing velocity from 3 to 8+ experiments per week by architecting SAFE experimentation pipeline across ingestion, processing, and visualization layers, eliminating Excel workflows.
- Enabled management of 17+ new metrics (non-zero e-commerce, account health KPIs) by orchestrating integration into SAFE pipeline, expanding measurement capabilities.
- Migrated 70% of experimentation codebase from Snowflake to GCP BigQuery while ensuring data integrity through comprehensive QA protocols, supporting increased scale.
- Reduced abandonment rate by 8+% by enhancing opportunity sizing framework with predictive forecasting models and business metrics, enabling prioritization toward high-impact tests.
- Doubled annual testing capacity by pioneering Bayesian experimentation framework that reduced test duration from 8 to ~3 weeks for 20% of experiments.

Data Scientist III

Aug 2022 – Oct 2023

Senior Data Scientist

Chicago, IL

- Generated \$114M in annual loan volume by leveraging XGBoost analysis to prove employment verification added no predictive value, then executing an A/B test on low-risk segments that increased conversions by 300+ customers monthly through friction reduction.
- Enabled \$1.8M in loan approvals by identifying and resolving critical iOS bugs through funnel analysis on the Automated Loan Approval platform, collaborating with engineering and UI/UX teams to restore conversion flow.
- Improved customer segmentation model accuracy by 12% by integrating GCP TensorFlow and collaborating with ML engineering team, enabling more precise risk assessment and targeted marketing strategies.
- Reduced infrastructure costs by 12+% by migrating legacy SAS workflows to modular SQL queries, improving cluster efficiency and data processing performance.

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.
- Increased subscription revenue by 28% by developing retention prediction models and recommending country-specific pricing strategies, collaborating with product teams to inform roadmap and go-to-market planning.

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

- Informed 3+ marketing campaigns by analyzing user clustering patterns to uncover distinct customer journey types, translating data insights into actionable segmentation strategies.
- 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.
- 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.

Awards & Projects

quizBeef – LLM-Powered Quiz App

- Built a TypeScript-based AI quiz app allowing users to upload PDFs and auto-generate Q&A challenges via LLM semantic parsing.
- Integrated vector retrieval and session-based logic using embeddings and FAISS, demonstrating interactive document comprehension.
- Conceptualized short-form video content for Instagram that increased website traffic by 1,800 visits within one month, leading to 700+ new user acquisitions for the quiz application.

Product Data Science via Chat-GPT - Identify product weaknesses

- Leveraged ChatGPT and OpenAI embeddings to analyze negative customer reviews across pet food brands.
- Cosine similarity revealed product weaknesses, benchmarked competitor issues, and visualized key sentiment drivers.
- Actionable insights (e.g., “hard treats” complaints) led to product improvement recommendations.

ChatGPT Q&A Knowledge Retrieval System

- Developed a semantic Q&A engine using OpenAI embeddings to answer data science product questions.
- Implemented document chunking, embedding creation, and cosine similarity ranking for high-precision retrieval.
- Demonstrated context-aware accuracy gains over zero-context prompts for AI-powered knowledge search.
- Showcased that AI can enhance domain-specific knowledge retrieval and interview preparation.

Ads Analysis and Clustering - E-Commerce Analytics (Data Science Club Hackathon 21') (1st place finish)

- Identified the 5 best ad groups using Profits and CTR. Clustering ads into 3 groups, based on Average Cost Per Click.

EDUCATION

The University of Texas at Dallas

Jun 2022

MS, Business Analytics

Dallas, TX

- Dean's Scholar (Top 20 %); President, Data Science Club — led workshops and speaker series.
- Top Student Mentor for Data Science track subjects. Answered 800+ questions and coached 70+ students.

Pune University

Jun 2016

BS, Computer Science and Engineering

Pune, IN

SKILLS

Programming & Data: Python, SQL, PySpark, BigQuery, Airflow, Pandas, NumPy, Scikit-learn, TensorFlow, LangChain

Machine Learning & Product: A/B Testing, Bayesian Methods, Machine Learning, Uplift Modeling, Prompt Engineering.

Infrastructure & Visualization: GCP (BigQuery, Vertex AI), AWS (S3, SageMaker), Azure, Docker, Streamlit, Tableau.

Certifications

[Google Analytics](#)