SIMRAN ABHAY SINHA

sinha.sim@northeastern.edu • +1 (857) 381-5028 • Boston, MA • LinkedIn • GitHub • Tableau • Portfolio

SUMMARY

Data Analytics Engineering grad student with experience in predictive modeling, data pipelines, and AI integration. Proficient in Python, SQL, Power BI, Tableau, and Snowflake. Skilled in building Machine Learning models, dashboards, and cloud-based solutions using AWS and OpenAI tools. Strong communicator with a focus on turning complex data into actionable insights. Passionate about leveraging data to drive innovation in financial services and investment analytics.

EDUCATION

Northeastern University, Boston, USA **May 2026 (Expected)**
Master of Science in Data Analytics Engineering GPA 3.8/4.0
Coursework: Data Management for Analytics, Data Mining, Computation and Visualization, Storytelling with Data, Structured Data Analytics, Deep Learning for AI, Applied Generative AI, Foundation Data Analytics

K. J. Somaiya Institute of Technology, Mumbai, India **June 2023**
Bachelor of Technology in Electronics and Telecommunication GPA 3.7/4.0
Coursework: Database Management System, Neural Networks and Deep Learning, Machine Learning, Big Data Analytics, Project Management

SKILLS

- **Programming Languages:** Python, SQL, R, MATLAB
- **ML Tools:** Scikit-Learn, XGBoost, Logistic Regression, Random Forest
- **Data Engineering:** Snowflake, MySQL, PostgreSQL, MongoDB, ETL, Data Modeling
- **Visualization:** Tableau, Power BI, Flourish, Datawrapper
- **Cloud Technologies:** AWS (S3, IAM, Bedrock, SageMaker, Lambda)
- **AI/LLMs:** Retrieval-Augmented Generation (RAG), OpenAI API, Vector DBs
- **Software and Tools:** Streamlit, GitHub, JIRA, PyCharm, VS Code

WORK EXPERIENCE

Data Science Lead - Crewasis, New York, USA (Remote) **Sept 2025 - Dec 2025**
- Analyzed consumer data across 10+ global projects using Python to identify spending patterns and behavioral trends supporting strategies
- Performed data cleaning, exploratory analysis, and visualization with SQL, and Power BI, enabling faster identification of key customer insights
- Built and maintained interactive dashboards on the Portal using Power BI, improving stakeholder reporting efficiency and project transparency
- Applied text analytics and structured literature review techniques to integrate qualitative and quantitative insights for improved decisions
- Collaborated with cross-functional teams to translate analytical findings into actionable business recommendations

Data Science Intern - Prodigy, Mumbai, India **Jun 2024 - Aug 2024**
- Developed machine learning models using Python and Scikit-Learn to predict customer purchases, increasing marketing ROI by 25%
- Cleaned, transformed, and validated customer datasets using Python, improving data consistency and model accuracy across workflows
- Designed interactive Power BI dashboards to visualize customer segments and KPIs, driving a 20% increase in campaign engagement
- Documented standardized data workflows and pipelines, ensuring reproducibility, scalability, and efficient knowledge transfer
- Assisted in model evaluation and feature selection to enhance prediction accuracy, model reliability, and overall business relevance

Data Science Intern- Champion Learns, Mumbai, India **Jan 2023 - Dec 2023**
- Optimized client targeting strategies using SQL and Python for campaign analysis, increasing qualified lead conversion by 25%
- Segmented and clustered customers using Tableau visualizations, improving marketing personalization and engagement by 30%
- Automated end-to-end data collection, cleaning, and reporting pipelines, reducing manual effort by 30% and accelerating insight delivery
- Presented data-driven insights and dashboards to management, supporting strategic marketing decisions and future campaign planning
- Monitored campaign performance metrics to identify optimization opportunities, emerging trends, and areas for continuous improvement

PROJECTS

Predicting Diabetes Risk using Health Indicator | *Python, Scikit-Learn, Logistic Regression, Random Forest, XGBoost* [Link]
- Developed classification models to predict diabetes risk, improving early diagnosis accuracy by 18% across health and lifestyle datasets
- Built full machine learning pipeline with preprocessing, tuning, and evaluation to ensure reliable, scalable prediction performance
- Engineered clinical features such as BMI, glucose, and age to boost model interpretability and reduce diagnostic ambiguity
- Deployed model using Streamlit for interactive testing and demo access by healthcare teams and technical reviewers

Global Consumer Spending Habits|*Power BI (DAX , KPI)* [Link]
- Created interactive dashboards with DAX formulas to reveal patterns in regional consumer behavior and spending by demographics
- Implemented real-time KPIs, filters, and slicers to enable decision-makers to track trends and adjust strategies dynamically
- Built scalable pipelines with Power BI dataflows to automate reporting and maintain consistent access to refreshed insights
- Analyzed behavioral drivers across markets to help identify high-growth opportunities and guide executive market entry decisions

Vital Sign Manager (Hospital Management System) |*SQL , Python , NOSQL (MongoDB)* [Link]
- Built a hybrid SQL and NoSQL database to manage 100+ patient records with fast, secure, and scalable access architecture
- Implemented role-based user access and authentication to reduce data latency and ensure compliance with healthcare standards
- Improved retrieval and display of patient vitals through optimized queries and structured schema designs for clinical use
- Enabled concurrent user access and audit tracking to support coordination among doctors, nurses, and admin personnel