HARI GEBRE

DATA SCIENTIST

Boston, MA | gebrehari01@gmail.com | 617-410-8617 | LinkedIn

PROFESSIONAL SUMMARY

Data Scientist with 8+ years of experience in developing analytical solutions using Python, SQL, machine learning, statistical analysis, and predictive modeling. Skilled in analyzing large-scale datasets, developing predictive models, and applying analytical techniques to financial services and customer behavior problems. Proficient in Scikit-learn, XGBoost, PySpark, TensorFlow, PyTorch, and modern AI technologies, with experience translating analytical findings into clear business insights.

TECHNICAL SKILLS

Programming & Data Analysis: Python, SQL, Pandas, NumPy, PySpark, Apache Spark, Statistical Analysis, A/B Testing, Hypothesis Testing

Machine Learning: Scikit-learn, XGBoost, CatBoost, Random Forest, Logistic Regression, Gradient Boosting, K-Means, Predictive Modeling, Feature Engineering, Time-Series Forecasting

Deep Learning & NLP: TensorFlow, PyTorch, Keras, CNN, RNN, GNN, Transformers, Hugging Face

Generative AI & LLM: RAG, Prompt Engineering, Fine-Tuning, AI Agents, Embeddings, Semantic Search, LangChain, LangGraph, Pinecone, OpenAI APIs, LLM Evaluation

Data Analytics & Visualization: Power BI, Tableau, Excel, Power Query, Matplotlib, Seaborn, Plotly, KPI Reporting, Data Validation

Big Data & Data Engineering: Apache Spark, Apache Airflow, Kafka, ETL/ELT, dbt, Data Warehousing, Snowflake

Cloud & Deployment: AWS S3, AWS EC2, AWS Lambda, Docker, Kubernetes, CI/CD, MLflow

Databases: PostgreSQL, Microsoft SQL Server, MySQL, Oracle, Snowflake

PROFESSIONAL EXPERIENCE

PNC - Data Scientist **Jan 2025 - Present**

Banking Customer Risk & Transaction Intelligence | Boston, MA

- Developed XGBoost models in Python to analyze 250K+ banking transactions, identifying unusual transaction behavior and customer activity patterns for risk assessment.
- Used SQL to analyze account activity, transaction amounts, frequency, and customer behavior, creating 10+ behavioral features for downstream risk analysis.
- Processed 500K+ financial records with PySpark, performing data cleansing, transformation, and aggregation to prepare reliable datasets for machine learning workflows.
- Performed feature engineering and variable selection across 15+ candidate attributes, improving the structure of transaction and customer datasets used for predictive modeling.
- Evaluated 5+ model iterations through hyperparameter tuning and cross-validation, comparing classification performance and selecting consistent approaches for customer and transaction risk analysis.
- Used MLflow to track 8+ model experiments, maintaining model parameters, evaluation results, and artifacts to support reproducible analytical development.
- Validated model outputs against business rules and historical patterns, investigating unexpected predictions and working with data teams to resolve input-data inconsistencies.
- Developed Power BI reporting covering 12+ financial and risk KPIs, helping business teams monitor transaction activity, customer behavior, and portfolio trends.
- Automated recurring data preparation and validation steps using Python and SQL, reducing repetitive manual checks before datasets were passed into modeling and reporting workflows.
- Presented model findings and customer behavior trends to business stakeholders, translating analytical results into insights used during monthly risk and transaction reviews.

Synchrony - Data Scientist **Jun 2022 - Dec 2024**

Consumer Credit & Customer Segmentation Analytics | USA

- Analyzed 200K+ consumer account records using Python and SQL to identify customer behavior patterns across credit and financing portfolios.
- Created 10+ behavioral features around spending, transaction activity, account usage, and engagement to support customer segmentation and portfolio analysis.

- Applied K-Means clustering to develop 4 customer segments based on behavioral patterns, helping analytics teams compare differences in consumer activity.
- Developed Logistic Regression models across 4+ iterations to analyze customer payment behavior and evaluate factors associated with different customer outcomes.
- Performed hypothesis testing and A/B analysis across 2 customer groups, providing statistical evidence for differences in customer response and behavior.
- Built ARIMA forecasting models using 18+ months of historical data to analyze transaction-volume and customer-activity trends for portfolio planning.
- Developed Tableau dashboards covering 10+ portfolio KPIs, combining customer segments, account activity, transaction trends, and portfolio measures.
- Partnered with business and analytics teams across multiple analytical requests, converting consumer-finance questions into datasets, statistical analyses, predictive models, and reporting.

DICEUS - Data Analyst **Jan 2021 - May 2022**

Client Data Analytics & Predictive Modeling | USA

- Reviewed and standardized 10K+ client records by identifying duplicate entries, incomplete information, and inconsistent data to improve overall data quality by 8%.
- Explored 5+ client-related variables to summarize common trends, unusual observations, and basic relationships that supported initial business analysis.
- Compared client characteristics and historical patterns to identify meaningful differences across customer groups and support ongoing analytical reviews.
- Organized and categorized client information into 3+ analytical groups, making information easier to review and helping teams identify relevant business patterns.
- Reviewed and validated client information across multiple sources, identifying missing or inconsistent records and preparing 2+ analysis-ready datasets for reporting and business review.
- Created reporting views covering 4+ business KPIs, helping senior team members monitor customer trends and reducing manual reporting effort by approximately 7%.
- Reviewed analytical findings and summarized key observations, trends, and customer patterns for discussion with senior data scientists and business stakeholders.
- Supported 2+ project iterations by working with senior data scientists and client stakeholders to clarify requirements, review findings, validate information, and incorporate feedback into analytical deliverables.

Allstate - Business Analyst **Jun 2018 - Dec 2020**

Insurance Operations & Customer Analytics | Lowell, MA

- Assisted with gathering and documenting requirements for insurance policy and customer service processes, supporting senior analysts during requirement discussions with business users.
- Used Excel and SQL to review 10K+ policy and customer records, identifying missing information, duplicate entries, and basic data inconsistencies.
- Prepared recurring reports covering policy activity, customer requests, and operational KPIs, helping business teams monitor day-to-day insurance operations.
- Supported UAT by preparing test cases, validating application results against requirements, and documenting 10+ defects for review by senior analysts and development teams.
- Maintained BRD/FRD documentation, process flows, and requirement trackers, coordinating with business users, QA, and IT teams throughout application enhancements.

EDUCATION

Computer Information System **Jan 2015 - May 2018**

Metropolitan State University, CO