

Hiba Khan

hiba56001@gmail.com | 919-819-7308 | [LinkedIn](#) | [GitHub](#)

EDUCATION

North Carolina State University, Raleigh, NC

January 2024 – June 2026

Bachelor of Science in Statistics, Minor in Biology

WORK EXPERIENCE

Data Automation and Analytics Intern, Greentree Capital, Chicago, IL

May 2025 – July 2025

- Improved performance monitoring and supported business decision-making by 45% by developing interactive Power BI dashboards and visual charts that let stakeholders track trends and act on emerging issues.
- Improved data accuracy by 75% and cut manual data entry by building automated Power Automate pipelines that reduced manual effort by collecting survey responses and Microsoft Power Apps data into SharePoint Lists and Excel for workforce analytics.
- Enabled consistent quality metrics reporting by defining and documenting metric calculations and recurring reporting processes for handoff to business users.
- Validated metric accuracy by cleaning and prepping raw data in Excel and Python, removing duplicate entries, correcting errors, and handling missing values before reporting.
- Collaborated with cross-functional teams to define reporting requirements, then communicated analytical findings to technical and non-technical stakeholders through dashboards and written summaries.

Merchandise Associate, HomeGoods, Apex, NC

August 2023 – December 2024

- Informed restocking decisions by using Excel to organize and analyze sales and inventory data, identifying trends in product movement and generating recurring reports.

PROJECTS

First Tee Triangle Donor Analytics (Client Engagement)

- Surfaced a funding dependency risk, with the top 10% of donors giving roughly 89% of dollars, for First Tee Triangle, a Raleigh youth development nonprofit, by building a donor-level data model from raw gift records using Python.
- Identified trends and risks for strategic planning by flagging donors at risk of lapsing, and delivered an end-to-end analysis report with interactive dashboards covering donor retention, concentration risk, and campaign performance.

Healthcare Claims Analytics

- Prioritized where intervention has the most impact by finding that 2 of 15 care types drive half of diagnosed spend and 86 of 3,918 facilities carry half of all spending, by writing SQL queries with multi-table joins, aggregations, and window functions against five years of claims data (\$99.1B billed) in a Snowflake cloud data warehouse.
- Detected and drove to closure five data defects, including one that overstated a key figure by roughly 543,000x, by investigating discrepancies, documenting root cause and the corrective SQL logic, and validating the fix.
- Kept 47 AI-assisted analysis queries reliable by specifying the plan and constraints up front, reviewing every output, validating against independent small-sample tests, and running a second AI agent as an additional review pass.

Clinical Risk Prediction Engine

- Built an early-warning predictive model in R to flag high-risk cases for intervention before failure; benchmarked Random Forest against Logistic Regression, Lasso, and XGBoost and selected on test PR-AUC of 0.876.
- Evaluated model performance across gender and race subgroups and tuned decision thresholds within each subgroup to keep predictions equitable across populations.

Early-Stage Diabetes Risk Detection Model

- Identified 81% of at-risk individuals as candidates for early screening outreach by building a threshold-tuned LightGBM model in Python on CDC survey data.
- Used SHAP to explain key risk drivers such as BMI, blood pressure, and income level, turning model output into factors public health teams can act on.

SKILLS

Languages: Python, SQL, R, SAS

Data and Visualization Tools: Power BI, Snowflake, Microsoft Excel, Power Automate, SharePoint

Statistics and Modeling: Descriptive Statistics, Hypothesis Testing, Regression, Bayesian Methods, Machine Learning

Data Quality and Documentation: Metric Definition, Data Validation and Discrepancy Checks, Technical Documentation