

Mariam Khan

Old Bridge, NJ | mariamkhan8805@gmail.com | 732-520-9055 | [LinkedIn](#) | [Portfolio](#)

SUMMARY

Recent Computer Science & Data Science graduate (Cum Laude, Rutgers University) who turns large, messy datasets into data-driven insights that stakeholders act on. Experience writing SQL and Python to query and clean data, running hypothesis and A/B tests to validate findings, and building interactive dashboards that communicate KPIs to cross-functional partners. Comfortable advising product, engineering, and business teams on improvement opportunities backed by data.

SKILLS

Data & Analysis: SQL & Relational Databases · Python (Pandas, NumPy, SciPy, Scikit-learn) · R · Statistical Analysis (Hypothesis Testing, A/B Testing, Regression, Cross-Validation) · ETL & Data Pipelines

Reporting & Tools: Tableau · Data Visualization & Dashboarding · KPI Reporting · Excel · Git/GitHub & Code Review

Also Familiar With: Cloud data platforms (AWS) from coursework and independent work.

EXPERIENCE

Product Specialist, Apple

July 2026 - Present

Freehold, NJ

- Diagnosed product and software issues for 100+ customers weekly, identifying recurring pain points and building structured troubleshooting workflows to improve resolution speed and accuracy.
- Connected 20+ local business owners to Apple Business resources based on individual needs assessments, reducing repeat support inquiries by 15%.

AI Automation Extern, Pfizer

June 2026 - Aug 2026

Remote

- Automated a manual, hours-long document search process by designing and building a RAG-based document extraction pipeline (OCR, open-source LLMs, vector retrieval), projected to reduce manual search time by 90%+.
- Evaluated multiple OCR and RAG approaches through systematic benchmarking, analyzing accuracy and performance trade-offs to select the optimal architecture, validated via an interactive demo.

AI Data Analyst, Model Evaluation (Contract), Browserbase

Oct 2025 - Nov 2025

Remote

- Evaluated 1,800+ LLM outputs across 6 models using Stagehand, labeling and classifying task outcomes as correct/incorrect based on accuracy, reasoning quality, instruction adherence, and safety compliance.
- Analyzed hallucinations and reasoning failures across complex, multi-step task environments, delivering structured feedback and evaluation data that informed model calibration and deployment readiness in workflows referenced by Anthropic, Google, and Microsoft.

Lead Process Analysis Extern, Harmony Biosciences (Rutgers MBS Externship Program)

Sep 2024 - Dec 2024

New Brunswick, NJ

- Led a cross-functional team of 5 to analyze a pharmaceutical client's contracting and procure-to-pay workflow, writing SQL queries to pull and review 50+ edge cases in the underlying records.
- Identified root causes from the query results and translated findings into data-driven process improvements and test scenarios that strengthened data accuracy.

PROJECTS

RxWatch: Drug Shortage Intelligence Platform | Independent Research | [GitHub](#)

August 2026 - Present

- Built a drug shortage risk-scoring dashboard after identifying a gap in pharmacy-facing shortage tools, cleaning/joining 36,000+ records across three FDA datasets using Python (Pandas); diagnosed and fixed a pagination bug that raised match rate from 15% to 90%, then validated a supply- vs. demand-side classification model via 5-fold cross-validation (AUC 0.81).
- Applied hypothesis testing (Mann-Whitney U, $p < 0.0001$) to uncover that manufacturer count significantly differs by shortage cause, then built an interactive dashboard letting pharmacies upload their inventory and receive per-drug risk scores.

Sales Performance Dashboard | Tableau Project | [View Project](#)

May 2026

- Built an interactive Tableau dashboard joining 4 relational tables (~10K transactions), engineering 9 calculated fields to surface a 5-point gap between 2023 sales growth (20%) and profit growth (14%), and identify 4 sub-categories at a net loss, including a -13% margin sub-category driving over \$8K in losses.

Early Alzheimer's Detection | [GitHub](#)

May 2025

- Achieved 83% accuracy using Python (NumPy, Pandas, Scikit-learn) on a 174-subject clinical trial dataset, engineering a pipeline to preprocess 25 multi-modal handwriting trials and concluding Random Forest outperformed Naive Bayes, Logistic Regression, and SVM.

LEADERSHIP

Co-founder & Marketing Director, Tech Collaborative

July 2025 - May 2026

New Brunswick, NJ

- Co-founded a tech initiative connecting students and mentors, raised \$2k+ in sponsorships, grew social media to 190k+ interactions, and attracted 70+ attendees per event, making it one of Rutgers' most active tech organizations.

EDUCATION

Rutgers University, New Brunswick, NJ

Sep 2023 - May 2026

Bachelor of Science (Cum Laude) in Computer Science and Data Science | Relevant Coursework: Machine Learning, Data Management, Regression Methods, Statistics