

SUMMARY

Data Engineer & Analyst (Python/SQL, AWS/Docker) building pipelines and dashboards for high-availability systems. Architected ingestion layers ensuring 100% data integrity and optimized ETL workflows that accelerated backend retrieval by 25%. Engineered scraping pipelines that slashed processing by 40%. [Microsoft Certified: Power BI Data Analyst Associate](#).

EDUCATION

M.S. in Data Science | Indiana University, GPA: 3.55/4 | Bloomington, IN

Aug 2024 – May 2026

B.Tech. in Computer Science | Manipal Academy of Higher Education, GPA: 8.4/10 | Dubai, UA

Aug 2019 – Jun 2023

TECH STACK

Languages: Python, SQL, R, HTML, CSS, JavaScript.
Data Visualization & Analytics: Power BI, Microsoft Excel (Advanced), Matplotlib, Seaborn, PyQt5 (Interactive Dashboards), Pandas, NumPy, Tableau.
Machine Learning & AI: Scikit-Learn, GenAI, SHAP (Explainable AI), NLP, Predictive Modeling.
Cloud, DevOps & Tools: AWS (EC2, App Runner, ECR), Docker, Linux, Git, Agile/Scrum, Fusion 360

EXPERIENCE

Software Engineer Intern (Data) | Ikabit Inc. | Folsom, CA | Remote

Jun 2025 – Aug 2025

- Ensured 100% data integrity and accelerated backend retrieval by 25% by engineering a Python UDP pipeline, optimizing SQL queries, and containerizing ETL workflows with Docker and AWS.
- Standardized data visualization layers and enabled real-time monitoring by transforming 50+ HTML/CSS/JS templates into dynamic PyQt5 layouts to create a unified Data Management Console.
- Reduced the development cycle by 20% and minimized sync errors by 90% by leveraging GenAI tools for logic generation and Git to version-control integrated Fusion 360 hardware specs.

Data Analytics Intern | Emirates Steel Arkan | Abu Dhabi, UAE | On-site

Mar 2024 – Jul 2024

- Achieved a 20% increase in data precision to facilitate a seamless ERP upgrade by orchestrating the cleaning and optimization of complex Materials Management datasets using Microsoft Excel (Advanced) and SQL.
- Reduced total processing time by 25% and enhanced data availability for stakeholders by streamlining procurement workflows and integrating Cloud platforms with SAP APIs.
- Eliminated reporting discrepancies and accelerated feature delivery by collaborating in an Agile/Scrum environment to develop and validate internal analytical applications against legacy systems.

Data Science Research Intern | New York University | Abu Dhabi, UAE

Sept 2023 – Dec 2023

- Ensured robust system interoperability and reporting accuracy by collaborating in an Agile Environment to develop and test in-house analytical applications leveraging Data Analytics.
- Cut manual data processing effort by 40% and improved actionable insights by building automated scraping and visualization pipelines using R and Python.
- Validated research findings for academic publication by applying rigorous statistical models to analyze Large-Scale Datasets and ensuring the highest standards of data reliability.

PROJECTS

Analyzing and Visualizing Human Reference Atlas (HRA) UI Usage | Python, Pandas, Plotly, AWS

- Enabled data-driven documentation improvements by processing 120K+ log records (665MB) by architecting a scalable analytics pipeline using AWS (EC2) to parse CloudFront logs with Pandas and NumPy
- Uncovered dominant user workflows and underutilized components by designing an interactive dashboard that visualizes session data using Matplotlib and Seaborn to map high-demand Knowledge Graph resources.

The Untold Stories of IPL: Advanced ML & Visualization | Python, Power BI, Tableau

- Identified key performance drivers beyond title counts and mapped player consistency by modeling win-to-brand-value correlations with Scikit-Learn ensemble models (SVM, Random Forest), and transformed aggregate statistics into compelling narratives utilizing SHAP, NLP, and advanced visualizations in Power BI and Tableau.