

## EDUCATION

---

### Northeastern University | Boston, MA | Sep 2024 – May 2026 (Expected)

- **Master of Science in Information Systems (GPA: 4.0)**  
*Relevant Courses:* Data Science Engineering Methods and Tools, Big-Data Systems and Intelligence Analytics, Application Engineering & Development, Programming Structures and Algorithm (PSA)

### Mar Athanasius College of Engineering (MACE) | Kerala, India | Aug 2016 – Jun 2020

- **Bachelor of Technology in Mechanical Engineering**  
*Relevant Courses:* Calculus, Linear Algebra and Complex Analysis, Probability Distributions

## PROFESSIONAL EXPERIENCE

---

### IQVIA | Kochi, Kerala, India

#### Software Developer | Oct 2020 – Apr 2023

- **Led the automation team** in my department, streamlining processes and reducing manual QA tasks by 250+ hours/month
- Streamlined data processing workflows by **implementing automated ETL pipelines** in Python, reducing manual data preparation time from 10 hours to 10 minutes per cycle
- **Automated weekly/monthly deliverables** using Python (Selenium), saving more than 75 manual hours per month
- Developed Python scripts to integrate data from multiple environments (API, Dashboards, Flat Files) and analyze anomalies using **NumPy, Pandas** and **Matplotlib** libraries reducing data processing time by 80%
- Eliminated manual QA bottlenecks by implementing **server-based automation** for end-to-end processes, increasing team productivity by 40%
- Delivered 15+ monthly reports through data analysis and validation using **Excel, Power BI, and SQL**, ensuring 100% on-time delivery
- **Optimized database performance** by maintaining and updating product database using **SQL (Toad)**, handling monthly additions of 100,000+ records
- Accelerated team capability by mentoring 3 interns in Excel data analysis and 4 team members in Python automation, reducing onboarding time by 60%

#### Software Developer- Intern | Jan 2020 – Apr 2020

- **Built scalable QA automation solutions** and extracted critical business data from client dashboards using Python libraries (Beautiful Soup, Selenium), improving data accuracy by 95%
- Transformed raw data into actionable business insights using Python and Excel, enabling data-driven decision making for pharmaceutical client projects

## PROJECTS

---

### Venture-Scope (Multi Agent - Agentic RAG): MCP, LangGraph, Pinecone, CI/CD Pipeline, LLM, GCP, Docker | Apr 2025

- Developed a comprehensive startup advisory platform integrating 3+ diverse data sources (Snowflake company datasets, yFinance, Google Maps API). Built a multi-agent Agentic RAG system with LangGraph and MCP servers, providing personalized business recommendations across 4 core functionalities. Implemented an end-to-end solution with FastAPI backend, Streamlit frontend, and automated workflows via Airflow and CI/CD pipelines, enabling data-driven market entry decisions

### Financial RAG Pipeline & Analytics Interface: Scikit-learn, Pinecone, ChromaDB, Airflow, LLM, GCP, Docker | Mar 2025

- Built end-to-end RAG system for 5 years of NVIDIA financial reports with automated extraction, custom PDF uploads, and OpenAI embeddings using Pinecone/ChromaDB. Implemented metadata filtering and context-aware retrieval, leveraging 2 parsing methods (Docling, Mistral OCR), 3 chunking strategies, and 3 vector database options. Orchestrated processing via Airflow DAGs and deployed containerized solution with FastAPI backend and Streamlit frontend

### Snowflake Pipeline - FRED: Snowflake, Snowpark, CI/CD Pipeline, Tasks(DAGs), Github Actions, S3 | Feb 2025

- Using Snowflake's Snowpark for Python, the system efficiently extracts, transforms, and validates financial data for advanced analysis and reporting. It builds a pipeline to create a dashboard that displays the inverse treasury yield curve, derived from Federal Reserve Economic Data's (FRED) U.S. Treasury yield data for 10-Year and 2-Year bonds

## TECHNICAL SKILLS

- 
- **Languages:** Python, SQL, Java, C++, HTML5, CSS3
  - **Tools:** Apache Airflow, Docker, MCP, Git, GitHub, Microsoft Excel, Power BI, Pinecone, ChromaDB, Mistral OCR
  - **Database & Cloud:** Snowflake, MySQL, dbt, GCP, Google Cloud Run, AWS S3, GitHub Actions, Redis Streams, Google Compute Engine
  - **Python Libraries & Frameworks:** NumPy, Pandas, Matplotlib, Scikit-learn, TensorFlow, Beautiful Soup, Selenium, LangChain, FastAPI, Streamlit, LiteLLM, LangGraph, spaCy, PyMuPDF, snowflake-connector-python