

Mike Rabayda

267-753-5299 · mrabayda4@icloud.com · Philadelphia, PA ·

[in linkedin.com/in/michael-rabayda](https://www.linkedin.com/in/michael-rabayda) github.com/mikerabs mikerabayda.com

SUMMARY

Software Developer and Data Scientist with expertise in full-stack development, data analysis, project management, and cloud architecture. Proven track record of managing development teams and delivering scalable systems. Strong background in machine learning and rapid product development. Former Division I Starting Pitcher.

EMPLOYMENT HISTORY

Upside Global - techstars_ funded Consulting Firm - Remote - *Data Scientist Intern* *April 2025 – Present*

- Architected **AI Chatbot** database as sole Data Engineer, enabling GMs to query and compare professional-grade sports products in real time.
- Normalized all vendor data submissions using **Python scripting**, ensuring high-quality inputs for chat responses.
- Automated marketing analytics with lightweight **JS** scripts to parse podcast and blog engagement, capturing which teams viewed content, eliminating manual tallying time by 1 hour per report and accelerating outreach.

HC Digital Scholarship - Web App Service - Remote - *Web App Consultant* *September 2023 – August 2024*

- Coordinated **MongoDB** database migration to store corpus of 270+ Latin texts for all webapps.
- Successfully **containerized** entire web application, ensuring compatibility and ease of setup for all team members across different environments.
- Promoted to team lead **managing four developers** after successful deployment of Latin statistics web app.
- Managed development team remotely using **Scrum** methodologies.
- Authored comprehensive guidelines for setting up development environments, reducing the onboarding timeline for new team members from 2 weeks to 3 days.

HC Digital Scholarship - Web App Service - Ardmore, PA - *Web App Engineer* *June 2023 – August 2023*

- Secured NIH-backed funding enabling next quarter hires
- Developed **full-stack** web application using **FastAPI** and **JS** to calculate student readability scores for Latin texts based on LexR, enhancing educational tools for students and educators in classics.
- Reduced ticket resolution time by **40%** by migrating project management system to Notion attributing ownership and centralizing project data.
- Established comprehensive project wiki with documentation to onboard new developers and streamline development processes, improving project organization and team efficiency.

International Data Corporation - IDG's Tech Consulting Firm - *CX Data Analyst Intern* *May 2022 – August 2022*

- Presented **churn analysis** to Global Vice President of CX, guiding focus initiatives within Customer Success related to logo retention.
- Identified product laggards utilizing **pivot tables** and **vlookup** on Salesforce data.
- Cleaned account bookings data to be used for customer retention rate models.
- Centralized significant international customers guiding a global customer survey, yielding over 1100 responses

TECHNICAL SKILLS

Languages: Python, Java, JavaScript, C#, HTML, CSS, Bash

AI Tools: ChatGPT, Copilot, Gemini, Cursor, Ollama, PyTorch

Data Analysis: Pandas, R, Tableau, Salesforce, Power BI, Office365, D3.js

Cloud & Tools: AWS, Google Cloud, GitHub, Docker, Docker Compose, DigitalOcean, Linux

Data Engineering: PostgreSQL, MySQL, MongoDB, CassandraDB, Hadoop(HDFS), Spark, Kafka, Airflow, Confluent

Methodologies: Agile, Scrum, CI/CD, Object Oriented Design, API Integration, Database Migration, Containerization

EDUCATION

Fordham University, Graduate School of Arts & Sciences

September 2024 - December 2025

M.S. in Data Science

3.875 GPA

Relevant Coursework: Machine Learning, ML in Finance, Big Data Computing, Relational Databases, NoSQL Databases, Data Visualization, Linear Algebra

PROJECTS

Stuff+ Model for Pitching Grades - (XGBoost, Pandas, Optuna)

- Achieved **0.133 R²** value explaining 13.3% of pitch-level Run Value on MLB 2024 data - **over twice** the signal of existing open-source models.
- Designed and implemented **xRV** and **xBatSpeed** features to denoise batted-ball outcomes, then standardized the result to an industry-scale Stuff+ index (mean = 100, SD = 10)
- Conducted cross-level validation by ranking all NCAA Division I programs by average Stuff+ scores — highlighting SEC (118.3) and ACC (113.9) conference means - demonstrating robust alignment with real-world pitching dominance.

Enron Email Network Analysis - (Networkx, Pandas, Louvain)

- Implemented **clustering**, **community detection**, and traversal analysis to uncover communication patterns, response times, and role-based networks among employees.
- Ranked **in- vs out-degree centrality** to spotlight prolific senders and central recipients.
- Generated **directed graph** visualizations to analyze frequency, burstiness, and anomalies in email communications, identifying key influencers and hidden hierarchical structures within the organization.

Transformer Neural Network for Stock Price Prediction - (PyTorch)

- **Transformer-based Neural Net** model engineered with technical, macroeconomic, and sentiment data and volatility-adaptive augmentation for robust, real-world allocation.
- 3905% portfolio return achieved (2019–2024 backtest), outperforming equal-weight benchmark (1069%) by over 3.6×
- 22% lower volatility delivered, improving stability during high-stress market periods (2020, 2022).
- 84% higher Sharpe ratio (2.740 vs 1.488), maximizing risk-adjusted returns through Sharpe-driven optimization.

Pitching Report Site - (Flask, Pandas, HTML)

- Developed **PDF pitching report** generation webapp from Trackman datasets using **Flask**, replacing a manual Jupyter workflow with a scalable, production-ready solution.
- Built modular **data pipeline** for ingesting uploaded Trackman files, dynamically generating per-pitcher visualizations and exporting standardized PDF reports.
- Integrated configurable Stuff+ metrics, enabling both manual input through the UI and automatic computation via persisted model weights, ensuring flexibility across coaching preferences.
- Implemented clean API endpoints and robust file handling to support seamless usage by non-technical staff, while following software engineering best practices for maintainability and deployment.

User Account Data Pipeline - (PySpark, Docker, Kafka, Airflow, CassandraDB)

- Automated **real-time** user account data pipeline for a **Google Cloud VM Cluster** using **Docker Compose**, promoting rapid prototyping and reducing operational complexity.
- Centralized system administration by leveraging UIs for **Airflow**, **Kafka**, **Confluent**, and **Spark Stream** Monitoring, simplifying troubleshooting, pipeline management, and user-access to live system metrics.
- Enabled seamless scheduling, triggering, and monitoring of ingestion workflows via Airflow DAGs and built-in web controls, minimizing manual effort and ensuring fast adaptation to business requirements.

ATHLETICS

Fordham University

- Weekend Starting Pitcher
- A10 Commissioner's Honor Roll

Haverford College

- Varsity Baseball Captain
- Centennial Conference Pitcher of the Year
- First-Team All-Centennial Conference
- Philly-SIDA Academic All-Area Team
- ABCA's Team Academic Excellence Award
- Centennial Conference Spring Academic Honor Roll