

Brian Donaghey

✉ briandonaghey21@gmail.com

in linkedin.com/brian-donaghey

🐙 [GitHub](#)

📁 [Personal Website](#)

EDUCATION

University of Pittsburgh

B.S Computer Science, Minor in Economics, CS Departmental Honors

GPA: 3.62/4.0

Pittsburgh, PA

Aug. 2021 – April 2025

Coursework: Data Structures and Algorithms, Intro to Operating Systems, Intro to Database Management Systems, Software Engineering, Programming Languages for Web Applications, Software Quality Assurance, Intro to Cloud Computing, Functional Programming, Data Science with Python

EXPERIENCE

Data Engineer Intern

May 2025 – Present

Sheetz

Pittsburgh, PA

- Currently developing, scheduling, and maintaining data pipelines in **Databricks** using **PySpark** and **SQL**, with cloud-based data storage and processing managed through **AWS**.
- Researched and integrated Census demographic data from various public **APIs** using **Python**, enabling site selection decisions for new store locations, BI/analytics insights, and ML modeling for data science teams.
- Deduplicated **4 billion+ row tables** and fixed pipeline logic in Databricks to ensure accurate and reliable data.

Data Structures and Algorithms Teaching Assistant

Aug. 2024 – April 2025

University of Pittsburgh

Pittsburgh, PA

- Taught **50+ students** core data structures such as linked lists, binary trees/heaps, hashmaps, and graphs while mentoring them in algorithm design, complexity analysis, and **object-oriented programming** in **Java**.
- Facilitate weekly recitations and host office hours for Pitt's **Data Structures and Algorithms** class.

Data Engineer Intern

May 2024 – August 2024

Federated Hermes

Pittsburgh, PA

- Engineered **15+ financial data pipelines** in Databricks to perform **ETL** using PySpark and **Pandas**, transforming raw data into structured **JSON** files and ensuring data integrity with **SQL DDL/DML** queries.
- Led the development of a variance reporting model in **Power BI** for a company hackathon that was projected to save **\$95,400 annually** by reducing manual report analysis time and streamlining financial reporting.
- Developed a Power BI workspace monitoring system that analyzes large-scale enterprise data stored in **Delta/JSON** files, providing real-time insights into report usage and access patterns within the company.

PROJECTS

AI-Spotify Recommendation Engine 🐙 | *Python, React, Flask, OpenAI, NLP*

- Built a **Flask API** for an end-to-end AI music recommendation system using **GPT-4o** and the **Spotify API**.
- Enhanced recommendation quality through **Natural Language Processing** (**spaCy**, **TextBlob**, **NLTK**) to better interpret user queries, implementing **feedback loops** for learning, and **caching** to prevent duplicates.
- Deployed the **Flask** backend and the **React** frontend as separate services, enabling clean integration with the API.

Digital Attendance System 🐙 | *Google Cloud Platform, Docker, Node.js, Javascript*

- Collaborated with others to design and deploy a full-stack cloud-based attendance tracking system using **Google Cloud Platform**, featuring QR code generation for attendance tracking.
- Developed **REST APIs** with **Node.js** and **Express** for database interactions, utilizing **Firestore's NoSQL** capabilities for data management, and hosted the APIs on **Google App Engine**.
- Automated a **CI/CD pipeline** with **Cloud Build** and **GitHub** integration and deployed the application to **Cloud Run** using **Dockerized containers**.

TECHNICAL SKILLS

Languages: Java, Python, C, SQL (Postgre), JavaScript, HTML/CSS, Haskell

Frameworks: JDBC, JUnit, Mockito, Selenium, Node.js, Flask

Developer Tools: Git, Azure DevOps, Databricks, Power BI, Linux, GCP, Docker, Maven, VisualVM, AWS

Libraries: pandas, NumPy, PySpark, matplotlib, scikit-learn, TensorFlow, Keras, React

Concepts: Unit Testing, Integration Testing, Performance Testing, Agile, TDD, SDLC, ETL, Data Lake Architecture