

# Jaylen Callahan

(865) 555-0147 jaylen.callahan@example.com

linkedin.com/example/jaylencallahan Knoxville, Tennessee 37919



## SUMMARY

Data Engineering Intern and 2026 computer science graduate with supervised experience building dependable cloud data solutions for academic operations and consumer analytics. Developed Python and SQL **ETL pipelines**, Azure Data Lake workflows, Databricks tables, validation checks, dimensional models, and Power BI reports. Applied probability, sampling, **confidence intervals**, regression, classification, data profiling, and basic A/B testing across hands-on labs and a capstone project. Documented transformations, standards, test evidence, and delivery progress in GitHub while communicating scope, limitations, and findings with faculty sponsors and project partners. Spanish-English communication supports business conversations across functions and cultures. Eager to contribute reliable **data foundations**, thoughtful analysis, and clear execution from day one while learning from experienced engineers and supporting practical **business decisions**.

## EDUCATION

### Bachelor's Degree in Computer Science

2026

University of Tennessee, Knoxville GPA: 3.8

Knoxville, Tennessee

*Coursework: Python, SQL, Probability and Statistics, Data Structures, Machine Learning*

## TECHNICAL SKILLS

- **Programming Languages:** Python, SQL, Jupyter
- **Azure Data Services:** Azure Data Lake, Azure Storage, Azure cloud
- **Databricks Platform:** Databricks notebooks, Delta tables, Spark SQL
- **Data Integration:** ETL pipelines, ingestion jobs, data transformations
- **Data Modeling:** Dimensional modeling, star schemas, source keys
- **Data Quality:** Schema checks, null-rate checks, referential integrity
- **Statistical Analysis:** Probability, sampling, confidence intervals
- **Machine Learning:** Regression, classification, holdout evaluation
- **Experimentation:** A/B testing, distributions, outlier analysis
- **Business Intelligence:** Power BI, dashboard design, report filters
- **Source Control:** Git, GitHub, branching
- **CI/CD:** GitHub Actions, pull requests, automated tests
- **Data Monitoring:** Pipeline latency, data drift, metric stability
- **Documentation:** Technical documentation, transformation logic, test evidence

## SKILLS

- |              |                         |                        |                  |
|--------------|-------------------------|------------------------|------------------|
| • Python     | • ETL pipelines         | • Probability          | • Classification |
| • SQL        | • Data modeling         | • Sampling             | • A/B testing    |
| • Azure      | • Data validation       | • Confidence intervals | • Power BI       |
| • Databricks | • Statistical profiling | • Regression           |                  |

## EXPERIENCE

### Data Engineering Intern – Supervised Research Placement

January 2026 - August 2026

University Data Engineering Lab

Knoxville, Tennessee

Supported academic operations analytics through cloud ingestion, quality controls, dimensional modeling, reporting, and documented delivery practices. Worked under faculty guidance while coordinating technical findings and priorities with research stakeholders.

- Built **Python** and **SQL** ETL pipeline for 185,000 records, loading Azure Data Lake and Databricks tables.
- Added schema, null-rate, duplicate, and integrity checks, stopping flawed files before reporting jobs ran.
- Created dimensional model for enrollment and advising activity, documenting transformations and reusable GitHub patterns.
- Profiled **distributions**, outliers, and **missingness** across eight fields, guiding revised rules with faculty sponsor.
- Configured GitHub Actions **CI/CD** for SQL checks and Python tests, creating repeatable pull-request reviews.
- Published **Power BI** prototype, sharing pipeline status, limitations, and next deliverables during weekly reviews.

## Capstone Project

*Knoxville, Tennessee*

- Scoped consumer-demand problem with three teammates, converting sponsor interviews into requirements and acceptance criteria.
- Developed **Azure** Python ingestion jobs for 42,000 transactions, loading curated tables into **Databricks**.
- Used SQL transformations and documented models, preserving source keys across product, store, date, and transaction entities.
- Evaluated missing values, unusual distributions, and duplicates, recording **validation** decisions and test evidence.
- Designed A/B test for dashboard layouts, calculating **confidence intervals** before recommending a clearer option.
- Presented Power BI report and technical documentation, translating **regression** and **classification** findings into practical actions.

## Student Analytics Lab

*Knoxville, Tennessee*

- Prepared 26,000-row retail dataset with Python and SQL for profiling, evaluation, and reporting exercises.
- Built repeatable ETL notebook, standardizing dates and categories while retaining an audit log.
- Compared **probability**-based sampling approaches, summarizing distributions, **outliers**, and confidence intervals in an instructor-reviewed report.
- Implemented regression and classification **experiments** with holdout evaluation and **metric stability** checks.
- Connected curated outputs to Power BI, creating four-page filters for product, region, and time.
- Used **Git** branching and pull requests, resolving SQL errors and improving final repository documentation.

## Consumer Demand Analytics Pipeline 📅 2026

Student Services Data Quality Monitor 📅 2026

Created a documented monitoring workflow for student-services data, checking missingness, duplicates, schema consistency, outliers, pipeline latency, and reporting readiness.

- Dean's List, University of Tennessee – Fall 2025 and Spring 2026
- Undergraduate Research Presentation Award, College of Arts and Sciences – 2026
- Finalist, Tennessee Student Data Challenge – 2025
- Workshop Coordinator, Data Science Club – 2025-2026
- Peer Mentor, Undergraduate Research Lab – Spring 2026

- Microsoft Learn: Azure Data Fundamentals Learning Path 📅 2026
- Databricks Academy: Fundamentals of the Databricks Data Intelligence Platform 📅 2026
- Google Data Analytics Professional Certificate 📅 2025
- Valid Tennessee Driver's License 📅 2026

- Data Science Club, Workshop Coordinator – 2025-2026
- Spanish Conversation Partner, Center for Global Engagement – 2024-2026
- Undergraduate Research Lab Peer Mentor – Spring 2026

- Spanish (Proficient)
- English (Native)

**Work Status** : Authorized to work in United States. No sponsorship required.

## REFERENCES

---

AVAILABLE ON REQUEST