

# DUNG C. LE

Kirksville, MO, United States

Phone: (660)-730-9576 (USA) or 0779518187 (Zalo VN) | Mail: [lechidung204@gmail.com](mailto:lechidung204@gmail.com) | Portfolio: [MyPortfolio](#)

GitHub: <https://github.com/DungLe-304> | LinkedIn: [www.linkedin.com/in/dung-le-data304](http://www.linkedin.com/in/dung-le-data304)

## SUMMARY

Data Science sophomore with hands-on experience across the full analytics workflow — from data cleaning and SQL-based feature engineering to statistical modeling and interactive dashboard deployment. Demonstrated through end-to-end projects: a reproducible OLS/ANCOVA regression study in Python (SciPy, Statsmodels) and a MySQL + Chart.js EDA pipeline with a live-hosted dashboard. Proficient in Python, R, SQL, and data visualization. Seeking internship opportunities in Data Analytics, Data Engineering, Data Science, or related fields.

## EDUCATION

Kirksville, Missouri, United States - Truman State University

August 2024 - Present

B.S. in Data Science

GPA: 3.73/4.0

## PROFESSIONAL EXPERIENCE

Mathematics Tutor, Truman State University

September 2025 - Present

- Tutored **30+ students** across Precalculus, Calculus I, and Calculus II through 1:1 and small-group sessions, with **24 students achieving an A grade** across all three courses.
- Developed targeted practice sets and concept recap sheets spanning key topics: trigonometric identities & the unit circle, limits & derivatives, integration techniques (u-substitution, integration by parts, partial fractions, trigonometric substitution), improper integrals, and sequences & series (convergence tests).

## PROJECTS

Walmart Sales Analytics (SQL and Dashboard) - [Explore Live Dashboard](#)

March 2025

- Engineered a MySQL analytics pipeline** on **1,000 Walmart transactions** across 3 branches, performing data cleaning and feature engineering (derived **time\_of\_day**, **day\_name**, **month\_name** columns) and writing **20+ EDA queries** to analyze revenue trends, customer segmentation, and product line performance.
- Surfaced key business insights through branch and product-level analysis: Naypyitaw (Branch C) led in total revenue (**\$110,568**), Food & Beverages ranked as the top product line (**\$56,145**, **avg. rating 7.11/10**), and afternoon hours drove peak sales (**454 transactions**, **~\$148K revenue**) with a consistent **4.76% gross margin** across all product lines.
- Deployed a fully self-contained interactive HTML dashboard** featuring **5 KPI cards**, **7 Chart.js visualizations**, and dynamic filter controls (Branch, Product Line, Customer Type, Gender, Payment) — no server or dependencies required, live-hosted via **GitHub Pages**.

SAT & Graduation Rate Analysis (Python, SQL, Statsmodels, Pandas) - [View on GitHub](#)

December 2025

- Engineered an end-to-end Python analytics pipeline** (pandas, SciPy, Statsmodels) on 47 U.S. colleges — converting raw Excel into a cleaned dataset, a SQL seed script, and 7 analytical SQL queries (sector statistics, SAT band breakdowns, outperformer detection, residual analysis, and imputation sensitivity checks).
- Conducted comparative statistical analysis using Pearson correlation, OLS regression, and ANCOVA-style interaction modeling across **20 public and 27 private institutions**; identified a significantly stronger SAT–graduation association in public schools ( **$r=0.701$ ,  $p<0.001$** ) vs. private ( **$r=0.572$ ,  $p<0.01$** ) with distinct slope coefficients (0.104 vs. 0.062); interaction term  $p=0.175$  (non-significant).
- Built a fully reproducible analytical workflow with regression diagnostics (**Residuals vs. Fitted**, **Q-Q plot**, **Cook's Distance**), an interactive Plotly scatter chart with hover tooltips, a sensitivity analysis excluding imputed SAT records, and a `predict_grad_rate()` function with extrapolation warnings — all documented in a structured methodology writeup.

## TECHNICAL SKILLS

- Languages:** Python, R, SQL, Java, Javascript, HTML, CSS, TypeScript
- Data/ML:** pandas, NumPy, scikit-learn, PyTorch, SciPy, Statsmodels
- Statistics/Modeling:** EDA, OLS/ANCOVA (interaction models), classification, model diagnostics, cross-validation, hypothesis testing
- Tools/Backend:** Jupyter, Git, GitHub, FastAPI, Flask, MongoDB, REST APIs
- Visualization/BI:** Tableau, Power BI, Excel

## HONORS & AWARDS

- President's List (Semester GPA 4.0), Truman State University** Spring 2025
- 3rd Place, TruHacks x Boeing 2026 (Hackathon Business Category)** March 2025
- Anthropic Claude Code in Action certification** March 2025
- Google Gemini Certified Student** December 2025