

Lavya Midha

✉ lavyamidha015@gmail.com | 📞 413-800-1278
🌐 github.com/lavyam | [in linkedin.com/in/lavya-midha-9ba408245](https://www.linkedin.com/in/lavya-midha-9ba408245)

Skills

Languages & Frameworks: Python, R, SQL, NoSQL, Rust, Shell Scripting, React, Git, Linux.

ML & AI: Supervised/Unsupervised Learning, NLP, LLMs (BERT, Hugging Face), Feature Engineering, Model Evaluation, TensorFlow, PyTorch, Scikit-learn, Keras, MLOps.

Data Analytics & Visualization: Excel (Advanced), Tableau, Power BI, Python Panel, R Shiny, Pandas, NumPy, Matplotlib.

Cloud & Big Data: AWS, Azure, GCP, Spark, Hadoop, Docker, API Integration, ETL Pipelines, HPC.

Work Experience

ReferU.AI, Lewes, Delaware

May 2025 – Present

Machine Learning Engineer

- Leading the development of an AI-powered marketing optimization engine, integrating Google Ads, Analytics, and Search Console APIs to automate campaign strategy and reduce manual overhead.
- Designing ML models to assess lead quality using first-party interaction data, while building Looker Studio dashboards that surface real-time KPIs and AI-driven actions to support strategic decision-making.
- Strategic problem solving, machine learning, marketing analytics, API integration, executive dashboarding, cross-functional collaboration, Python, SQL, Looker Studio, Google Ads, Google Analytics, SEM Rush, GCP.

BU Spark!, Boston, Massachusetts

January 2024 - May 2024

Project Manager

- Led cross-functional teams on data innovation projects, including public budget transparency, voter engagement analytics, and social impact dashboards, translating insights into tools that enhanced strategic planning for partner organizations.
- Introduced Agile delivery practices as Scrum Master, streamlining execution and boosting team productivity by 40%, while ensuring continuous client alignment through structured communication and feedback loops.
- Innovation strategy, operational efficiency, stakeholder management, Agile project delivery, data visualization, Google BigQuery, Tableau, Python, cross-sector collaboration, Trello.

BU School of Public Health, Boston, Massachusetts

February 2023 - December 2023

Data Science Research Assistant

- Advanced air quality monitoring across 125+ Boston Public Schools by developing regression-based models and time-series visualizations to assess CO₂ and PM2.5 exposure, directly supporting district-level infrastructure planning.
- Built a computer vision solution to quantify environmental disparities between classrooms (e.g. window access), enabling equity-driven improvements in school facility standards.
- Delivered a custom Python Panel dashboard that improved real-time temperature monitoring, enhancing data accessibility for the operations team and reducing reporting lag by over 35%.
- Applied ML, environmental analytics, public sector innovation, dashboard design, cross-functional communication, Python, SQL, R, Python Panel, Linux, scikit-learn.

Education

New York University

Expected May 2026

M.S. in Data Science (Industry Concentration)

GPA: 3.72/4

Relevant Coursework: Probability, Statistics, Programming, Intro to Data Science, Text as Data, Big Data, Machine Learning

Boston University

September 2021 - May 2024

B.S. in Data Science, Minor: Statistics

GPA: 3.75/4

Relevant Coursework: Data Visualization, Machine Learning and AI, Algorithms, Statistical Modeling, Probability, Big Data, Data Mechanics, Programming, Optimization, Regression Analysis, NLP, Biological Data Science, Linear Algebra, DEI in Tech, Civic Tech, DS for Good, Inference, Foundations of Data Science, Analysis of Variance, Calculus.

Leadership and Awards: Undergraduate Research Grant, Dean's List, Secretary – BU Data Science Association, International Peer Mentor, Peer Tutor, Admissions Ambassador

Project Work

- Boston Police Department Budget Analysis (RiseUp):** Conducted a forensic analysis of Boston Police Department's overtime budget across units and ranks, identifying inefficiencies and structural overspending. Delivered interactive Tableau dashboards to advocacy partners, directly supporting data-driven policy recommendations and public sector budget reallocation strategy.
- COVID-19 Policy Analysis:** Built an end-to-end analytics pipeline using Azure and Power BI to consolidate and standardize pandemic policy data from ten countries. Produced comparative visualizations and strategy briefs identifying top-performing containment measures, supporting cross-border policy evaluation and global health preparedness consulting.
- NYC Subway Sentiment Analysis:** Engineered a sentiment analysis pipeline by scraping and processing 10K+ tweets via the Twitter API, leveraging BERT-based NLP to assess public perception of NYC subway services in real time. Mapped sentiment to borough-level ridership trends using geospatial analysis, generating actionable insights for urban mobility strategy and transit experience optimization.
- Recommender System:** Developed a scalable recommendation and customer segmentation system using PySpark on Dataproc, analyzing data from 330K users and 86K movies. Identified top "movie twin" user pairs with MinHash, implemented a tuned ALS model for personalized recommendations, and evaluated performance using Spark ranking metrics against a popularity-based benchmark.