Yash Bheke

linkedin.com/in/yash-bheke|github.com/yashbheke2000|yash.bheke2000@gmail.com|yash-bheke.lovable.app

SUMMARY

Data scientist and ML-focused graduate student with end-to-end experience building scalable data pipelines, accelerating model inference by 87%, and driving six-figure cost savings through production ML systems, A/B testing, and cloud-based analytics on Databricks, AWS, and MLOps stacks. Passionate about building production-ready solutions that drive measurable business impact.

EDUCATION

Master of Science in Information Systems - University of Cincinnati Aug 2024 - Apr 2026
Relevant Coursework: Gen AI, Statistical Computing, Data Mining for BI, Data Analysis, AI & Machine Learning, Data Visualization

Bachelor of Engineering in Electronics and Telecommunication - Mumbai University Aug 2018 - May 2022

TECHNICAL SKILLS

Languages: Python, SQL, R, Bash, C++, TypeScript
Data Science & Machine Learning: PyTorch, TensorFlow, Keras, scikit-learn, NLP, LLMs, RAG, A/B Testing
Cloud, Big Data & Databases: Databricks, PySpark, Snowflake, AWS (S3, RDS), BigQuery, Oracle, MySQL, PostgreSQL
Data Engineering: ETL (Batch & Real-Time), Feature Engineering, PCA, Operational Data Store, Data Warehousing
Visualization & Analysis: Tableau, Power BI, Looker, Excel, Zoho, Matplotlib, Seaborn
DevOps & Deployment: Docker, FastAPI, Flask, Git, GitHub, CI/CD

WORK EXPERIENCE

Parking Base (Data Science Intern) Feb 2026 - Apr 2026
Tech Stack: Looker, BigQuery, Gemini SDK, Python, LookML
- Developed a **Vendor Profitability Matrix** using Looker and BigQuery to analyze transaction volume vs. net profit, identifying discrepancies in commission structures to optimize inventory allocation across various parking reservation partners.
- Designed a **Manual Adjustment Audit** correlating **Biometric** vs manual **Backoffice** clock-outs, to enhance payroll integrity.
- Architected a payment-performance analytics workflow that surfaced declines by card type and channel, flagging high failure rate cards for review and enabling management to raise overall payment gateway approval rate to **98.7%** (+5.7% relative).

University of Cincinnati (AI Engineer Intern) Sept 2025 - Dec 2025
Tech Stack: Python, Tensorflow, Keras, N8N, RAG, vector store, LLM
- Built a custom n8n-based RAG workflow that ingests MSIS program FAQs from Google Drive, chunks content, creates OpenAI embeddings stored in a Pinecone vector database; enabling faster semantic retrieval with dynamic retraining-free updates.
- Achieved **95%** validation accuracy on a custom 30-intent classification model using a TensorFlow/Keras neural network.

University of Cincinnati (Graduate Assistant (IS 8034: Big Data Integration)) Sept 2025 - Dec 2025
Tech Stack: Databricks, DBeaver, ETL, Python, PySpark, SQL, AWS, S3
- Optimized legacy data pipeline on Databricks, improving processing efficiency by **31%** across 15 complex research workflows.
- Designed scalable data ingestion pipelines integrating AWS RDS/S3 with Unity Catalog for reproducible ML experiments.

Accelya (Data Scientist) Sept 2022 - Jul 2024
Tech Stack: Oracle, Python, SQL, ETL, A/B Testing, Fast API, Process Automation
- Accelerated model inference time by **87% (from 15+ minutes to under 2 minutes)** for the predictive analytics pipeline with PCA-based dimensionality reduction and XGBoost hyperparameter tuning which aided business analysts across 20+ clients.
- Saved **$370K+ annually** in operational costs by developing and deploying ensemble classification models (Random Forest, Gradient Boosting) via FastAPI, achieving **81%** accuracy in flight delay predictions for 12 clients with **1k+ daily inferences**.
- Improved user engagement by **18%** and reduced reporting errors by **16%** by designing A/B testing frameworks that combined Bayesian inference and hypothesis testing; identified **9 statistically significant improvements** in financial workflows and UI/UX.
- Automated data pre-processing, eliminating **60+ hours** of monthly manual effort by streamlining cleaning, validation, and EDA.

PROJECTS

Airbnb Optimum Pricing Tool [DataBricks, Python, sklearn, SHAP] Oct 2024 - Dec 2024
- Developed a regression model to predict Airbnb listing prices, improving metrics by **35.8%** compared to baseline models.
- Performed data cleaning, pre-processing, and feature engineering (outlier removal, imputation, one-hot encoding), applied SHAP to identify key price drivers and deployed via Streamlit for hosts to tweak parameters and view optimum pricing.

EXTRACURRICULAR ACTIVITIES

Project Lead, Neo Initiative (Pro-bono Consulting), University of Cincinnati Sept 2025 - Dec 2025
Student Ambassador, MSIS, University of Cincinnati Sept 2024 - Aug 2025
Technical Head, Tech Team, SIES Graduate School of Technology May 2020 - Jan 2022
Flagship Events Organized: ByteCamp, Hackathons, Dev Summit, TEDxSIESGST May 2020 - Jan 2022

Certificates

Python, Hackerrank
Advanced SQL, Hackerrank
Google Data Analytics, Google
Relational Database Design, Pluralsight