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
Courses: 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
Machine Learning & Statistics: TensorFlow, Scikit-learn, PyTorch, NLP, A/B Testing
Big Data & Cloud: Databricks, PySpark, Snowflake, AWS, BigQuery
Visualization & Analysis: Tableau, Power BI, Looker, Matplotlib
Deployment: Docker, FastAPI, 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-level profitability matrix** leveraging BigQuery to analyze **450K+ transactions** to quantify key margin drivers across booking channels, identifying commission misalignments contributing to **8.3% leakage in profit margins**.
- Reduced reconciliation review time by conducting anomaly detection analysis on payroll discrepancies (biometric vs. manual clock-outs), detecting outlier cases with **>15% time variance** and improving payroll review controls.
- Raised payment gateway approval rate to **98.7%** (+5.7% relative) by building a statistical segmentation of payment failure patterns that surfaced declines by card type, reason and channel, flagging high failure rate methods for review.

University of Cincinnati (AI Engineer Intern)   Sept 2025 - Dec 2025
Tech Stack: Python, Tensorflow, Keras, N8N, RAG, vector store, LLM
- Developed a **30-intent text classification model** for University admissions chatbot using TensorFlow/Keras on 1.4K+ labeled queries; improved weighted F1-score from 0.82 to 0.91 through hyperparameter tuning and class rebalancing.
- Automated ingestion and embedding updates for new FAQ documents, eliminating the need for full model retraining.

University of Cincinnati (Graduate Assistant)   Sept 2025 - Dec 2025
Tech Stack: Databricks, DBeaver, ETL, Python, PySpark, SQL, AWS, S3, RDS
- Optimized large-scale data transformation workflows in Databricks (PySpark/SQL) processing **15M+ records**, reducing end-to-end runtime by **31%** and improving data availability for faster downstream analysis.
- Built analytics-ready datasets by integrating AWS RDS and S3 sources; standardized schemas, implemented data validation checks and engineered derived features to support statistical modeling and 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 **83%** 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 listing prices, improving R^2 by **35.8%** (0.41 to 0.56) 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