(513) 276-5130
Cincinnati, OH
yash.bheke2000@gmail.com

# Yash Bheke
Data Scientist / Data Analyst

Portfolio: yash-bheke.app
github.com/yashbheke2000
linkedin.com/in/yash-bheke

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.

## TECHINCAL SKILLS

**Languages:** Python, SQL, R, Bash, C++
**Data Science & Machine Learning:** PyTorch, TensorFlow, Keras, scikit-learn, NLP, LLMs, RAG, A/B Testing
**Cloud & Big Data:** 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

## TECHNICAL EXPERIENCE

**AI Intern** **Sept 2025 - Dec 2025**
*University of Cincinnati* *Cincinnati, Ohio*

- 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.

**Graduate Assistant (IS 8034: Big Data Integration)** **Sept 2025 - Dec 2025**
*University of Cincinnati* *Cincinnati, Ohio*

- 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.

**Data Scientist** **Sept 2022 - Jul 2024**
*Accelya* *Mumbai, India*

- Accelerated model inference time by 87% (from 15+ minutes to under 2 minutes) by enhancing the predictive analytics pipeline with PCA-based dimensionality reduction and XGBoost hyperparameter tuning which aided business analysts across 20+ clients.
- Saved $200K+ 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 global airlines 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.
- Automated data pre-processing, eliminating 60+ hours of monthly manual effort by streamlining cleaning, validation, and EDA.

**Airbnb Optimum Pricing Tool** **Oct 2024 - Dec 2024**
*Academic / Personal Project* *Cincinnati, Ohio*

- Developed a regression model to predict Airbnb listing prices, improving metrics by 40% 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.

## EDUCATION

**Master of Science in Information Systems**, *University of Cincinnati, Carl H. Lindner College of Business* 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 Telecommunications**, *University of Mumbai* Aug 2018 - May 2022

## ACTIVITIES

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

## CERTIFICATES

**P**ython, Hackerrank (LINK)
**A**dvanced SQL, Hackerrank (LINK)
**G**oogle Data Analytics, Google (LINK)
**R**elational Database Design, Pluralsight (LINK)