

# Yash Krishna Bheke

(513) 276-5130 | yash.bheke2000@gmail.com | Cincinnati, OH | [linkedin/yash-bheke](#) | [github/yashbheke2000](#) | [Portfolio](#)

## Professional Experience

**Teaching Assistant (Big Data Integration IS8034)** | University of Cincinnati, Ohio September 2025 – Present

- Enhanced learning experience for 20+ students by collaborating with faculty to update course materials for the Big Data Integration class.
- Converted legacy code and labs from outdated formats to modern Python and SQL codes, authored step-by-step instructional guides and designed walkthroughs for data import via AWS RDS, catalog management, and hands-on labs, ensuring compatibility with Databricks.

**Software Developer (Data Analytics and Engineering)** | Accelya, Mumbai September 2022 – July 2024

- Reduced analytics dashboard **latency by 97%** (from 60+ min to <2 min, 39x faster) for **200+ enterprise clients** by collaborating with cross-functional teams, architecting materialized views and developing Java REST APIs, enabling real-time business reporting during peak loads.
- Slashed SLA **lag by 27%** by automating end-to-end ETL pipelines (Python, SQL, TypeScript), **empowering 30+ analysts** across 13 teams.
- Led **9 A/B testing experiments** on process and model improvements, analyzing airline revenue, reporting and accounting data to drive **adoption of 5 model enhancements**, achieving **22% increase in predictive accuracy** and a **16% reduction in financial reporting errors**.
- Resolved and deployed fixes for **52 prod bugs** and drove **16 Change Requests** (primary owner for 7) across C-based airline revenue module, TypeScript microservices and Python/SQL data pipelines. Documented root cause analysis in Confluence for knowledge transfer.
- Cut infrastructure spending by 23%** by leading the POC for containerized reproducible Docker analytics and model environments.
- Saved 28-30 hours** of team manual labor monthly by developing automated Python/Bash scripts to scrap and validate SFTP/CSV data.
- Ensured **99.95% on time availability** of **KPI-grade data over 6 months**, by designing ETL/ODS/DWH workflows **processing 500GB+** of airline's financial and operational data, model performance monitoring, automated data quality checks and error retry logic systems.

## Projects

**SafeStreet Chatbot** January 2025 – February 2025

- Architected a multimodal, real-time safety chatbot utilizing Gemini-1.5 Pro with Google News and Brave Search APIs via Dify framework; implemented retrieval-augmented generation (RAG) pipelines to provide live crime statistics, benchmark neighborhood risk indices, and delivered **lowest-risk housing recommendations** for homebuyers, realtors and property investors with AI-driven map visualizations.
- Integrated **Vision-LLM** pipelines that parse user-supplied crime-scene images (YOLOv8 + Gemini Vision) to perform scene understanding and deliver step-by-step mitigation guidance; extended **NLU stack** to support **20+ languages**, broadening access for diverse communities.
- Implemented rate-limited API wrappers, **JSON-schema outputs**, and geo-scoped prompt templates, establishing the groundwork for push alerts, 911-service hand-offs and complaint-triage extensions, achieving **sub 2 seconds response time** on standard **500-token** queries.

**Airbnb Pricing Tool** October 2024 – December 2024

- Cleansed and shaped **7k+ listings** across **106 columns**, stripping currency symbols from numeric fields, parsing bathroom counts from free-text, and **capping outliers**. Engineered revenue-signal features such as bed-to-guest ratio, shared-bath flag and stay-length.
- Optimized a **600-tree** Random Forest via RandomizedSearchCV (80/20 split, 5-fold CV), reaching  $R^2$  0.83, MAE \$23.9, RMSE \$65.6 and achieved a **40% error drop** versus a regularized linear model and consistently out-performed Gradient Boosting in grid evaluations.
- Uncovered the five strongest price driving features through permutation importance and **SHAP**. Also designed a self-serving pricing tool via a shareable Google Colab that lets hosts tweak listing attributes and instantly view model-backed pricing (**93% prediction band**).

## Leadership Experience

**Technical Head** | SIES Graduate School of Technology, Mumbai May 2020 – January 2022

- Upskilled a team of 20 through weekly workshops in Unity, C#, Git, and Blender, boosting project delivery speed by 35%.
- Directed end-to-end production of flagship tech events (ByteCamp, Hackathons, Developer Summit, TEDxSIESGST).

## Education

**University of Cincinnati, Carl H. Lindner College of Business** August 2024 - May 2026

Master of Science, Information Systems

Courses: Generative AI, Statistical Computing, Datamining for BI, Data Analysis, Web development, XML, Database Modeling & Design

**University of Mumbai** August 2018 - May 2022

Bachelor of Engineering, Electronics and Telecommunications

## Technical Skills

- Programming Languages:** Python, TypeScript, Java, JavaScript, C, C++, C#, Bash, Pro\*C, SQL, R
- Data Science & Machine Learning:** GenAI, NLP, LLMs, PyTorch, TensorFlow, Scikit-learn, RAG, Statistical Modeling, A/B Testing
- Databases & Big Data:** Oracle, MySQL, PostgreSQL, Snowflake, Databricks, BigQuery, Data Warehousing
- Data Engineering & Integration:** ETL Pipelines, REST APIs, Data Cleansing, Feature Engineering, Batch & real-time Processing
- Data Visualization & Analysis Tools:** Tableau, PowerBI, Looker, Excel
- DevOps, Infrastructure and OS:** Docker, Git, GitHub, Containerization, CI/CD, Windows, Linux