Yash Krishna Bheke

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

Professional Experience

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

- Built and deployed near real-time analytics dashboards by architecting Oracle materialized views and developed Java REST APIs, reducing dashboard latency from 60+ minutes down to under 2 minutes **(39x faster, -97% CPU usage)**, serving **200+ business clients** at peak loads.
- Designed, implemented, and maintained **end-to-end data pipelines** (ODS, DWH, ETL) utilizing advanced SQL, Python and TypeScript, automating data extraction, transformation, and feature engineering for machine learning models and analytics dashboards, reducing SLA lag by 27% and **empowering over 20+ business analysts** across 13 teams with faster, more reliable and SLA compliant data delivery.
- 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, resulting in **22% increase in predictive accuracy** and a **16% reduction in financial reporting errors**.
- Resolved and deployed fixes for **52 production issues** across C-based airline revenue modules, TypeScript microservices and Python/SQL data pipelines, ensuring high data integrity and model reliability. Documented comprehensive root cause analyses in Confluence for reproducibility and knowledge sharing and **led 16 process improvements** (primary owner for 7) from scoping through deployment.
- Built and automated analytics/testing environment replication using Docker/RedHat, creating Python scripts for reproducible model deployment and validation, cutting setup times from 1 day to 30 minutes **(95% faster)** and reducing infrastructure spend costs by 23%.
- Eliminated repetitive tasks by developing a library of Python and Bash automation scripts, which scrapped SFTP feeds, validated CSVs, incorporated **error-handling and retry logic,** and pushed API/ETL driven data to staging tables, **saving 2-3 hours of team time** every day.
- Engineered robust **ETL workflows processing over 500 GB** of airline financial and operational data, automated data quality checks, model performance monitoring and alerting systems for KPI-grade data delivery, **achieving 99.95% on time availability** over 6 months.

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 — May 2020 – January 2022

SIES Graduate School of Technology — Mumbai

- Upskilled a twenty-person team 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, SQL, Data Warehousing (DWH), Operational Data Store (ODS)
- **Data Engineering & Integration:** ETL Pipelines, SFTP, REST APIs, Data Cleansing, Feature Engineering, Batch & real-time Processing
- **Data Visualization & Analysis Tools:** Tableau, PowerBI, Excel (VBA), Matplotlib, Seaborn, ggplot2
- **DevOps, Infrastructure and OS:** Docker, Git, GitHub, Containerization, CI/CD, Windows, Linux