

MAYANK BAMBAL

Chicago, IL

(765) 767-1067

mbambal@purdue.edu

[linkedin.com/in/mayankbambal/](https://www.linkedin.com/in/mayankbambal/)

EDUCATION

Purdue University, Daniels School of Business

West Lafayette, IN

MS in Business Analytics & Information Management

August 2025

Indian Institute of Technology

Indore, India

Bachelor of Technology in Mechanical Engineering

May 2019

DATA SCIENCE EXPERIENCE

CIBC USA

Chicago, IL

Data Science Intern (Industry Practicum)

January 2025 – May 2025

- Developed an LLM-powered data cleaning and extraction pipeline to transform key information from private financial reports, cutting manual effort by 60% and significantly improving data accuracy.
- Added an LLM-driven validation step that contextually flags incorrect data extractions, cutting manual QA work by 30%.
- Built a RAG-based Q&A system using Azure AI Search and Azure AI Foundry, delivering context-aware insights from company filings and cutting compliance reporting time by 40%.

Accordion India

Hyderabad, India

Associate 1 (Data Scientist 2)

September 2022-July 2024

- Led the churn-reduction effort on Databricks, training a Random Forest using customer data stored in a Delta Lake and using SHAP to reveal key metrics that drove attrition, cutting churn by 5% year over year.
- Developed a live KPI dashboard using Databricks SQL and Power BI with automated daily email alerts to center managers, which increased operational efficiency by 3%.
- Developed an automated claims processing pipeline that ingested scanned forms into a Delta Lake, used Spark OCR to extract key fields, and ran time-series models to detect anomalies – enabling targeted interventions that recovered ~1% of annual revenue.

Analyst 2 (Data Scientist 1)

September 2021- September 2022

- Built an automated A/B testing pipeline on AWS (S3, Glue, EMR, Lambda orchestrated by Step Functions) that cut analysis turnaround from ~3 days to under 4 hours.
- Built Spark/Python workflows for statistical tests and loaded results into Redshift to power self-service Tableau dashboards, enabling a 12% jump in campaign conversion rate.

Capgemini Invent

Mumbai, India

Associate Consultant (Senior Data Analyst)

October 2020-September 2021

- Replaced a lexicon-based sentiment classifier with a fine-tuned BERT model, boosting sentiment analysis accuracy by 13% and enabling more data-driven marketing budget decisions.
- Engineered an end-to-end keyword extraction pipeline on Azure Synapse Spark to process customer reviews (using TF-IDF features and Azure Data Factory for ingestion).
- Delivered near real-time insights via Power BI dashboards, spotlighting emerging customer themes to guide product strategy.

Senior Analyst (Data Analyst)

July 2019-October 2020

- Co-developed Azure Data Factory pipelines and Python Azure Functions to ingest 100K+ daily Amazon bestseller rankings and review data into Delta Lake (Databricks), cutting data delivery time by 80%.
- Supported development of Power BI dashboards by preparing curated data tables in Databricks, creating DAX measures, and setting up role-level security; enabling category managers to monitor sudden ranking “bursts” and shifts in customer sentiment.

SKILLS AND CERTIFICATIONS

Core Competencies

- Statistical Modeling, Data wrangling, data visualization, statistical analysis, experimentation, model deployment, MLOps.
- Machine learning, deep learning, natural language processing, Large Language Models ,explainable AI, generative AI, retrieval-augmented generation (RAG), autonomous AI agents.
- Communication, critical thinking, problem solving, data storytelling, domain expertise.

Tools & Technologies:

Scripting Languages: Python, SQL.

Frameworks: Scikit Learn, TensorFlow(Keras), PyTorch, PySpark, Pandas, NumPy, NLTK, SpaCy, Git, Docker.

Tools: Tableau, Microsoft Power BI, Microsoft Excel, Microsoft Azure, AWS, SAS, Databricks.

ML/DL Models: Linear regression, logistic regression, decision trees, random forest, bagging & boosting, naïve Bayes, k-nearest neighbors (KNN), neural networks (CNNs, RNNs/LSTMs), deep learning, Transformers.

Certifications

- Visualization: Microsoft Certified: Power BI Data Analyst Associate (PL-300), Tableau Desktop Specialist.
- AI Engineer: Microsoft Certified: Azure AI Engineer Associate, AWS Certified: Machine Learning Engineer Associate.
- Data: IBM Data Science Professional Certificate.
- Cloud: Microsoft Azure Fundamentals (AZ-900), Microsoft Azure AI Fundamentals (AI-900), Microsoft Azure Data Fundamentals (DP-900), AWS Certified cloud Practitioner.

OPEN-SOURCE PROJECTS

- **Unstructured Data Extractor** ([Link](#)): Created simple data extraction pipeline skeleton on public reports using LLMs to automate data extraction.
- **SIC code web scraper** ([Link](#)): Created web scraper to pull SIC code for all the public companies from SEC website.
- **Airbnb Customer Retention Strategy** ([Link](#)): Predicted host attrition using ensemble models (Random Forest, Gradient Boosting, MLP) to inform targeted retention strategies.
- **Los Angeles Crime Analysis** ([Link](#)): Visualized and analyzed crime patterns in Los Angeles using Tableau dashboards.