

# Madhumitha Sekamuri

[madhutrish06@gmail.com](mailto:madhutrish06@gmail.com) | +1-614-483-7876 | [LinkedIn](#) | [GitHub](#) | Ashburn, VA(open to relocation)

## Summary

Product Data Analyst with 3+ years of experience uncovering customer behavior patterns, optimizing digital experiences, and delivering data-driven recommendations. Proficient in SQL, Python, Tableau, and statistical modeling with a strong background in experimentation and retention analysis. Known for transforming complex datasets into insights that shape product roadmaps and accelerate growth

## Experience

Moxie

Ashburn, VA

Product Data Analyst

Feb 2025 – Present

- Developed and optimized omnichannel data pipelines integrating online clickstream and in-store transaction data with PySpark, Airflow, and AWS Redshift; reduced query times by 40% and enabled near real-time analytics for merchandising and marketing teams.
- Delivered product and campaign insights through SQL, Python, and Tableau dashboards that identified customer journey drop-offs, informing retail strategies and driving a 15% improvement in digital ad ROI.
- Designed and executed controlled experiments (A/B and multivariate tests) to validate feature enhancements in checkout and ad placement; provided post-launch insights that improved conversion rates by 12% and enhanced customer engagement.

Behavioral Neuroscience – The Ohio State University

Columbus, OH

Product Data Analyst

May 2024 – Dec 2024

- Analyzed high-frequency interaction data using SQL, Python, and statistical modeling to uncover engagement patterns and feature adoption trends, providing insights that informed product design and improved decision-making.
- Designed and executed controlled experiments (A/B and cohort analyses) to measure the effectiveness of new tracking methods; delivered recommendations that increased measurement accuracy and enhanced usability for researchers.
- Built and deployed interactive dashboards in Tableau and Python to monitor engagement KPIs and behavioral metrics, reducing manual reporting effort by 40% and accelerating insight delivery to product stakeholders.
- Developed predictive models with regression and clustering techniques to identify retention drivers in sequential behavior data, generating insights that guided prioritization of experimental features and improved long-term engagement.
- Automated analytics workflows using Python and Airflow to streamline experiment tracking, data ingestion, and reporting, enabling reproducible analysis and scalable delivery of product insights.

Pelotonia Research Center – Cancer Research

Columbus, OH

Data Analyst

Jan 2023 – May 2024

- Built and automated ETL pipelines using Python, SQL, Apache Airflow, and Azure Data Factory to ingest and clean large-scale structured and unstructured EHR/genomic datasets; improved data accessibility and reporting efficiency by 50%, enabling faster analytics delivery.
- Developed medical image processing workflows with MONAI, PyTorch, and SimpleITK on Azure ML, applying diffusion, affine, and deformable models to achieve 97% alignment accuracy across histopathology slides and improving reliability of downstream analysis.
- Designed predictive and survival analysis models (scikit-learn, XGBoost, Lifelines) to evaluate patient outcomes and treatment response; applied Kaplan-Meier and log-rank tests to identify survival patterns that guided clinical research decisions
- Created interactive Tableau dashboards deployed on Azure App Service to visualize survival curves, treatment efficacy, and patient stratification, empowering clinicians and researchers with self-service insights for decision-making.

Cognizant Technology Services

Chennai, India

Machine Learning Engineer

Mar 2021- Dec 2022

- Built and deployed **risk prediction models** using **LightGBM and pandas** on **Amazon SageMaker** to classify high-risk insurance policies, enhancing fraud detection and underwriting decisions with a **30% improvement in accuracy** across 100K+ policy records.
- Deployed ML models as **REST APIs using FastAPI**, containerized with **Docker**, and hosted on **AWS ECS**, while managing experiment tracking and version control via **MLflow on S3**, streamlining the MLOps lifecycle in a regulated insurance environment.
- Designed and orchestrated **scalable data pipelines** using **Databricks (PySpark)** and **dbt**, integrated with **AWS Glue and Redshift**, automating ingestion and transformation of **1M+ insurance records** and reducing model retraining time by **40%**, supporting real-time policy risk scoring workflows.
- Conducted comparative benchmarking of machine learning algorithms (**LightGBM, XGBoost, Logistic Regression**) using cross-validation and **AUC/F1** metrics to determine optimal model for policy risk classification.
- Built and automated end-to-end **ETL pipelines** using **Python, SQL, Apache Airflow, and Pandas** on **Azure Data Factory**, to ingest, clean, and integrate large-scale structured and unstructured **EHR and genomic datasets**, improving data accessibility and reporting efficiency by **50%**.

## Skills

- Product Analytics & Experimentation:** A/B Testing, Cohort Analysis, Funnel Analysis, Regression Modeling, Survival Analysis, Kaplan-Meier, Log-rank Test
- Programming & Languages:** Python, R, SQL, Java, JavaScript, HTML/CSS
- Data Analytics & Visualization:** Pandas, NumPy, Tableau, Matplotlib, Seaborn, Plotly, Metabase, Exploratory Data Analysis (EDA)
- Machine Learning & AI:** Scikit-learn, XGBoost, LightGBM, PyTorch, TensorFlow, Keras, H2O, MONAI, Lifelines, Transfer Learning
- LLMs & NLP:** OpenAI (GPT-4, ChatGPT APIs), Google PaLM, LLaMA 2, Mistral, LangChain, Retrieval-Augmented Generation (RAG), FAISS, ChromaDB, Prompt Engineering
- Data Engineering & Pipelines:** Apache Airflow, dbt, Azure Data Factory, AWS Glue, PySpark, ETL, Data Cleaning, Data Integration, Data Modeling
- MLOps & Deployment:** FastAPI, Flask, Docker, MLflow, REST APIs, AWS SageMaker, Azure ML Studio, GCP AI Platform, Cloud Composer
- Cloud & Databases:** AWS (Lambda, ECS, S3, Redshift), Azure (App Service, Data Factory), GCP (Compute Engine, TPU), MySQL, PostgreSQL, DynamoDB, Elasticsearch, Neo4j
- Data Governance & Compliance:** Data Governance & Privacy (HIPAA, PII), Cross-Validation, Hyperparameter Tuning, Model Benchmarking
- Certifications:** Udacity – Product Analytics Nanodegree, Coursera – Data Analysis with Python, Microsoft Azure Cloud Fundamentals, Azure AI Fundamentals, Google – Python for Data Science Certification, CyberArk Trustee Certification, NPTEL – Java, Machine Learning, CXL Institute – A/B Testing Mastery

## Projects

AI-Powered Business Idea Evaluation System:

Built an end-to-end pipeline to evaluate 500+ business ideas using LLMs (LLaMA 2, Mistral, ChatGPT APIs) and classical ML models; performed text preprocessing, EDA, and model benchmarking with Logistic Regression and Decision Trees, achieving over 85% agreement with expert-labeled criteria through cross-validation and hyperparameter tuning.

Text-to-SQL App:

Built a **Streamlit application** that translates natural language into SQL queries using **LangChain and Google PaLM**, enabling seamless interaction with **MySQL databases**. Integrated prompt chaining, query execution, and result visualization to support real-time data exploration for business users.

Smart Mobility - [Smart-Mobility-GNN](#)

Built a GNN-based model using **PyTorch Geometric** on the **Open Traffic Dataset**, modeling road networks as graphs for route optimization and traffic prediction; achieved a **15% improvement in path prediction accuracy** over baseline models.

Face-X - [Face Recognition](#)

Developed a high-accuracy facial recognition attendance system using OpenCV, Haar Cascade, and LBPH, achieving 98% accuracy in real-time identification enhanced with Dlib and FaceNet for robust feature extraction. **Presented and Published at ICIVC 2022.**

## Education

Master of Science in Computer Science and Engineering – Ohio State University – CGPA: 3.53

Dec 2024

Coursework: Neural Networks, Data Mining, Fairness in Artificial Intelligence and Databases, Data Visualization, Parallel Computing, Advanced OS, Cybersecurity

Bachelor of Technology in Computer Science and Engineering – Anna University – CGPA: 4

June 2021

Coursework: Problem-Solving, Object-Oriented Design, Data Structures & Algorithms, Databases, Operating Systems, System Design, Networking, Machine Learning