

Soumya Bhate

soumyabhate19@gmail.com | +1 (667) 452-0629 | [linkedin.com/in/soumyabhate19/](https://www.linkedin.com/in/soumyabhate19/) | github.com/soumyabhate | [soumyabhate-portfolio](#)

PROFESSIONAL SUMMARY

Data Science graduate student with hands-on experience in machine learning, NLP, and statistical analysis, and a track record of building scalable data pipelines and BI dashboards using Python, SQL, and Tableau. Skilled at translating complex data into actionable insights that support data-driven decision-making, with a strong interest in applying analytics to guest experience and media/entertainment platforms.

EDUCATION

Master of Professional Studies, Data Science | University of Maryland, Baltimore County (UMBC), USA | 2024–2026

CGPA 3.74/4.0 | Specialization: Data Modeling/Warehousing, Database Administration, Statistics & Probability, Statistical Modeling

Bachelor of Technology, Information Technology | Medi-Caps University (MU), Indore, India | 2019–2023

CGPA 3.51/4.0 | Specialization: Web Technologies & IoT

EXPERIENCE

Business Intelligence Analyst & IT Assistant | UMBC | Baltimore, MD

March 2025 – May 2026

- Built Tableau dashboards and SQL/Python workflows as part of a student team that migrated 170+ legacy reports to Tableau, with 100+ more in progress across a ~300-report initiative.
- Built and maintained ETL pipelines and data warehouse schemas for scalable reporting.
- Wrote optimized SQL queries to process large datasets, improving reporting efficiency.
- Partnered with stakeholders to translate business requirements into data solutions.

Data Engineer & Business Consultant | OMEG Digital Technology Solutions | Remote

Jan 2023 – July 2024

- Built end-to-end data pipelines for web/mobile analytics on large-scale interaction data, leading 2 projects including a business banking client in Malaysia.
- Ran statistical analysis and A/B tests to optimize user engagement and performance, contributing to a 12% increase in user engagement across client projects.
- Built KPI dashboards tracking user behavior, funnel performance, and cross-platform trends, completing 20+ Agile/Scrum tasks and 45+ Technical Specification Documents (TSDs).
- Cleaned, validated, and transformed large datasets using SQL to ensure data quality, resolving recurring tracking failures that eliminated data loss in downstream reporting.

PROJECTS

AI Legal Advisor (RAG + Voice Assistant) | UMBC

- Led Streamlit UI development and Groq LLM (Llama 3.3 70B) integration for a RAG-based legal chatbot.
- Integrated MongoDB Atlas vector search over 470+ embedded legal-text chunks for context-aware answers.
- Built a PDF Q&A pipeline with automated PII redaction across 10 categories of sensitive data.
- Added a full voice interface (speech-to-text/text-to-speech) and deployed the app to Streamlit Cloud.

Physical Fitness & Diabetes Prediction - Statistics | UMBC

- Analyzed NHANES survey data (1,028 observations) to study lifestyle predictors of fitness and diabetes.
- Built a logistic regression model with VIF multicollinearity checks to predict diabetes status.
- Used Welch's t-test to confirm activity level differs significantly by diabetes status ($p < 0.001$).

Crime & Economic Analysis - Machine Learning | UMBC

- Cleaned and integrated 29K+ DC crime records with Census ACS economic indicators.
- Built and compared 4 model families (Logistic Regression, Random Forest, XGBoost, PyTorch FFNN).
- Applied SMOTE and class-weighting, raising violent-crime recall from 3% to up to 70%.
- Excluded demographic data from modeling to reduce bias, following responsible-AI practices.

Stock Market Trends and Volatility Analysis - Big Data | UMBC

- Analyzed 5 years of daily price data for 8,000+ U.S. tickers (~700MB) using Apache Spark.
- Engineered returns, moving averages, and rolling volatility using Spark window functions.
- Built a Logistic Regression model (Spark MLlib) predicting next-day price direction at 53.7% accuracy.

SKILLS

Programming & Databases: Python, SQL

Machine Learning & AI: Scikit-learn, TensorFlow, PyTorch, NLP, Deep Learning, Feature Engineering, LLMs, RAG

Data Engineering: ETL Pipelines, Data Warehousing, Apache Spark, PySpark

Analytics & BI: Tableau, Power BI, Adobe Analytics, Adobe Launch, Adobe Target, Adobe Experience Manager (AEM), Adobe Dashboards

Tools & Libraries: Pandas, NumPy, OpenCV, Streamlit, LangChain, FAISS

Other: Statistical Modeling, A/B Testing, Prompt Engineering, Data Analysis