

BHANU SHARMA

+1 617-560-0716 | sharma.bhan@northeastern.edu | [Linkedin](#) | [Portfolio](#) | Boston, MA

EDUCATION

MS in Bioinformatics (Data Science Focus) (GPA: 4.0)

Northeastern University

Relevant coursework: Foundations of AI, ML, Statistics, Algorithms

Sept 2024 - Dec 2026

Boston, MA, USA

Bachelor of Technology in ECE (GPA: 7.7/10)

Punjab Engineering College (PEC)

Relevant coursework: DBMS, Data Structures Algorithms, Operating Systems

Aug 2016 - May 2020

Chandigarh, India

TECHNICAL SKILLS

Programming Languages: Python, R, SQL, UNIX/Shell, Scala, Java, SAS

Libraries: Pytorch, Tensorflow, Numpy, Pyspark, Pandas, Matplotlib, ggplot2, scikit-learn

Tools: Jupyter, Power BI, Tableau, Azure, Databricks, Snowflake, Spark, Docker, ADF, AWS, MS Excel, Github

Skills: Machine Learning, Artificial Intelligence, NLP, Deep Learning, Data Visualisation, Data Wrangling, Big Data, Data modelling, Data Warehousing, Business Intelligence,, ETL pipelines, CI/CD, Agile, Cloud Technologies

EXPERIENCE

Northeastern University, Boston, MA: Teaching Assistant

Sept 2025 - Present

- Assisted Professor Valerie Hower with MATH 7340: Statistics for Bioinformatics by conducting weekly office hours and supporting **90+ graduate students** in R programming and core statistical concepts

UnitedHealth Group (Optum), Gurgaon, India: Data Engineer 2

Sept 2023 - Aug 2024

- Designed numerous interactive dashboards using Power BI and MySQL delivering essential business insights and KPI results for **50+ clients spread across 4 departments**
- Developed a comprehensive notification & alerting system in Python that enabled users to send and schedule SQL result sets as emails via Outlook, **saving an estimated 6 hours weekly** for end-users

UnitedHealth Group (Optum), Gurgaon, India: Data Engineer 1

Aug 2020 - Aug 2023

- Migrated legacy ETL pipeline to Spark/Scala and Snowflake framework on Microsoft Azure, enabling a transition to a cost-efficient cloud-based platform and **reducing database costs by up to 90%**
- Redesigned on-prem ETL architecture from SAS to Spark/Scala, PySpark, and MongoDB(NoSQL) architecture hosted on Azure Databricks and Azure Data Factory allowing clients to adopt an efficient workflow **handling 3TB+ of daily data**
- Engineered novel Data Quality system using Great Expectations Python Library, achieving **accuracy of 99%**

PROJECTS

Rating Aware Beer Recommendation System - [Github](#)

Aug 2025

- Built and trained a beer recommendation system using KNN and Gradient Boosted Trees with LLaMA LLM wrapper to interpret natural language queries, recommending beers based on flavor profiles and predicted user ratings

Cervical Cancer Risk Prediction using Machine Learning - [Github](#)

Feb 2025

- Built, trained and evaluated decision trees based XG-Boost classifier ML model using Python's Scikit-learn library to predict whether a person has a risk of having cervical cancer with an **accuracy of 97%**

DNA Methylation Analysis Pipeline - [Github](#)

Nov 2024

- Reproduced DNA methylation analysis pipeline using Singularity/Docker from msPipe [research paper](#) hosted on HPC to generate DNA methylation profiles for any species and sample for advanced genomic analysis

PUBLICATIONS

Some Patterns of Sleep Quality and Daylight Savings Time Across Countries

Aug 2025

International Journal of Data Mining & Knowledge Management Process (IJDMP), 15(4). <https://doi.org/10.5121/ijdkp.2025.15401> Analyzed sleep quality patterns across 61 countries using predictive modeling and exploratory ML techniques.