

ANAGHA VEENA SANJEEV

Boston, MA | (857) 398-5307 | sanjeev.an@northeastern.edu | linkedin/anagha-v-s | github/anaghavs09

EDUCATION

Northeastern University, Boston, MA

December 2025

Master of Science in Data Analytics Engineering, **GPA: 3.73/4.0**

Related Courses: Applied Generative AI, Applied NLP, Data Mining Engineering, Statistical Learning, Computation and Visualization

NMAM Institute of Technology, Karnataka, India

June 2023

Bachelor of Engineering in Information Science and Engineering, **GPA: 8.66/10.0**

Related Courses: Machine Learning, Statistics, Database System Models, Data Science, Deep Learning, Python Programming

SKILLS

Data Engineering: dbt, Airflow, ETL/ELT Pipelines, Data Modeling, PySpark, Data Quality

Cloud & Databases: AWS (Athena, Glue, S3, Redshift), Snowflake, PostgreSQL, MySQL, MongoDB

Analytics & BI: Tableau, Power BI, SQL (Advanced), Excel, Statistical Analysis, KPI Development

Programming: Python (Pandas, NumPy, Scikit-learn), R, SQL

ML & AI: Predictive Analytics, NLP (spaCy, NLTK, LangChain), Time Series, Clustering

Tools: Git, Jupyter, Docker, Streamlit, FastAPI

Soft Skills: Communication, Analytical Thinking, Problem Solving, Leadership, Adaptability, Business Acumen, Curiosity

Certifications: [JPMorgan Chase & Co. Excel Skills Job Simulation](#), [Managing Big Data with MySQL](#), [AWS Cloud Practitioner Essentials](#)

PROFESSIONAL EXPERIENCE

Cohere Health, Boston, MA

January 2025 – June 2025

Business Intelligence Engineer Co-op

- Engineered scalable dbt data models processing 50M+ records with comprehensive testing framework, achieving 98% data quality and enabling reliable analytics for 15+ clients.
- Reduced reporting time by 40% through building self-service Tableau dashboards and optimizing SQL queries in AWS Athena, improving stakeholder decision-making speed.
- Automated daily data workflows using PySpark and Airflow, eliminating manual processes and ensuring timely delivery of complex datasets to external clients.
- Enhanced data governance by creating standardized models, comprehensive documentation, and quality checks supporting compliance.
- Collaborated with cross-functional teams (product, data, design, engineering) to translate business requirements into data solutions in agile environment.

Exposys Data Labs, Karnataka, India

June 2022 – June 2023

Data Scientist

- Delivered 5+ end-to-end analytics projects for clients across e-commerce, finance, and logistics, building automated ETL pipelines that reduced manual processing time by 60%.
- Created 15+ Power BI dashboards providing real-time business intelligence and KPI tracking, enabling data-driven decision-making for stakeholders.
- Built predictive analytics models for customer segmentation, demand forecasting, and trend analysis using Python (Pandas, Scikit-learn) and statistical methods.
- Processed 10M+ records daily through automated data pipelines, implementing data quality checks that improved accuracy by 70%.
- Applied NLP techniques to analyze 50K+ customer reviews using Python, extracting insights that informed product strategy.
- Conducted exploratory data analysis and presented findings to business teams, translating technical results into actionable recommendations.

PROJECTS

AI-Powered Investment Portfolio Analyzer | Python, LangChain, OpenAI, yFinance

August 2025 – October 2025

- Built intelligent portfolio analysis tool using LangChain combining quantitative metrics with real-time news sentiment analysis.
- Integrated multiple financial APIs (yFinance, NewsAPI) providing comprehensive investment insights and rebalancing recommendations.
- Application: Financial analytics, AI-driven decision support

Customer Segmentation Using RFM Analysis | Python, K-Means, Pandas, Tableau

June 2025 – August 2025

- Performed RFM analysis and K-Means clustering identifying 5 customer segments, revealing top 2 segments drive 60% of revenue
- Created interactive Tableau visualizations enabling targeted marketing strategies and customer retention initiatives
- Impact: Actionable segmentation for e-commerce marketing optimization

Crime Data Analysis & Forecasting | Python, ARIMA, Prophet, Time Series

September 2024 – December 2024

- Developed time series forecasting models for crime prediction supporting law enforcement resource allocation planning
- Performed geospatial analysis identifying crime hotspots and temporal patterns for proactive intervention strategies
- Application: Predictive analytics for public safety and resource optimization

Real Time Fraud Detection for Secure Financial Transactions | Python, Scikit-learn, Pandas, Statistical Analysis

June 2024 – August 2024

- Built anomaly detection system for identifying fraudulent financial transactions using statistical methods and pattern recognition
- Implemented data preprocessing and feature engineering techniques improving detection accuracy
- Application: Financial security and transaction monitoring