

# Aditi Neema

Seattle, WA | 4084625759 | [aneema099@gmail.com](mailto:aneema099@gmail.com) | [LinkedIn](#) | [Tableau](#)

## SUMMARY

Data enthusiast with 4+ years of experience in transforming legacy systems into modern, automated data solutions. Proficient in Python, SQL, Azure, AWS, and Snowflake, with hands-on experience building scalable data pipelines and predictive models. Also skilled in forecasting and analytics, bridging business strategy with engineering execution to turn complex requirements into data-driven systems that improve accuracy, speed, and decision-making.

## EDUCATION

### University of Connecticut

Master of Science, Business Analytics and Project Management

Aug 2022 - May 2024

Connecticut, USA

### Rajiv Gandhi Technical University

Bachelor of Technology, Computer Science and Engineering

Aug 2018 - May 2022

MP, India

## TECHNICAL SKILLS

- **Programming & Scripting Languages:** Python (pandas, NumPy, scikit-learn, Selenium), SQL(PostgreSQL, MySQL), PySpark, C++, SAS
- **Cloud Platforms:** AWS, Azure
- **Data Engineering & Warehousing:** Snowflake, Redshift, ETL/ELT Pipeline Development, Data Modeling, Stream Processing (Kinesis)
- **Machine Learning & Analytics:** Time-Series Forecasting (ARIMA, SARIMA, LSTM), Feature Engineering, Dimensionality Reduction, Predictive Modeling, Statistical Analysis
- **Data Visualization & Business Intelligence:** Tableau, Looker
- **DevOps & Version Control:** Git, GitHub, GitHub Actions, CI/CD Pipelines, Airflow (Basic)
- **Project Management & Collaboration:** Agile Methodologies, JIRA, Stakeholder Communication

## PROFESSIONAL EXPERIENCE

### A2R Software Solutions

Data Engineer

Jul 2025 - Nov 2025

Remote, USA

- Transformed a 25,000+ line legacy SAS codebase into Python data infrastructure, building Azure-Snowflake pipelines to power product performance and pricing analytics.
- Engineered 9,000+ features from 100+ event-level variables and reduced them to 15 key predictors to improve outcome and user behavior models.
- Built automated ingestion with Selenium and SFTP into Azure Blob Storage and Logic Apps-Snowflake workflows, eliminating 20+ hours of manual work weekly.
- Integrated AI model outputs into dashboards showing real-time probabilities, trends, and cohorts to guide product strategy and operations.
- Created technical documentation for migration, feature engineering, and pipeline architecture to support team adoption and maintenance.

### Zavvis Technologies

Data Engineer

Aug 2024 - Jun 2025

Remote, USA

- Engineered scalable streaming and batch data pipelines using AWS Glue and Amazon Kinesis to process 20k+ daily financial transactions in real-time, enabling near-instant analytics for revenue forecasting and risk assessment.
- Designed interactive Tableau dashboards connected to RDS databases, providing C-suite executives with real-time visibility into 15+ key financial metrics.
- Integrated AI-powered time-series forecasting models (ARIMA, XGBoost, LSTM) into production data pipelines, achieving 73% prediction accuracy and automating revenue/expense forecasts to enable proactive business decisions.
- Established Git workflows and CI/CD automation through GitHub Actions for model deployment and dashboard updates, reducing deployment errors and accelerating release cycles.

### University of Connecticut

Tutoring Assistant: Mathematics and Statistics

Mar 2023 - May 2024

Connecticut, USA

- Tutored undergraduate students in mathematics and statistics courses, including calculus, linear algebra, probability, and statistics. Conducted individual and group sessions tailored to diverse learning styles.
- Created study guides, practice problems, and visual aids to reinforce complex concepts. Fostered an inclusive environment that built student confidence in quantitative coursework.

### University of Connecticut: Budget Planning and Institutional Research

Data Analyst Intern

Aug 2023 - Dec 2023

Connecticut, USA

- Analyzed student lifecycle data to improve graduation and retention outcomes, contributing to a predictive model that led to a 20% increase in graduation rates.
- Processed 2M+ rows of student-related institutional data across nine datasets, conducting full-scale EDA using SQL.
- Identified and resolved intricate data quality issues using Python (pandas, NumPy) and SQL, establishing standardized cleaning procedures that streamlined data processing.
- Acted as liaison between analysts and university stakeholders, conducting recurring meetings to translate institutional goals into data strategies.

## Infoserve Consultants

Jun 2020 - Jun 2022

### Data Analyst

MP, India

- Conducted in-depth EDA to identify critical drivers of loan defaults, including credit score, income level, and debt-to-income ratio, enabling the development of targeted risk mitigation strategies.
- Collaborated with risk management teams to validate predictive models, ensuring alignment with RBI guidelines and regulatory requirements, and enhancing the accuracy of credit risk assessments.
- Monitored and evaluated model performance using KPIs such as precision, recall, and F1-score, achieving 10% improvement in recall and enhancing the detection of high-risk loan applicants.
- Utilized JIRA to track project milestones, deliverables, and timelines, ensuring on-time completion of analysis.

## Community Leadership

### Mental Health Listening Initiative | Instagram

Apr 2020 - Mar 2021

#### Founder & Community Engagement Lead (Volunteer)

- Founded an Instagram-based peer support initiative, leveraging platform analytics to identify high-engagement times and optimize outreach, reaching 250+ individuals during COVID-19 lockdown.
- Analyzed engagement metrics and audience demographics to refine content strategy and improve response rates.

## Projects

### Getting Started with UFC | [Looker Dashboard](#)

- Designed and deployed Snowflake ETL pipeline processing 15K+ UFC records using Python and SQL with window functions and dimensional modeling to create analytics-ready data warehouse for Looker dashboard visualization.
- Implemented data quality framework and performance optimization including Python-based standardization of 100+ weight class variants and Snowflake table clustering/partitioning.

### Insurance Policy Cancellation ML Pipeline | [GitHub](#)

- Developed multiclass classification model using PySpark Random Forest Classifier to predict insurance policy cancellations across 1M+ records, implementing hyperparameter tuning and custom threshold optimization for imbalanced classes.
- Engineered distributed data pipeline with PySpark SQL and MLlib transformers including feature imputation, OneHotEncoder/StringIndexer for categorical variables, VectorAssembler for feature vectorization, and EDA with statistical analysis to identify cancellation drivers.

### AI Generated Text Detection | [GitHub](#)

- Trained deep learning classifier to detect AI-generated vs. human text using Keras/TensorFlow neural networks with document embeddings, word count, and punctuation features on imbalanced dataset.
- Executed end-to-end ML pipeline including exploratory data analysis, feature correlation analysis, data preprocessing with StandardScaler, dropout regularization, and model evaluation with confusion matrix and ROC-AUC metrics.

## CERTIFICATIONS

- **Snowflake:** Snowpro Associate - Platform
- **HackerRank:** SQL
- **Google:** Introduction to Generative AI