

# JAYASHREE JOHNSON (She/Her)

One Pace Plaza, NY, 10038 | [jayashree.j@pace.edu](mailto:jayashree.j@pace.edu) | 201-736-4914 | [LinkedIn](#) | [GitHub](#)

## PROFILE

Graduate student pursuing a Master of Science (MS) in Data Science. Proficient in Python, SQL, and Machine Learning. Experienced in exploratory data analysis, ML algorithms, DBMS, & predictive modeling, and building end-to-end data pipelines for real-world applications.

## EDUCATION

<b>Pace University, Seidenberg School of Computer Science and Information Systems</b>	New York, NY
Master of Science (MS) in Data Science   GPA: 3.90/4	May 2026
<b>Justice Basheer Ahmed Sayeed College for Women</b>	Chennai, Tamil Nadu, India
Bachelor of Science (BS) in Advanced Zoology & Biotechnology   GPA: 8.1/10	May 2024

## RELEVANT COURSEWORK

Python Programming | Mathematical Foundation of Analytics | DBMS | Scalable Databases | ML | Data Mining | Data Algorithms

## TECHNICAL SKILLS

**Languages:** Python, SQL

**Libraries & Frameworks:** Pandas, NumPy, Scikit-learn, Matplotlib, Seaborn, Plotly, Statsmodels, XGBoost, TensorFlow (basic)

**Big Data & Cloud (SaaS):** AWS EMR, EC2, Hive, HDFS, PySpark, GitHub, Google Colab

**Modeling & Techniques:** Supervised & Unsupervised Learning, Classification & Regression, Time Series Forecasting, Dimensionality Reduction (PCA, SVD), Ensemble Methods, Feature Engineering, Cross-Validation, Hyperparameter Tuning

**Data Handling & Tools:** BeautifulSoup, Regex, Featuretools, Git, Excel

## PROFESSIONAL EXPERIENCE

<b>Data Science Intern – Big Data   ML Pipeline</b>	New York, New York
Pace University	Mar 2025 – May 2025
<ul style="list-style-type: none"><li>Designed a distributed pipeline on AWS EMR to process 7M+ DOT flight records using Hive and HDFS.</li><li>Engineered flight features and trained SGD, KNN, and Decision Tree classifiers for delay prediction.</li><li>Tuned hyperparameters with GridSearchCV, improving F1-score by <b>26%</b> and reducing false positives by <b>18%</b>.</li><li>Identified top delay-heavy airports (ORD, ATL) and airlines (WN), supporting better route planning with data visualization tools.</li></ul>	
<b>ML Research Intern - Time Series Forecasting</b>	New York, New York
Pace University	April 2025 – May 2025
<ul style="list-style-type: none"><li>Modeled NYC electricity usage trends using Prophet with linear/logistic growth curves on 15 years of data.</li><li>Forecasted 20-year demand across daily, monthly, and yearly horizons, improving long-term planning accuracy by <b>38%</b>.</li><li>Detected seasonal peaks (summer/winter) accounting for <b>60%</b> of total consumption variation.</li><li>Enabled a <b>30%</b> reduction in overcapacity planning errors, supporting smarter grid load allocation.</li></ul>	
<b>HeteroGene</b>	Chennai, Tamil Nadu, India
Research Intern	Jan 2024 – April 2024
<ul style="list-style-type: none"><li>Developed predictive models for genetic sex identification in Agapornis species, improving accuracy by <b>25%</b>.</li><li>Crafted PCR primers and achieved <b>90%</b> alignment accuracy using NCBI, PubChem, and UniProt datasets.</li><li>Automated analysis pipelines using Knime and Biovia Studio, reducing processing time by <b>12%</b> and enhancing reproducibility.</li></ul>	

## SELECTED PROJECTS

<b>NYC Traffic Volume Prediction &amp; Classification</b>	New York, New York
Pace University	April 2025 – May 2025
<ul style="list-style-type: none"><li>Built classification and regression models using hourly NYC traffic data to predict congestion levels and total traffic volume.</li><li>Tuned KNN, SGD Regressor, and Decision Tree models; post-tuning performance improved prediction accuracy by <b>28%</b>.</li><li>Uncovered peak-hour traffic contributing to <b>65%</b> of daily congestion, supporting smarter routing and infrastructure planning.</li></ul>	
<b>SQL Logic Programming – UDF</b>	New York, New York
Pace University	Feb 2025 – Mar 2025
<ul style="list-style-type: none"><li>Formulated SQL UDFs for automated age-based messaging, mean calculation, and variance analysis by IF-ELSE and WHILE loops.</li><li>Reduced manual computation time by <b>40%</b> and improved statistical query accuracy by <b>22%</b> in simulated business logic pipelines.</li><li>Enabled real-time data logic execution, scalable across high-volume inputs in analytics workflows.</li></ul>	