

Kshamaa Suresh

(313) 549-3330 | ks4423@columbia.edu | www.linkedin.com/in/kshamaasuresh

EDUCATION

Columbia University Masters of Science Data Science, 3.50 Coursework: Exploratory Data Analysis, Algorithms for Data Science, Applied Machine Learning, Machine Learning for Functional Genomics.	New York, New York Sep 2025 - Dec 2026
MS Ramaiah University of Applied Sciences Bachelors of Science Computer Science, 9.74 / 10. Coursework: Database Systems, Artificial Intelligence, Computational Intelligence, Deep Learning, Data Analytics, Python for Data Science, Internet of Things, Data Structures and Algorithms, Design and Analysis of Algorithms.	Bangalore, India Aug 2019 - Aug 2023

EXPERIENCE

Barnard College (Columbia University) Data Analyst Graduate Assistant	New York, New York Jan 2026 - Present
<ul style="list-style-type: none">Research position analyzing Drosophila connectome data under Dr. Gabrielle Gutierrez. Responsibilities include performing statistical and network analyses on neural connectivity data, implementing spatial analysis methods (k-nearest neighbor algorithms, centroid models, distribution models), applying community detection algorithms (modularity maximization, stochastic block models), developing Python-based analysis pipelines, consolidating and optimizing existing research code, and documenting all processes to meet reproducible research standards.	
Ecolab Associate Business Process Analyst	Bangalore, India Jan 2025 - Jul 2025
<ul style="list-style-type: none">Led strategic data integrity initiatives using Generative AI (Claude, Cortex) to computerize error detection in Snowflake, slashing manual validation by 80% and boosting data reliability.Managed full development lifecycle for AI analytics solutions, from prompt engineering on Databricks to deploying user-facing Streamlit apps and ThoughtSpot reports, helped analytics be completed in a matter of 3 hours.Partnered with GBS business units to develop a centralized Snowflake stage and ThoughtSpot dashboards, streamlining daily metric reporting and source-to-target validation.	
Ecolab Associate Data Engineer	Bangalore, India Jul 2023 - Dec 2024
<ul style="list-style-type: none">Automated testing of complex SAP-to-Snowflake data pipelines for Supply Chain projects, ensuring integrity across all transformation layers. Ensured 98% accuracy between data.Developed automated scripts to validate millions of records during Snowflake migrations, slashing issue resolution from 2-3 days to 4-5 hours.	
Ecolab Intern	Bangalore, India Jan 2023 - Apr 2023
<ul style="list-style-type: none">Designed and built a regression automation tool in Python to validate data transformations across Snowflake layers (ILA, ILB, Base Model).Delivered a key internship project having enhanced data quality assurance and was submitted as a final university thesis. Internship project covered 50% of final semester grade.	

PROJECTS

Ecolab: Generative AI for Data Analytics	
<ul style="list-style-type: none">Engineered a Generative AI data validation system using Claude API on Databricks (70% better accuracy) and Cortex AI on Snowflake to automatically identify variances and errors in Snowflake tables via advanced prompt analysis.Constructed an NLP pipeline to structure and process generative AI outputs, storing results in Snowflake and automating ThoughtSpot reports for business user consumption.Architected a full-stack data quality application with Snowflake Cortex, Mistral-large2, and prompt engineering to detect column-level data anomalies, processing JSON outputs into structured tables.Designed and deployed an interactive Streamlit UI on Azure, enabling business users to self-serve and visualize data mismatches, reducing manual validation burden.	
Ecolab: Email Alerting System	
<ul style="list-style-type: none">Devised an computerized pipeline monitoring system leveraging Python within Snowflake to query execution logs and track daily status of all data pipelines.	

- Engineered a proactive email alert mechanism automatically notifies stakeholders of pipeline failures, including specific details including failed job names and timestamps. Email alert mechanism ran in less than 120 seconds.
- Enhanced operational visibility and reduced downtime by providing real-time failure notifications, eliminating need for manual log checks and accelerating response times.

Ecolab: GBS Reporting

- Architected a centralized Snowflake stage environment to consolidate data from multiple subject area tables, streamlining access to vital business metrics for the GBS team.
- Developed automated ThoughtSpot dashboards to visualize key metrics and perform daily source-to-target data comparisons, enabling real-time error detection.
- Enhanced data integrity and reduced manual reporting effort by providing self-service analytics, eliminating need for tedious daily table iteration.

Ecolab: One Customer BCA Validation

- Constructed an ML classification framework in Python to identify and correct erroneous Ultimate Parent Code assignments in Snowflake data sourced from MDM.
- Engineered a data validation system computerizing detection of incorrect customer-site matches, ensuring accuracy of critical "One Customer" initiative data.
- Delivered actionable business intelligence by generating detailed reports on data quality issues, enabling proactive corrections before erroneous data reached customers.

Ecolab: Institutional BI Migration

- Automated validation for a large-scale data migration from SQL Server to Snowflake by developing a Python script to compare millions of records using primary key joins, ensuring data integrity.
- Engineered a data reconciliation system automatically identifying mismatched records and generated detailed result reports, pinpointing exact discrepancies for developers.
- Slashed bug-resolution timelines from 1-2 days to a few hours by providing developers with precise primary keys and data variances, dramatically accelerating migration process.

Ecolab: Snowflake and View Regression Automation

- Constructed an automated regression testing framework in Python to certify data integrity across multiple Snowflake layers (ILA, ILB, Base Model), streamlining data quality assurance process.
- Engineered a system to monitor over 1,000 production database views, automatically categorizing by project and logging performance impacts in Snowflake to prevent customer-facing issues.
- Formulated a user-friendly automation tool generated and executed complex validation queries based solely on a table name, simplifying testing for non-technical users.
- Delivered a key automation project cut down manual testing efforts and was recognized as a final university thesis, demonstrating strong applied data engineering skills.

MS Ramaiah University: Smart Irrigation Using Sensors

- Formulated and prototyped a smart IoT irrigation system leveraging NodeMCU ESP8266, integrating soil moisture and DHT11 sensors to collect and transmit real-time environmental data to a cloud interface.
- Engineered an computerized watering solution optimizing agricultural water usage, supporting sustainability goals by minimizing waste through data-driven triggers.
- Published system architecture and results as a research paper in the IFERP journal, contributing academic insights to field of IoT applications in sustainable agriculture.

Columbia University: Skincare Recommender

- Built LLM-based skincare product recommendation system as a part of AML course.
- Accepts user questions and provides personalized suggestions.

SKILLS

Technical Skills: Applied skills like: Python, Java, SQL, Deep Learning, Generative AI, Model Development, Snowflake, Azure, ThoughtSpot, Git, Excel, UI UX design, Graphs, Transformers, BERT.

Libraries Used: Employed tools like: TensorFlow, Scikit-learn, PyTorch, Pandas, Numpy, Matplotlib.

PUBLICATIONS

- Research Paper in the IFERP journal titled Smart Irrigation System using Soil Moisture sensor as a part of undergraduate degree.
- Self published anthology titled A Memo of Memories available on Amazon.
- A compilation of poems and stories titled World With Words available on Amazon.