

# Mateus Silva Aragao

(917) 732-3371 | [mateusilvaragao@gmail.com](mailto:mateusilvaragao@gmail.com) | [linkedin.com/in/mateus-silva-aragao](https://linkedin.com/in/mateus-silva-aragao) | [mateusaragao.com](http://mateusaragao.com)

## EXPERIENCE

<b>Data Engineer</b>	Oct 2024 – Present
<i>Amazon Ads (Amazon Marketing Cloud)</i>	Boulder, CO
<ul style="list-style-type: none"><li>• Spearheaded a solo data migration project that deprecated a legacy dataset, reducing daily processing from hundreds of billions to 20 billion records by eliminating duplication.</li><li>• Optimized a Scala/Spark ETL workflow, shrinking the AWS EMR cluster from 300 to 50 nodes and slashing annual operational costs by \$52K (74%), from ~\$70K to ~\$18K.</li><li>• Designed and deployed a high-throughput internal API using Kotlin, materializing 10 million daily records from an Apache Iceberg table into DynamoDB to serve data to internal services.</li></ul>	
<b>Junior Laboratory Associate</b>	Feb 2024 – Oct 2024
<i>New York University, Courant Institute</i>	New York, NY
<ul style="list-style-type: none"><li>• Developed a web interface using JavaScript, Flask, HTML/CSS for visualizing RNA splicing predictions, which was selected for presentation at the IEEE VisXai Workshop, improving research workflow for 5+ lab members.</li><li>• Built automated data analysis pipelines in Python using CellProfiler, Pandas, and Matplotlib to process and visualize cell profiling data from 200+ biological samples, generating actionable insights from large bio datasets.</li><li>• Mentored three undergraduate students in biological computing research, leading them presenting their work at an IEEE workshop, securing industry internships, and advancing to additional research opportunities.</li></ul>	
<b>Full Stack Software Engineer</b>	Jul 2024 – Sep 2024
<i>Freelance</i>	Remote
<ul style="list-style-type: none"><li>• Led the development of abilio.ai, an AI-powered business assistant platform for Brazilian franchisees, serving 40+ users and processing hundreds of daily LLM API calls for franchise operations support.</li><li>• Developed a full-stack application with Python/Flask backend and React frontend, implementing a RAG pipeline that processed document uploads, vectorized content using embeddings, and stored knowledge in Pinecone vector database for intelligent query retrieval.</li><li>• Designed dual-knowledge-base chat system with Pinecone for vector storage, enabling AI-powered responses grounded in both general franchise operations knowledge (administrative, financial, bureaucratic) and franchise-specific documents, handling 30+ daily chat sessions.</li></ul>	
<b>Data Engineer Intern</b>	May 2023 – Aug 2023
<i>Amazon Inc.</i>	Boulder, CO
<ul style="list-style-type: none"><li>• Engineered an observability framework in Scala for Spark jobs on AWS EMR, monitoring 500+ daily Spark tasks processing petabytes of data, providing detailed performance insights into shuffle, spill, and task execution metrics for distributed data pipelines.</li><li>• Developed real-time analytics dashboards in Quicksight, monitoring Spark application health, tracking performance metrics across 2 production pipelines, allowing the team to proactively identify performance bottlenecks.</li><li>• Implemented cloud infrastructure with AWS CDK in TypeScript, designing automated deployment pipelines integrated with production and beta environments, enabling near real-time data ingestion and processing of hundreds of records per minute for applications serving millions of customer data views.</li></ul>	
<b>Product Data Analyst Intern</b>	May 2022 – Aug 2022
<i>Moises.ai</i>	Remote
<ul style="list-style-type: none"><li>• Orchestrated A/B testing analysis for a novel AI-assisted songwriting editor across 2,000 users over three months, using Python (NumPy, SciPy, Pandas) and statistical methods including One-Way ANOVA, Kruskal-Wallis, and post-hoc tests to evaluate model efficacy and measure feature impact.</li><li>• Developed and maintained SQL ETL pipeline in Google BigQuery with optimized queries and CRUD operations, processing 500K+ daily user events to enable streamlined data preparation for Tableau dashboards monitoring user engagement and demographics.</li><li>• Built product analytics dashboards in Tableau for key stakeholders monthly reports synthesizing insights on user behavior, directly informing product development decisions and feature prioritization for the AI music startup.</li></ul>	

## EDUCATION

<b>New York University</b>	New York, NY
<i>Bachelor of Arts in Computer Science and Data Science</i>	Dec. 2023

## TECHNICAL SKILLS AND TOOLS

Kotlin, Scala, Python, JavaScript, SQL, TypeScript, Apache Spark, Data Engineering, Data Analytics, Software Engineering, API Development, Apache Iceberg, Flask, AWS DynamoDB, PostgreSQL, MongoDB, Google BigQuery, AWS CDK, AWS EMR, AWS Quicksight, AWS S3, AWS Lambda, Git, SQL ETL, Big Data, Portuguese, Spanish