Andres F. **Cano Botero**
Data Scientist | Researcher

About me

I am a Data Scientist with over two years of experience in a client-facing role. I've co-developed and delivered production-ready data science projects for the world's largest companies across multiple industries, including airlines, retail, manufacturing, pharma, banking, and insurance. I'm passionate about optimizing workflows and helping others increase their productivity using AI & ML. My science background allows me to think critically and solve problems methodically, I'm eager to apply my skills to accelerate progress and make a meaningful impact on people's lives!

Contact

- From Manizales, Colombia
- andrescano30@gmail.com
- +1 404-956-0320
- Conway, MA
- andresfelipecano
- andresfcano

Languages

- Spanish - Native Language
- English - Fluent
- French - Conversational

💼 WORK EXPERIENCE

2022–2025

Data Scientist – Services 📍 Atlanta, GA (Remote)
Dataiku
- Led the scoping, development, and deployment of 15+ production-grade AI solutions for Fortune 500 clients across aviation, banking, insurance, and retail—leveraging NLP, computer vision, forecasting, and LLMs to drive ML adoption, automate workflows, and improve decision-making, generating over $1M in client value through delivered services.
- Trained 100+ client stakeholders on Dataiku's platform—translating complex data science concepts into accessible learning paths while helping organizations identify and procure the right services to accelerate AI initiatives and build self-sufficient internal teams.
- Collaborated with internal R&D and product teams to develop and release custom Python plugins that extended Dataiku's functionality—empowering clients to apply advanced techniques, integrate external tools, and accelerate adoption across use cases.
- As Data Scientist in Residence at a major airline, formalized an MLOps framework with CI/CD and security standards—leading to the deployment of 3 AI projects within the first 6 months and building a reusable pipeline that scaled data science adoption enterprise-wide.
- Mentored and onboarded two new data scientists through shadowing sessions, technical guidance, and stakeholder coaching—enabling them to ramp up and transition to leading client projects independently quickly.

2020-2022

Graduate Researcher 📍 Boulder, CO
Labbe Lab, University of Colorado Boulder
- Leveraged university supercomputing resources and legacy software (written in C++ and Python) to build reaction mechanisms that uncover the thermal decomposition of soot-reducing diesel additives.
- Optimized research methods by writing a backward-compatible, high-performance computing pipeline that used Python and Unix shell scripts to generate, submit, and run quantum chemistry calculations on the RMACC Summit supercomputing cluster, saving days to weeks of computational work.

2019-2020

Data Scientist - Market Research 📍 Bogota, Colombia
Opera Urban Investment Firm (OIU)
- Developed a Streamlit app that displayed market trends, sales forecasts, and client activity from OIU's website. This established direct communication between internal teams and improved client outreach.
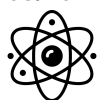
2018-2019

Undergraduate Researcher 📍 Rochester, NY
Extreme State Physics Laboratory (XSPL), University of Rochester
- Built a high-voltage transformer to supply ±200 kV to the High Amperage Driver for Extreme States (HADES), a 250 GW pulsed-power driver used to model matter at high temperatures and pressures.

🎓 EDUCATION

08/2020-08/2022

M.S. Mechanical Engineering 📍 Boulder, CO
University of Colorado, Boulder
Thesis: *Alkyl End Group Effects on the Thermal Decomposition of Fuel Additives.*
Advisor: Dr. Nicole J. Labbe
My work focused on optimizing and leveraging electronic structure theory methods and kinetic modeling to identify the decomposition products of two potential diesel additives. Studying the unimolecular decomposition of these molecules provides valuable insights into the formation mechanisms of soot particles in the exhaust of diesel engines.
GPA: 3.95/4.0

08/2015-05/2019

B.S. Astrophysics 📍 Rochester, NY
University of Rochester
Sigma Pi Sigma (National Physics Honor Society)
Lorenzo de Zavala Scholar, National Hispanic Institute
American Physics Society Member

——— Page Summary ———

This page highlights the various projects I have led as a Data Scientist. It showcases my experience building data science pipelines, creating LLM applications, and establishing MLOps frameworks. It also briefly describes my academic experience building automation tools for computational chemistry and using quantum chemistry to model the combustion of biofuels. These projects demonstrate my ability to work collaboratively and solve complex problems by leveraging computational tools, and together they paint a picture of my professional journey thus far.

——— Skills and Interests ———
——— Soft Skills ———

Curiosity   Self Motivated

Driven   Independent Thinking

Problem Solving   Team Work

Communication

——— Professional Skills ———

Data Science   Machine Learning

LLM Application Development

MLOps   Statistical Modeling

Web Development & Data Viz

Computational Physics & Chemistry

Electronic Structure Modeling

Astrophysics & Mathematics

Engineering Design

——— Other Interests ———

- Guitar 🎸
- Chess ♟
- Gym 🏋
- Travel ✈

- Triathlon 🏃
- Cooking 🍳
- Calisthenics 🤸
- Books 📚

🗁 RELEVANT PROJECTS & RESEARCH

Industry Projects - Dataiku

Automating RFPs with Self-Hosted LLMs   📍 Dataiku

Motivation: Field experts at a major insurance company manually handled hundreds of requests for proposals (RFPs), many of which had recurring themes, leading to inefficiencies and a loss in market competitiveness.

Solution: I designed and led the development of one of the first retrieval-augmented generation (RAG) pipelines for a customer at Dataiku that utilized open-source embeddings and large language models (LLMs) on local GPUs. By collaborating closely with stakeholders to ensure the solution met their needs, I built a Dash web application that enabled real-time querying, significantly improving RFP response times and overall efficiency.

MLOps and CI/CD of Data Science Projects   📍 Dataiku

Motivation: A U.S. airline's data science team was developing multiple projects using Dataiku but lacked a structured MLOps process, which caused projects in development and production to be in the same environment. This led to security risks and inefficiencies.

Solution: I led the effort to design and implement a clear MLOps strategy that adhered to the company's CI/CD process. I spearheaded the development of an Azure DevOps pipeline in conjunction with the airline's Cloud Architecture and Security departments, automating project deployments and ensuring that security and data quality checks were in place. This solution not only streamlined project delivery but also enhanced transparency and operational efficiency across the organization.

AI for Science - ML for Observation Quality   📍 Dataiku

Motivation: After each science observation, astronomers on duty (AoD) at a large observatory need to assess the quality and usability of the data. This was done by manually reviewing data and categorizing it into three groups: Pass (usable data), Semi-Pass (data may be useful but needs to be repeated), or Fail (unusable data requiring a repeat observation). This process is time-consuming, especially for uncertain cases.

Solution: To streamline this process, I spearheaded the development of a classification model to automate the assessment of observational quality. The model was robust and aligned with operational needs, allowing AoDs to focus only on uncertain cases. This solution greatly increased the efficiency of the data review process, reducing the manual workload.

Academic Projects - CU Boulder

Combustion Modeling for Biofuels   📍 CU Boulder

Motivation: Novel diesel additives have shown promise in reducing pollution from engines. Molecular modeling is crucial for understanding how these additives dilute soot and enhance engine performance; however, it requires expensive, manual, and time-consuming computational workflows, which limit scalability and efficiency.

Solution Developed a Python-based automation suite to optimize computational chemistry calculations, reducing manual setup time from months to minutes. Applied these tools to study the combustion behavior of diesel additives, uncovering key radical decomposition pathways relevant to soot and NOx formation. My M.S. thesis work contributes to cleaner fuel design.

🖥 INFORMATION TECHNOLOGY SKILLS

Data Science and Web Development	**Python** : *Higly Specialized* **HTML/CSS/JavaScript** : *Intermediate*
AI/ML	**Sci-kit Learn, Keras, PyTorch, MLFlow** : *Advanced* **Hugging Face & Langchain** : *Intermediate*
Data Visualization Tools	**Matplotlib, Streamlit, Seaborn, Dash**: *Advanced* **Tableau**: *Intermediate*
CI/CD Pipelines	**Azure DevOps**: *Intermediate* **Jenkins**: *Intermediate*
Version Control	**Git & Github**: *Intermediate*
Big Data	**Spark** : *Intermediate* **SQL** : *Intermediate* **Kubernetes & Cloud Deployments** : *Intermediate*
Linux	**Shell Script**: *Intermediate* **Slurm**: *Intermediate*
Typesetting	**LaTeX**: *Intermediate*