

# MARTINA PAEZ-BERRU

(646) 246-5581 | mp4395@columbia.edu | www.linkedin.com/in/martina-paez-bERRU

## EDUCATION

### Columbia University

#### Master of Science in Business Analytics, GPA: 3.66/4.00

New York, USA

Expected Dec 2025

Coursework: Artificial Intelligence/Deep Learning (NLPs, GenAI), Applied Machine Learning, Optimization, Digital Marketing, Marketing Analytics, Probability, Statistics & Simulation, Data Analytics, Managerial Negotiations, Capital Markets

Certificates: Google Analytics Certificate

### Ecole Polytechnique de Paris

#### Bachelor of Science in Mathematics & Economics, GPA: 3.58/4.00

Paris, FR

Jun 2024

Coursework: Asymptotic Statistics, Machine Learning, Convex Optimization and Optimal Control

## LANGUAGE AND IT SKILLS

- Programming Languages: Python, R, Julia, SQL, Stata, C++
- Machine Learning & AI: PyTorch, Scikit-learn, Sentence Transformers, LangChain, TensorFlow, Keras, OpenCV
- Data Engineering and Analysis: ChromaDB, LightRAG, Databricks, Spark, Scrapy, Playwright, Pandas, NumPy, TQDM, SymPy, SciPy, Statsmodels, NetworkX, Jupyter Notebook, Stata, Microsoft Excel
- Database & Retrieval Systems: PostgreSQL, vector databases, similarity search, embedding-based retrieval pipelines
- Visualization & Reporting: Matplotlib, Seaborn, PowerPoint, Google Slides API (automated generation)
- Infrastructure & Automation: Docker, Git, FastAPI, Celery + Redis, RESTful APIs, OAuth2
- Full-Stack Tools: React (frontend), FastAPI (backend)
- Languages: Spanish (Native), French (Professional Proficiency DELF B2)

## WORK EXPERIENCE

### Deep Venture Partners LLC

New York, USA

#### Data Engineer

Jun 2025 - Aug 2025

- Built full-stack AI system to automate VC tech scouting; orchestrated multi-agent workflows with real-time GPT-4 inference, Fast API micro services, and Google Slides API, sourcing academic research via Scrapy + Playwright
- Engineered end-to-end automation pipeline consolidating LLM outputs with Google Slides API to generate investor decks; took initiative in implementing OAuth 2.0, template selection logic, and dynamic content positioning
- Spearheaded design and deployment of normalized PostgreSQL schema and Entity-Relationship Diagram to structure data on universities, TTOs, research outputs, inventors, and categories; enabled scalable ingestion across 25+ institutions

### Columbia Engineering | Quantitative Research Lab MathWorks

New York, USA

#### Project Manager & Data Scientist

Jun 2025 - Aug 2025

- Led and collaborated with cross-functional team to develop an AI-powered investment ideation system, reducing research time by over 100 hours per month and accelerating alpha generation workflows
- Architected a modular ingestion-to-retrieval pipeline processing 3.41 GB of financial research using the Light/Simple RAG stack; deployed Sentence Transformers, LangChain, and ChromaDB for semantic search
- Engineered a Neo4j–ChromaDB graph database unifying 5,000+ academic papers and 20,000+ social media posts; collaborated with lab peers to enable multi-modal retrieval and powering ML-driven backtests across 12 strategies

### Ardian

New York, USA

#### Data Scientist - Capstone Project

Jan 2025 - May 2025

- Extracted insights from 1.5TB+ of European electricity market data spanning (10+ years, 20+ variables); partnered with experts and team members on exploratory data analysis and time series feature engineering for price behavior modeling
- Built an end-to-end Databricks pipeline for 30-year electricity-price forecasting, training RNN and XGBoost ensembles; cut overall error by 21 % (MSE 24.2 to 19.1 € / MWh) on 7.5 k forward predictions

### Center for Research in Economics and Statistics (CREST)

Paris, FR

#### Data Scientist & Economics Researcher

Jan 2024 - Mar 2024

- Developed hypotheses and employed estimations alongside researchers, applying statistical modeling, including Nonlinear Least Squares and Generalized Method of Moments, analyzing peer effects in Social Program involving 18,918 people
- Preprocessed large dataset of over 18,000 observations and 30 variables, spanning from 1997 to 2017, ensuring clean data for statistical analysis, and communicate data-driven insights and recommendations to collaborators

## ACADEMIC PROJECT & EXPERIENCE

### Stratyfy | Digital Marketing

New York, USA

#### Digital Marketing Freelancer

Jan 2025 - Mar 2025

- Proposed and designed headline A/B testing strategy (ethical vs. functional framing), carried out causal inference and statistical analysis; improved campaign decision-making and lowered CPC by 18%
- Performed advertising analytics through full-funnel analysis and causal inference, boosting ROI and channel effectiveness with 65+ clicks and a 2.10% CTR