Miguel Rivera-Lanas

Phone: (703) 981-0554 **Email:** mriveral@andrew.cmu.edu
Website: mriveralanas.com **GitHub:** rivera-lanasm

Data Engineer passionate about designing reproducible and scalable research methods and data pipelines with 5+ years of experience in asset management and academic research settings, working autonomously and in project management roles. Experience designing and implementing various aspects of cloud based research pipelines, from web scraping, data transfer, feature engineering, data warehousing, exploratory data analysis, machine learning, and data visualization. Passionate about public policy research investigating the impact of algorithmic decision making systems.

EDUCATION

2017 **— B.A. Economics,** University of Pennsylvania, (GPA: 3.5/4)
- Minor(s): Statistics, International Relations

2025 **— M.S. Public Policy Management - Data Analytics,** Carnegie Mellon University

EXPERIENCE

UPenn Computational Social Science Lab **Philadelphia, PA**
Research Data Engineer **Apr. 2022 -**
- Spearheaded cloud resource utilization and open-source software development standards for diverse research use cases across the lab
- Implemented **AWS** based automated research pipeline for streaming, multi-terabyte TV transcript text database, including exploratory analysis, data architecture design, and feature engineering methods
- Developed **Selenium** based web scraping toolkit to simulate human activity and gather user data from Youtube

Point72 Asset Management **New York City, NY**
Long/Short Energy & Industrials Portfolio Data Scientist **Sept. 2019 - Apr. 2022**
- Developed regression based models **(sklearn, scipy)** to interpolate missing values and identify anomalies across thousands of individual Texas Permian Basin oil wells, and forecasted aggregate monthly production
- Developed novel feature engineering methods over large datasets, such as credit card transaction panels, to supplement investment decisions with signals correlated with company and sector KPIs
- Delivered real time analytics using event driven ETL job orchestration based on **Apache Airflow** and **AWS (Lambda, SNS)**, used **Spark** for scaling feature engineering and model estimation methods

AllianceBernstein Asset Management **New York City, NY**
Investment Research Associate, Fixed Income **June 2017 - August 2019**
- Contributed to portfolio construction and risk discussions with weekly data analysis and commentary
- Maintained R and Python libraries aimed at reproducible research methods, and authored documentation explaining output of proprietary Generalized Additive Model **(mgcv R library)** based bond pricing model
- Developed and maintained novel market liquidity signals from streaming bid/offer orderbook database, utilized by traders to assess real time supply and demand levels in subsets of the market.

PUBLICATIONS

- H.Hosseinmardi, **M. Rivera-Lanas,** A.Ghasemian, M. Ribeiro, R. West, D. J. Watts "Disentangling YouTube's Recommender System from User Preferences with Counterfactual Bots", ***Work in Progress***

CORE COMPETENCIES

Scripting: Python, R, JavaScript, Scala, Bash **Devops:** Git, Airflow, Docker, Jenkins **Cloud:** AWS (S3, EC2, Lambda, SNS, SES, Glue) **Database:** PostgreSQL, MongoDB **Distributed Computing:** Spark **Web Development:** Flask, D3, Hugo, Plotly/Dash **Web Scraping:** Selenium

SELECTED PROJECTS

Volunteer Developer — Statisticians without Borders **Nov. 2020 - March 2021**

Working with non-profit, India Excellence Forum (IEF), our team achieved development of an AWS based web dashboard used by municipal health organizations in India to evaluate COVID-19 related epidemiological model results. See citation in resulting project article here.

Longitudinal System Monitor **2023**

Linux tools like htop, du, and ps monitor system and process status and resource utilization at a point in time. This project provides a longitudinal perspective for these measures. Tool provides a CLI entry point that manages data collecting services and a lightweight web dashboard.

Coursera: Functional Programming Principles in Scala **2021**

Covered principles of functional programming, program design using immutable data structures, and combining functional and object-oriented programming principles. See course certificate here.

LANGUAGES

English (Fluent), **Spanish** (Fluent), **French** (Working Proficiency)