

# GEORGE LEWIS

220 East 18th Street, New York, NY

ghl227@gmail.com

(203) 918-4859

<https://spontaneoussymmetry.com>

**Summary:** Data scientist and software engineer experienced in building and deploying statistical models, designing efficient data pipelines, and creating effective developer tools.

## EXPERIENCE

---

### LendUp

2013 - Present

Head of Risk and Analytics

Principal Data Scientist

- Founded the Risk and Analytics Team, later the Data Science Team. Designed and implemented the risk and underwriting program, growing company's loan portfolio by a factor of 200 while decreasing loss rates and achieving unit profitability. Led the team from seed round funding through series B.
- Designed, trained, deployed, and monitored statistical models using logistic regression, generalized additive models, decision trees, and random forests to evaluate the credit risk of all applicants and to identify and mitigate fraud.
- Grew the Data Science Team to more than 10 members. Established best practices for building statistical models. Developed techniques and tools for analytics and monitoring. Enforced high software standards and helped develop a culture of strong programming practices within the Data Science Team.
- Built production-quality software in Java, Python, and Clojure to implement our risk program, serve and score models statistical models, and to ingest and process data in real time.
- Recruited and trained the original Data Engineering Team. Collaborated with Data Engineering Team to build an ETL framework, jobs, and pipelines for supporting machine learning and analytics using Python, Scala, Airflow and Spark. Developed tooling to enable streamlined model deployment for data scientists.

### ATLAS Experiment, CERN

2008 - 2013

Graduate researcher on the Large Hadron Collider (LHC)

- Built sophisticated models using custom likelihood functions and modern frequentist inference techniques to produce world accurate measurements of fundamental particle properties.
- Developed and maintained a statistical framework in C++ that implemented novel modeling techniques and was used extensively throughout a 3,000 person collaboration.
- Used parallel batch computing to analyze petabytes of data distributed worldwide across data centers.
- Wrote production C++ code to select interesting collisions of the 10 million created per second.
- Enforced coding standards, conventions, and best practices across a large C++ and Python code base.
- Recipient of National Science Foundation US LHC Graduate Student Support Award.

## EDUCATION

---

### Ph.D. in Experimental High Energy Particle Physics

2013

New York University

### B.A. in Physics and Mathematics

2007

Columbia University

## SKILLS

---

**Modeling:** Regression, decision trees, bagging, boosting, neural networks, MCMC, likelihood-based inference.

**Languages:** Python, Java, SQL, Scala, Clojure, C++. Some experience with R, C, Cython.

**Tools:** Scikit-Learn, Pandas, Scipy, Jupyter, Postgres, Redshift, BigQuery, Spark, Airflow, AWS, Weka, XGBoost, Matplotlib, Seaborn, Flask, Play, Ring, Tensorflow, Theano, Tidyverse, PyMC3.

---

For a sample of my work and code, please see: <https://spontaneoussymmetry.com/work>