

Overview:

Experienced four-person remote data science and engineering team with expertise in financial applications of machine learning, Bayesian statistics, streaming big-data services, DevOps, and software engineering. Domain expertise in finance, travel, and weather. Worked together over three years in a startup environment.

Team:

1. **Sam Zimmerman** ([linkedin](#), [resume](#)) : Cofounded [Freebird](#) in 2015. Supported fundraising \$16.5M and recruited and managed a full-stack 15-person tech team (engineering, product, design, security, ds). Average tenure of the tech team was nearly three years. Eight years of fintech experience spanning quant finance, machine learning R&D, consumer tech, and risk management. Research at MIT/NASA. Fintech advisor. Most capable in Python (some Java/Scala).
2. **Paul Kernfeld** ([website](#), [github](#), [linkedin](#)): Principal Software Engineer who architected and implemented the majority of Freebird services. Passionate about machine learning, distributed systems, and all things back end. Specializes in Scala, Java, Python and Rust. Technical culture leader.
3. **Max Livingston** ([website](#), [github](#), [linkedin](#)): Director of Data Science with six years DS experience (previously in research at NYFed). Deep expertise in Bayesian modeling and recommender systems. Lead risk modeling research, strategy, development, and communication at Freebird, with an approach that combines modern techniques with traditional actuarial practices. Capable developer in Python, Scala, and Java. Highly versatile data scientist.
4. **Camen Piho** ([github](#)): Data Scientist and Software Engineer with two years of experience. Brown Classics/Math. Built and maintained production services in Python and Scala. Exceptional communicator. Goes toe-to-toe with the rest of the (more senior) team.

Artifacts:

- Blog post: [Solving Flight Monitoring with CRDTs](#)
- Blog post: [Analyzing a Billion Files in S3](#)
- Personal Notebook: [Analyzing the Green Line Extension with OpenStreetMap](#)
- Talk: [Scaling Bayesian Inference](#)
- Personal notebook: [Using Integer Programming to Win your Office NFL Survivor Pool](#)
- Talk: [Deep Learning and Real-Time Flight Prediction](#)
- Personal blog: [Paul's personal blog](#)
- Published paper: [Prediction and Uncertainty Quantification of Daily Airport Flight Delays](#)
- Afternoon Team Hackathon: [Covid19 and Air travel](#).

Strengths:

1. **Velocity:** A Kanban process and practical attitude have enabled the team to produce and maintain an outsized list of deliverables.
2. **Code quality:** Rigorous code review processes and engineering practices allow each team member to develop, deploy, and maintain production services.
3. **Independence:** Each member of the team has a multidisciplinary skill set composed of data science, engineering, and dev ops, allowing them to work without external support.
4. **Bayesian knowledge:** The team specializes in using Bayesian techniques such as MCMC to capture tail risk.

Capabilities:

1. **Design** models collaboratively to solve new problems
2. **Prototype** models quickly using **pandas**, **PyMC3**, **Postgres**, and **Presto**.
3. **Validate** models with scoring functions appropriate for **Bayesian** analysis
4. **Deploy** models using **Docker**, **Python**, and **Scala** best practices in **AWS**.
5. **Iterate** on models following **Kanban** practice
6. **Maintain** data pipelines using monitoring tools like **Datadog** and **Papertrail**
7. **Present** findings externally (executives, partners, conferences, journals, blogs...)

Accomplishments:

Modeling: Developed and implemented a novel probabilistic mathematical risk framework and models for estimating, forecasting, fine-tuning, and experimenting with the financial risk of a completely novel insurance-like product. Performed well over operating history (consistently <7% aggregate error) and positively evaluated by third-party actuarial consulting firm.

Engineering: Built and sold a proprietary air travel data feed product for a third party, including historical and live components. Setup TB-level big data streaming and querying pipeline.

Contact Us [Here](#).