

Sam Kritchevsky

Location: New York, NY

[GitHub](#) | [LinkedIn](#) | sam.kritch@gmail.com | 336-414-8694

ABOUT ME

I'm a software engineer and data generalist, with 5 years of experience across a modern cloud-based data stack.

TECHNICAL SKILLS

- Languages** : Python, Scala, SQL, some Javascript / Typescript
- Technologies** : PostgreSQL, Redshift, DBT, Redis, Spark, Flink, RabbitMQ, Elasticsearch, Docker, Nomad, Datadog, Grafana, AWS, EMR, Lambda, Kinesis, S3, Neptune, Jupyter, Dagster, Luigi, Akka
- Roles** : Data Science, Data Engineering, Data Platform, Backend, Analytics

EXPERIENCE

Seatgeek New York, NY
Sr. Software Engineer, Data Platform, Tech Lead 2020-2021

Tech lead of team of six engineers. Projects I was personally involved with included:

- Rewrite of our in-house event ingestion service in Flink + Kafka. Our primary goal was scalability—the earlier RabbitMQ-based system was very unstable—but we also added a number of features:
 - * Fanout to multiple downstreams: front page recommendations, push notifications, and hourly batch loads into Redshift
 - * Realtime sessionization implemented as a stateful Flink job
 - * Validation against JSONSchema with configurable alerting in Datadog

I advocated for the project, designed the system, and implemented most of the core validation and transformation jobs. Others on the team deployed Flink and Kafka, and built the sessionization job.

- Deployment of Amundsen for an OSS data-discovery and documentation platform. We built an AWS Neptune backend and added a number of extensions to ingest lineage metadata from internal services.
- Standardized Jupyter Notebook infrastructure based on Docker, Papermill, and NBViewer. This let users use notebooks in production workflows, and supported local development in an identical environment to production.
- Extensive improvements to the core data stack: [Druzhba](#) (in-house DB-to-DB ETL tool, now open-source) , Kinesis, DBT, Luigi (heavily modified), and Redshift.

Sr. Data Scientist 2019-2020

Data Scientist II 2018

I specialized in recommendation signal and user activity data, working in Spark, SQL, and Python. Projects included:

- Algorithms for the weekly newsletter, cart abandonment and price drops notifications, in Spark + SQL. The newsletter also tripled the revenue attributed to the weekly newsletter immediately, to about \$700k/week.
- Models to determine event and performer popularity in Keras for event recommendations and search.
- Customer entity-linking algorithm to deduplicate devices and users in marketing attribution and funnel KPIs.
- Ongoing involvement in the design and measurement of recommendation, search, and CRM marketing systems.

Software Engineer, Discovery Sept 2016-2017

- I built features for search, marketing, and frontpage recommendations in Python (Tornado), Elasticsearch, Postgres, Redis, and RabbitMQ.

Greater Harlem Coalition, Volunteer 2022-

- Contributed writing and data analysis to advocate for a just distribution of social services. Code [here](#).

EDUCATION

University of Wisconsin Madison, WI
Ph.D. Physics (coursework) 2013-2015

University of North Carolina Chapel Hill, NC
B.S. Physics 2012