

Erin SHELLMAN

Data Scientist | Leader

📍 Seattle, WA 🌐 www.erinshellman.com

✉ erin.shellman@gmail.com

in linkedin.com/in/erinshellman 🐙 github.com/erinshellman



I'm a data scientist with a track record of shipping data-driven software products. I bring scientific reasoning and technical depth to every project, with a focus on analytical processes and automation that make advanced methods like machine learning and AI practical and reliable. I work best in environments that prioritize rapid, incremental delivery and value direct communication. Throughout my career, I've emphasized continuous learning, presenting, and sharing knowledge with others. I love data science because it demands constant adaptation and creative problem solving.

EXPERIENCE

January 2026
Present

Data Science Architect, DEVOTED MEDICAL, Seattle, WA | Waltham, MA

Devoted is a Medicare Advantage provider that uses technology and data to deliver better care to seniors. As a Data Science Architect, I focus on analytics engineering, forecasting, and AI-driven development practices.

- Build and maintain analytics engineering pipelines with dbt, Airflow, and Snowflake.
- Own the daily call center forecast model that drives staffing decisions for customer care.
- Own the employee scorecard used to track customer care outcomes across the organization.
- Lead AI-driven development on the team, including custom Claude Code skills and agents that accelerate our development workflow.

dbt Airflow Snowflake Python Claude Code Codex GitHub

July 2025
December 2025

Director of Data Science, KING COUNTY REGIONAL HOMELESSNESS AUTHORITY, Seattle, WA

Individual contributor role (non-managerial) driving data science strategy and execution for the agency responsible for coordinating regional homelessness response. With a small team and limited resources, focused on building sustainable open-source analytics infrastructure and improving data fluency across departments.

- Deployed and maintained cost-efficient, open-source analytical tools.
- Prepared analyses and reports in response to questions from journalists and policy-makers and to support KCRHA leadership in legislative sessions.
- Drafted data science roadmap including organizational structure of data team, and cross-departmental initiatives.

Azure R JupyterHub Superset GitHub

September 2024
January 2021

Senior Director / Head of Data, GINKGO BIOWORKS, Seattle, WA | Boston, MA

Ginkgo Bioworks is a synthetic biology company applying microbial engineering and industrial fermentation at scale. Built and led a multi-team data organization (data science, platform engineering, enterprise analytics, and data product & strategy) delivering analyses, software products, and infrastructure across the company.

- Defined the vision for data at Ginkgo and scaled the team from 3 to 21 employees.
- Drove quality and continuous improvement in laboratory services by establishing templated reports, gold standard data sets, and formalized data processes/best practices.
- Partnered with the software organization to implement vendor solutions that improve instrumentation and data quality.
- Implemented vendor and in-house tools to strengthen regulatory compliance (SOX, finance, and ops).
- Launched a powerful, custom web application that allowed scientists to explore and retrieve datasets, build custom analyses, and download presentation-ready visualizations.

Snowflake dbt Airflow SQL GitLab Jira

December 2020 December 2018	Senior Director, Data Science and Analytics, NORDSTROM, Seattle, WA Directed a 40-person organization delivering data products for merchandising, e-commerce, operations, and executive strategy. ➤ Led several Analytics functions including A/B testing and optimization, data engineering, digital applications (recommendations, search and browse, content optimization), and corporate strategy (scaled events, traffic estimation, store strategy). ➤ Launched two merchandising teams supporting the Nordstrom Product Group with forecast-driven intelligence including price and markdown optimization. ➤ Built a digital operations program to reduce fraud and improve customer care, automating fraud/abuse decisions in call centers and forecasting call volume for scheduling. ➤ Developed a SQL curriculum for non-technical audiences outside the CoE. R SQL GitLab Jira event-driven systems
November 2018 October 2015	Senior Data Scientist / Engineering Manager, ZYMERGEN, Seattle, WA Emeryville, CA Zymergen was a biotechnology company at the intersection of data science, microbiology, and materials science. As the founding data scientist and manager, I established the team's strategy and led execution. ➤ Researched and implemented a data processing pipeline supporting high-throughput screening platform with analytical automation by creating data pipelines to clean, normalize, and analyze data with pandas and Airflow. ➤ Developed and patented predictive models to recommend genetic edits likely to result in high-performing strains. Patent ID : 10047358 ➤ Built dashboards to monitor operational metrics in near real-time with Superset. ➤ Built and lead distributed teams as I grew the data science organization from 1 to 14 people including launching an office in Seattle. Python Pandas R tidyverse Apache Airflow Superset AWS GitHub
May 2018 November 2014	Program development consultant Instructor, UNIVERSITY OF WASHINGTON, Seattle, WA ➤ Developed a 3-course series in applied deep learning covering fundamentals through NLP, image recognition, and reinforcement learning. ➤ Developed and taught a 10-week machine learning course for business intelligence covering regression, decision trees, association rules, and clustering. TensorFlow Deep Learning R tidyverse caret arules
October 2015 March 2015	Research Scientist II, AMAZON WEB SERVICES S3, Seattle, WA First member of an embedded research team with a mission to inform critical S3 engineering decisions with data. ➤ Mined and aggregated terabytes of S3 web server request logs using Scala Spark in the AWS ecosystem, specifically EMR, S3, EC2, and RedShift. ➤ Applied a broad set of analytical techniques to inform engineering decisions including segmentation of customer usage patterns, survival analysis to estimate time-to-network-failure, and disk capacity / latency simulation. Scala Spark R tidyverse AWS Survival Analysis simulation study
March 2015 March 2013	Data Scientist II, NORDSTROM DATA LAB, Seattle, WA The Data Lab is a small team of data scientists and software developers who build data-driven software products. Our focus was high-risk, high-reward projects with the ultimate goal of making the best customer experiences in retail. ➤ Designed and developed novel product recommendation algorithms, including recommendations based on real-time customer behavior. ➤ Filed Nordstrom's first technology patent for recommendations based on stylist domain expertise and color harmony (Patent ID : 480467). ➤ Built a personalized algorithm and email campaign to remind customers to replenish beauty products based on individual consumption patterns. ➤ Mentored an intern who became a permanent member of the team. Python Pandas Seaborn R tidyverse AWS

EDUCATION

2012	PhD Bioinformatics — University of Michigan, Ann Arbor, MI
2008	MS Biostatistics — University of Michigan, Ann Arbor, MI
2006	BA Economics, Evolutionary Biology — Case Western Reserve University, Cleveland, OH