

# Matthew McAnear

Senior Data Scientist

Resume

July 2023

 Ann Arbor, MI  
 +1 423-277-0411  
 mcanearm@gmail.com  
 mcanearm

## Summary

Experienced full stack data scientist adept at using Python, R, AWS, and SQL to create cost-effective and scalable solutions to business problems.

## Professional Experience

2021 - 2023	<b>Point Predictive</b> <i>Senior Data Scientist</i> <ul style="list-style-type: none"><li>➤ Fit statistical models for predicting early loan default and chargeoff for auto lenders in both object oriented and functional programming paradigms.</li><li>➤ Wrote ETL pipelines for securely loading sensitive lender data (PPI) into multiple database environments, improving load times from 12 hours to less than 10 minutes.</li><li>➤ Implemented DynamoDB-backed remote key-value store to improve AWS Lambda cold start times and move large files out of application layer.</li><li>➤ Wrote performant Postgres stored procedures for real-time calculation of historical borrower data, returning business-critical aggregate statistics in sub-200ms query times.</li><li>➤ Produced a PostgreSQL-based fuzzy matching probabilistic record linkage algorithm for retrieving prior credit applications of incoming borrowers, reducing query times by 50%.</li><li>➤ Designed and automated Snowflake queries to detect model distribution drift, saving 5 hours per week in manual performance monitoring and proactively identifying model miscalibration for clients.</li></ul>	San Diego, CA (Remote)
2020-2021	<b>Clear Capital</b> <i>Senior Machine Learning Engineer</i> <ul style="list-style-type: none"><li>➤ Served as principal developer for Clear Capital's AVM, a system that produces 150 million new predictions and 150GB of new data each week.</li><li>➤ Built a Bayesian hierarchical model and accompanying Docker-based webser-vice to predict home complexity</li><li>➤ Spearheaded the creation of a quantile-regression error prediction proce-dure and accompanying Java implementation to recalibrate model prediction intervals.</li><li>➤ Implemented Amazon Quicksight dashboards for model validation at multiple levels of analysis, including each individual home in the United States.</li><li>➤ Designed, launched, and administered Redshift data warehouse to power machine learning models, ETL workflows, and ad-hoc analytical queries.</li><li>➤ Created custom PySpark classes and ETL pipelines for model fitting and predic-tion.</li></ul>	Reno, NV
2017-2019	<b>Clear Capital</b> <i>Data Scientist II</i> <ul style="list-style-type: none"><li>➤ Led team of five to streamline original AVM model, reducing build time and costs by over 90%</li><li>➤ Improved AVM performance from last place of 31 vendors to industry leader in 6 months (based on absolute mean prediction error), leading to annual revenues of over \$1M.</li><li>➤ Designed and deployed an S3 and DynamoDB backed application that manages 35+ terabytes of MLS photos and their metadata.</li><li>➤ Promoted to Senior Machine Learning Engineer in 2020.</li></ul>	Reno, NV

2015-2017	<b>Clear Capital</b> <i>Data Scientist I</i> <ul style="list-style-type: none"><li>➤ Built an automated valuation model (AVM) cascade on commodity hardware to predict home prices using distributed, high performance R and PostgreSQL.</li><li>➤ Created custom mixture methodology based on empirical and historical model performance for final value prediction.</li><li>➤ Built webservice in Python, Flask, and AWS Lambda for serving AVM model predictions and internal company data through a RESTful interface.</li><li>➤ Promoted to Data Scientist II in 2017.</li></ul>	Reno, NV
2014-2015	<b>Seer Interactive</b> <i>Data Scientist</i> <ul style="list-style-type: none"><li>➤ Designed and carried out web-based experiment on domain recognition using multivariate hierarchical regression.</li><li>➤ Supported analytics account managers and external clients through project planning and automated reporting.</li></ul>	Philadelphia, PA

## Tools

### Languages

- R, Python, bash, SQL

### Technologies

- AWS, S3, RDS, Lambda, Cloudformation, DynamoDB, Redshift, Snowflake, Linux, Docker, PostgreSQL, Quicksight, Looker, Slurm, Spark, Sagemaker, REST, Flask

### Packages

- scikit-learn, PyMC, numpyro, jax, pandas, numpy, ggplot2, pyspark, PyTorch, sqlalchemy

### Algorithms

- RandomForest, Regression, Clustering, XGBoost, MCMC sampling, Variational Inference, Bayesian Deep Learning, Hierarchical Modeling, Record Linkage

## Education

2022-Present	<b>M.S., Applied Statistics</b> University of Michigan	Ann Arbor, MI
2013-2014	<b>M.S., Nonprofit/NGO Leadership</b> University of Pennsylvania <ul style="list-style-type: none"><li>➤ Donald J. Deutsch Endowed Graduate Fellowship</li></ul>	Philadelphia, PA
2009-2013	<b>B.A., Mathematics &amp; Economics</b> Bucknell University <ul style="list-style-type: none"><li>➤ <i>Magna cum laude</i></li><li>➤ National Merit Finalist Scholarship</li><li>➤ Dean's Scholarship</li><li>➤ Omicron Delta Epsilon Economics Honor Society</li></ul>	Lewisburg, PA