Zachary **Barry**

Data Scientist · M.S. Applied Mathematics

(541) 678-2898 | zackbarry13@gmail.com | ZackBarry.github.io | github.com/ZackBarry | linkedin.com/in/zackbarry

Summary

Data enthusiast with a passion for big data, machine learning, and well-documented code. 3+ years experience applying computational methods to numerical problems, focusing on scalability and reproducability.

Technologies

Languages	R (tidyverse, sparklyr, mlr3, shiny), SQL, Python (for ETL), Scala (Apache Spark)
Machine Learning	Regression, KNN, K-means, SVM, Bagging, Boosting, Random Forest, XGBoost, lightGBM
Skills	Data Cleaning/Wrangling/Visualization, Airflow, Docker, Google Cloud

Experience

G5
Bend, OR

Data Scientist
Oct. 2019 - Present

- **Attribution Platform** (R, Spark, Airflow):
  - Use Markov Chain Attribution to enable cross-channel advertising spend optimization, leading to 37% more conversions at 45% lower cost per conversion
  - Leverage Spark to handle 100 million distinct user paths (1+ billion observations) in a scalable manner
  - Implement Airflow DAGs to deploy models in a Google Cloud production context
- **Model Lifecycle**: (Docker, Kubernetes, MLflow)
  - Use Google Kubernetes Engine to deploy MLflow server for model parameter and metric tracking
  - Containerize multiple production R models for transition from Airflow to Kubeflow / AI Platform Pipelines
- **Customer Data Platform** (Scala, Spark):
  - Reduced run time and cost of production user model by 95% (6hr to 35min); maintained 99.98% accuracy
  - Delivered as an executable JAR for deployment in a GCP Dataproc context

G5
Bend, OR

Data Analyst
Jan. 2019 - Sep. 2019

- **Business Intelligence** (R, Shiny):
  - Reduced report creation time by over 80% while establishing a single source of truth for performance data
  - Created and maintained the back- and front-end of an in-memory performance benchmarking application
  - Applied Mann-Whitney U test and Global Moran's I test to identify statistically relevant comparisons
  - Encouraged adoption by providing use cases and technical documentation for internal training course
- **Operator Response Time** (R, Tidyverse):
  - Analyzed wave patterns to determine caller wait time for call scoring project (85% accurate within 3 sec)
  - Presented results as a POC Shiny application to internal stakeholders in Product Management

Northwest Mathematics, LLC
Bend, OR

Independent Contractor (C#)
Aug. 2018 - Nov. 2018

- Expanded a probabilistic model to support the concurrent simulation of slot machines across casinos
- Used Windows forms and C# package EPPlus to programmatically create excel spreadsheets for end users

Education

Oregon State University
Corvallis, OR.

B.S. and M.S. in Applied and Computational Mathematics — 3.96 GPA
Sept. 2014 - June 2018

- Researched numerical techniques to reduce solution error for systems of nonlinear PDEs (NSF grant)
- First student to complete Accelerated Masters in Mathematics (5 year program) in 4 years
- Courses in numerical linear algebra, finite element methods, computational tomography, and probability