Bryan Whiting

bryan.g.whiting@gmail.com | 203.822.4355

Python, R, SQL

BryanWhiting.com | GitHub | LinkedIn

SUMMARY | Data scientist with 8 years of experience building complex, large-scale data pipelines and machine learning models for inference and prediction. Experience running online A/B experiments, managing complex compute infrastructure, and confidently presenting to VP-level audiences. Passionate about innovation, creativity, tech, open source, engineering, leadership, and team building. Startup co-founder for 18 months.

Experience

Senior Data Scientist — Hopper, Hotels — Feb 2022-Present

Increased conversion rate by 5% and drove over $1.3M in annual revenue:

- Launched four recommendation iterations to millions of daily active users. Drove an increase in hotel booking conversion rate of 5% across all online experiments. Built several large-scale data pipelines to manage recommendations. **SQL, GCP**
- Optimized hotels pricing strategy by building and deploying a model on the cloud that processes 40 million transactions a day. **GCP, Python**
- Scaled up our competitor pricing awareness strategy and delivered monthly reports with actionable insights on how we compare to our customers. Built data pipelines and two dashboards to bring transparency to revenue, competitive pricing and purchase trends. **Google Studio**

Data Scientist, Engineering — Google, YouTube Music — Apr 2021 - Feb 2022

Built data-driven innovations that enhance music recommendations on YouTube:

- Designed a novel technique to **improve recommendations for music discovery for one billion users**. Used regression techniques and online A/B experiments to drive improvements. Presented findings to VP-level audience. **Processed, combined, and intelligently sampled billions of rows** across 15 data sources of video metadata and user-event logs. **R**, **SQL**
- Drove re-engagement with user-artist listening by pioneering novel techniques and finding signals in the noise of 80 billion events (120TB of data) over 90 days of YouTube history. **R, SQL**

Data Scientist, Engineering — Google Maps — Mar 2020 – Apr 2021

Increased volume of user-generated content (photos/reviews) and quality:

- Drove **10% growth in photo upload volume** and **5% growth in review contribution** from millions of users by providing insights and ideas for re-designing contribution UX and designing and analyzing **six online A/B tests.**
- **Drove awareness to stakeholders of the freshness/quality of Maps photo and review corpus by** systematically measuring data quality of billions of global locations and producing a daily-updating dashboard used by 20+ engineers. Presented 15+ times to key stakeholders. **SQL**
- Clarified to stakeholders how users were engaging with their product by segmenting user accounts into four behavioral groups using clustering, regression, and feature importances. **Python, Regression, Clustering**

Manager Data Scientist, Valuations — Capital One — Jun 2019 – Mar 2020

Rebuilt a five-year-old customer valuation framework from scratch in the cloud in six months using engineering best practices:

- Increased present value of credit card application program by **$80M by improving model predictions by 5%**. Replaced old modeling system (trained on 5-year-old data that had 115+ manual model adjustments) with freshly-trained models.
- Reduced model training cadence from **2 years to 2 weeks**. Built modeling platform capable of automatically retraining **12 machine learning models** that estimate the lifetime profitability. Implemented feature selection, hyperparameter tuning, and model validation techniques.
- Optimized across **decision trees and logistic regression** to select the best modeling methodology. **Python, XGBoost, H2O, Dask**
- Processed 20x more data (**1.6 billion rows, 1.7 terabytes**) by designing and developing a **data mining pipeline** that combines eight data sources using distributed computing techniques. **Python, pandas, Dask, SQL, Docker, Linux, EC2, Kubernetes**
- Coordinated across **20+ teams** to consolidate redundant technologies through rigorous prioritization and clear communication.
- Automated software packaging and delivery by implementing 15 code quality checks using CI/CD. **Python, Jenkins, Docker, GitHub**
- **Supervised three data scientists** by providing daily feedback, mentorship, code reviews, and development opportunities.
- Evangelized learnings to diverse groups of stakeholders and community by sharing dozens of resources and providing multiple trainings.

Principal/Mgr Data Scientist, Risk — Capital One — Jul 2017 – Jun 2019

Bryan Whiting

bryan.g.whiting@gmail.com | 203.822.4355

Python, R, SQL

BryanWhiting.com | GitHub | LinkedIn

Developed an end-to-end data processing and machine learning classification pipeline to predict customer behavior in real-time:

- Designed and developed a **data mining pipeline** from scratch that queried, cleaned, and combined **2 billion rows from 17 tables** into a single view of customer behavior. Developed pipeline as a **Python** package with over **26,000 lines of code** complete with logging, configuration files, code quality (unit tests, code coverage, etc.), and command-line tools. **Python, PySpark, SQL, shell scripting**
- Trained **1000+ machine learning models** and identified top 65 predictors of customer behavior. Developed feature selection, model selection, and model validation methods. **Python**, **H2O**, **Docker, Databricks, decision trees, logistic regression**
- Simplified insight-generation process by building an interactive web application to visualize and compare model results. **R, R Shiny**
- Coordinated implementation of real-time model with key stakeholders, business analysts, and software engineers. Presented findings weekly.
- Promoted to Manager within one year of hire; management considered it a 'rare event', citing exceptional leadership and innovation. **Jul 2018**

Consultant II — **Bates White Economic Consulting** — **Aug 2015 – Jul 2017**

- Saved colleagues **110+ hours per month** by developing 67 tools from scratch that were used over 20,000 times in six months. **Excel VBA**
- Forecasted LIBOR interest rates using 13,000 time series regression models and produced 30,000 plots to validate performance. **R, MATLAB**

Startup Co-Founder — **Novi Security** — **Jan 2013 – Aug 2014**

- Achieved **top 98.7% of all-time grossing** Kickstarter projects by selling $175,681 of product to 848 customers in 30 days (link).
- Raised **$560,000 of seed investment** by pitching business model to 15+ angel, venture capital, and private equity investors across four states.

Passion Projects — **Open Source Contributions** — **Nov 2016 – Present**

- Scraped 3,500 articles, performed **text analytics**, and used **GitHub actions** to post insights to social media on a daily schedule. **R, Docker**
- Created five **open source data science blogs**, sharing demos in R and Python with 23+ posts and **20,000+ all-time page views**.
- Scraped 15,000 used car prices to research and identify undervalued cars and optimal buying prices (see post). **R, dplyr, ggplot2**
- Coded **Bayesian hierarchical model** (Gibbs sampler) to forecast revenue of golf course tee times. **Python, R**
- **Build NBA game prediction** model that performs as well as Nate Silver's 538 Elo Model. **Python, XGBoost, CI/CD**

Leadership, Awards, and Service

- **Peer Bonuses, Google** (2020-2021): Recognized **five** times for exceptional collaboration and work under difficult circumstances.
- **Accelerated Talent Management, Capital One** (2019): Selected as one of 30 high-achieving data scientists with leadership potential.
- **Congregation Leader** (Feb 2018-Aug 2019): Served 200+ members, conducted 15 personal interviews per quarter, and mentored youth.
- **Crocker Innovation Fellow, BYU** (Jan 2013): Selected as one of 20 university students to receive $10,000 to pursue entrepreneurial ideas.
- **Volunteer missionary, Honduras** (Aug 2009-Aug 2011): Taught self-reliance principles to thousands of people across six cities. Gained empathy for those living in extreme poverty. Becoming fluent in a foreign language and culture.

Education and Skills

Education	**Brigham Young University (Provo, UT)**: B.S. in Statistics (April 2015), **M.S. in Statistics** (April 2015). Masters Project: Built statistical framework to identify outliers in count data with class imbalance (link).
Programming	**Python**: H2O, PySpark, pandas, Dask, XGBoost; **R**: ggplot2, Plotly, Shiny, dplyr, tidyverse; **SQL**; **git**; **shell scripting**; with prior experience in **MATLAB**; **Stata**; **Excel VBA**; **C.**
Cloud	**Google Cloud Platform**; **Google Studio; AWS**: EC2, S3, Redshift, EMR; **Linux**; Databricks; **GitHub**; Jenkins; Docker; distributed computing
Machine Learning	Regression, classification, decision trees, statistical methods (frequentist, Monte Carlo, Bayesian).
Projects/Examples	GitHub, Covid19 Dashboard, XGBoost model