BRYAN WHITING

bryan.g.whiting@gmail.com | 203.822.4355

DATA SCIENCE LEADER

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.

EXPERIENCE

SENIOR DATA SCIENTIST   HOPPER, HOTELS   FEB 2022-PRESENT

Drove $5M in incremental annual revenue and increased booking conversion rate by 5% through online experimentation:

- Drove **$5M in annual revenue by identifying user segments with sub-optimal pricing** and launching new models to correct mispriced hotels. Built and deployed API to process 40 million pricing requests per day. **GCP, VertexAI, FastAPI, Docker, Python**
- Increased hotels **conversion rate by 5% by building Hopper's first hotel recommender**. Designed and launched four recommendation systems using A/B testing. Deployed several large-scale data pipelines to manage recommendations. **BigQuery, Airflow**
- Scraped competitor prices to identify opportunities for competitors. Delivered monthly reports with actionable insights. Built data pipelines and two dashboards to bring transparency to revenue, competitive pricing and purchasing trends. **Google Studio, BigQuery, R**
- Developed custom statistical analysis tooling for A/B testing in BigQuery**.** Became Hopper's experimentation resident expert.

DATA SCIENTIST, ENGINEERING   GOOGLE, YOUTUBE MUSIC   APR 2021 - FEB 2022

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

- Pioneered novel recommendation techniques to **improve music discovery experiences for over one billion users**. Used regression models and online A/B experiments to discover what users like to listen to. **Presented research to multiple VP-level audiences.**
- **Processed, combined, and statistically sampled billions of rows** across 15 data sources of video metadata and user-event logs. Found actionable insights 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. Managed **six online A/B tests**. Gave product insights on user experiences and funnels, which helped improve how the UX was designed.
- **Drove awareness to stakeholders of the quality of Maps photo and review corpus**by systematically measuring data quality of billions of global locations. Produced 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. 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. Built python packages, managed CI/CD pipelines. **Python, pandas, Dask, SQL, Docker, Linux, EC2, Kubernetes, Jenkins**
- 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. Compared decision trees and regression models. **Python, XGBoost, H2O, Dask**
- **Supervised three data scientists** by providing daily feedback, mentorship, code reviews, and development opportunities.

PRINCIPAL/MGR DATA SCIENTIST, RISK   CAPITAL ONE   JUL 2017 – JUN 2019

Built real-time default prediction model that powers Capital One's no present spend limit business card, a credit card with no credit limit:

- 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 a 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**
- Engineered 500+ time-series features. Trained **1000+ machine learning models** and identified top 65 predictors of customer behavior. Developed novel feature selection, model selection, and model validation methods. **Python, H2O**, **Docker, Databricks, decision trees**

BRYAN WHITING

DATA SCIENCE LEADER

bryan.g.whiting@gmail.com | 203.822.4355

BryanWhiting.com | GitHub | LinkedIn

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 a 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**
- Built 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 hundreds 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, BigQuery**; **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