

JACOB COLVIN

(707) 628-2987

jbcollvin@gmail.com

Bay Area, California

[linkedin.com/in/jacobcolvin](https://www.linkedin.com/in/jacobcolvin)

DATA SCIENTIST

Applied Statistician with a Computer Science background

A Data Scientist with a Statistics and Computer Science background, passionate about compute and big data, seeking to answer the right questions using data.

- Three degrees in Statistics (Frequentist and Bayesian)
- Primary programming languages include R, SQL, Python, C++/C#
- Big data experience with Azure Cosmos DB, BigQuery, PySpark, Airflow.

EXPERIENCE

FANDOM

June 2021 – January 2026

Senior Data Scientist

Fandom is a commercial wiki provider focusing on games, fictional media, and their fans.

- Implemented data engineering pipelines as needed, including a keyword and hyperlink database extracted from 40+M files with 1.6+B hyperlinks stored on S3.
- Developed and pipelined dashboards evaluating page level classifiers, wiki level cross visitation, user engagement, and ad revenue metrics.
- Developed infrastructure and dashboards for A/B testing after identifying a user level randomization error significantly biasing experimentation results.
- Assisted with various ad hoc requests from advertising campaigns including trust and safety.

MICROSOFT

May 2013 – March 2020

Applied Statistics Engineer

Provided statistical support for Xbox hardware performance and reliability testing, including silicon yield and power consumption. Developed ETL pipelines for PowerBI dashboards for console usage, boot times, error rates, etc. Reported on monetary behavior of Xbox users with respect to social actions.

- Lead analysis efforts of a console manufacturing defect discovered postproduction affecting more than \$300M worth of product by combining telemetry usage data with accelerated life testing results to predict console

lifetime stress. Results predicted that accumulated stress would result in a negligible failure rate after 1-5 years, confirmed by monitoring telemetry after shipping.

- Identified a partitioning issue with the Xbox Cosmos cluster when the job queue was overflowing and implemented a fix which cleared the job queue using only 50% capacity. This provided efficiency gains worth \$200k+ annually and allowed Xbox developers to submit new jobs.
- Educated and helped engineers automate analysis and debugging efforts with an R package, including a console timeline tool that replaced two previous efforts.
- Developed a causal inference model relating unmeasured disposable income to subscription status, usage, and digital revenue.
- Worked with Director and Senior Director level executives across 3 different organizations (Devices, Silicon, and Experiences).
- Statistical Analyst (Protgent), May 2013 – December 2014

UC SANTA CRUZ

September 2007 – June 2011

Teaching Assistant

Led discussion sections for lower division statistics, calculus, and linear algebra.

GRAPHICS MICROSYSTEMS INC

September 2004 – September 2005

Data Analyst

GMI provides closed loop control equipment for industrial printing presses. Provided ad hoc analysis for executives, sales, and support staff.

- Developed a system of HTML reports summarizing printing press performance which was later productized for a large customer.
- Automated the collection of data from diverse sources including unstructured text files and repurposing product code to access binary files.

EDUCATION

Master of Science (MS), Statistics and Applied Mathematics, 2011

University of California Santa Cruz, CA

Boost Smoothing and Sequential Markov Random Field (SMRF) Models

Master of Science (MS), Statistics, 2007

California State University East Bay, CA

Bachelor of Science (BS), Computer Science, 2003

Bachelor of Science (BS), Statistics, 2003

University of California Davis, CA