

Gregory Faletto

Data Scientist | Statistician | Ph.D. in Statistics

Sherman Oaks, CA | 973-349-2153 | gregory.faletto@marshall.usc.edu | [linkedin.com/in/gregfaletto](https://www.linkedin.com/in/gregfaletto) | gregoryfaletto.com | github.com/gregfaletto

SUMMARY OF QUALIFICATIONS

3+ years full-time expert-level experience in analyzing data and building machine learning models with tested, performant code. Clear, open communicator who values prioritizing highest value-adds, focus, and hitting deadlines. Extensive knowledge of statistical & causal inference. Committed to ongoing learning (cutting-edge statistics along with technologies and other tools). Detail-oriented.

WORK EXPERIENCE

VideoAmp

Los Angeles, CA

Intermediate Data Scientist

July. 2023-present

- Use advanced but easily interpretable causal inference techniques to estimate lift on observational data. Crafted methods, proved consistency, derived uncertainty estimates, and wrote code. Collaborated across teams to better understand data and clients' priorities to direct my efforts.
- Created tool to compare viewing metrics under actual ad schedule vs. alternative to demonstrate the value of targeting.
- Produce & maintain documentation. Communicate complex data analysis results to stakeholders (presentations, writing).
- Proactive identify problems and clearly communicate early to solve. Identify and communicate blockers.

Part-Time Apprentice (Engineering department)

Mar. 2023-Jun. 2023

- Proposed a novel model using causal inference techniques (propensity score matching), survival analysis (accelerated failure time model), and classification to estimate a KPI which previously had no estimator. Implemented method in Python (PySpark).
- Use SQL to collect big data (Snowflake). Clean data and code Python models using Jupyter notebooks, Spark (PySpark).

Department of Data Sciences and Operations, University of Southern California

Los Angeles, CA

Research Assistant, Graduate Assistant Lecturer

Jan. 2019-May 2023

- Full-time data science researcher and lecturer Jan. 2021-May 2023; part-time research assistant Jan. 2019-Dec. 2020.
- Designed, coded, and tested novel methods for top venues (International Conference on Machine Learning, PNAS). (1) [ICML 2023] PRESTO estimates rare event probabilities, like probability of purchase after viewing an ad (github.com/gregfaletto/presto). (2) Fused extended two-way fixed effects is a panel data causal inference (difference-in-differences with staggered adoptions) ML method (arxiv.org/abs/2312.05985). (3) Cluster stability selection is a feature selection method for clustered features (github.com/gregfaletto/cssr-project).
- Created a novel recommendation system with a startup. Used matrix completion to estimate factors of an approximately low-rank matrix, and harnessed learned factors in a model estimating click probabilities, improving probability estimation by 5.7%.
- Taught 100 students \$375,000 worth of courses on analytics in Excel & JMP (SAS); making dashboards; communicating results.

Google

Los Angeles, CA (remote)

Data Scientist Intern (Chrome Analytics Team)

May 2021-Aug. 2021

- Designed, programmed simulation studies in Python to quantify flaw in prior method for estimating A/B test treatment effects. Crafted specific solution from problem description. Created a new method for treatment effect estimation & coded in Python.
- Coordinated with team, responding to and incorporating broad-strokes objectives, informal feedback, and formal code reviews.
- Reduced bias and MSE of treatment effect estimates by over 99% in simulations, while controlling Type I error rate much better.
- Submitted 4000 lines of documented, reviewed Python code to Google codebase implementing method and simulation studies.

ZipRecruiter

Santa Monica, CA

Data Analytics Research Intern

Jul. 2017-Jan. 2018

- Developed adaptive lasso logit model in R estimating probability a job seeker will apply to a job listing to infer preferences for listed perks. Improved over baseline accuracy by 6%. Estimated cash value of benefits for job seekers conditional on observables with model parameters.
- Accessed data via SQL in Periscope/Sinsense. Final deliverable: approximately 3500 lines of R code, 39-page written report.

Live Nation

Los Angeles, CA

Machine Learning Intern

Feb. 2017-May 2017

- Created time series model to predict future concert set lists of musical artists from available setlist data. Details on my GitHub.

PERSONAL PROJECTS

cssr R Package

Aug. 2022-Jan. 2023

- Created an R package for cluster stability selection (method I developed; see above). Wrote over 12,000 lines of code including over 3300 tests and user-friendly wrappers. Download, documentation, and description: gregfaletto.github.io/cssr-project.

2020 COVID-19 Computational Challenge (City of Los Angeles, Global Association for Research Methods and Data Science) Jun. 2020

- Won 2nd place in open-ended challenge to estimate “risk of exposure to COVID-19” in LA without specific directions.
- Interpreted daily neighborhood test results across Los Angeles as delayed, noisy measure of infections. Estimated real-time infections using data and available COVID research. Forecasted new infections in each neighborhood using a Poisson generalized linear panel data model. Converted forecasts to interpretable risk scores. Written report and code: grmds.org/2020challenge.
- Collaborated with partner on setting goals and dividing work. Presented solution with partner at IM Data Conference 2020.

Orange County R User Group Hackathon 2019

May 2019

- Won “Best Model.” Hackathon topic was water usage in California. With team, trained model establishing a significant association between county-level health outcomes and water pollutants. Link: gregoryfaletto.com/2019/05/19/our-entry-in-the-ocrug-hackathon-2019

TECHNICAL SKILLS

- Languages: Python (proficient; pandas, scikit-learn, matplotlib, etc.), R (proficient; tidyverse, caret, Keras, etc.), SQL (proficient).
- Software: Spark (PySpark), Snowflake, Jupyter/JupyterLab, Microsoft Excel, Periscope/Sinsense, Google Colab, Amazon Redshift, GitHub/GitHub Enterprise, ChatGPT (GPT-4).

EDUCATION

University of Southern California, Marshall School of Business

Los Angeles, CA

Doctor of Philosophy, Data Sciences and Operations (Statistics Group); Advisor: Prof. Jacob Bien

Aug. 2018-Aug. 2023

- 3.90 GPA. Selected coursework: regression & generalized linear models, econometrics, panel data models, statistical inference, deep learning theory & practice with convolutional neural networks, dynamic programming & reinforcement learning theory.
- Awarded “Top Reviewer” for 2023 International Conference on Artificial Intelligence and Statistics (AISTATS).
- Gave invited talk on a novel fairness proposal at the 2020 [Copenhagen Workshop on Algorithmic Fairness](#).

Washington University in St. Louis

St. Louis, MO

Bachelor of Arts, Physics

Aug. 2006-Dec. 2009

- 3.9 GPA. Graduated Phi Beta Kappa, College Honors. Dean's List every semester. National Merit Scholarship.