Greg **Faletto**

DATA SCIENTIST · PH.D. STUDENT

| ✉ | 🌐 gregoryfaletto.com/ | 🐙 github.com/gregfaletto | 🔗 https://www.linkedin.com/in/gregfaletto/

Employment History (Summary)

[Undisclosed Tech Startup]   *Los Angeles, CA*
RESEARCH CONSULTANT   *Oct. 2019 -*

- Developing an algorithm for a core business objective (can't disclose details under agreement).
- Working with adviser and communicating regularly with key startup data scientists.

ZipRecruiter   *Santa Monica, CA*
DATA ANALYTICS RESEARCH INTERN   *Jul. 2017 - Jan. 2018*

- Developed a model that predicts whether a job seeker will apply to a given job listing to infer preferences for forms of compensation.
- Chose observational study design and model form. Accessed data via SQL. Conceived outcome metric, constructed from raw data.
- Worked mostly independently under light supervision from ZipRecruiter and Dr. Sanjog Misra (Prof. of Marketing, U. of Chicago).

Education

University of Southern California Marshall School of Business   *Los Angeles, CA*
PH.D. IN DATA SCIENCES AND OPERATIONS (STATISTICS TRACK)   *Aug. 2018 -*

- Advisor: Dr. Jacob Bien
- Current research: developing a novel feature selection method; solving applied business problem with tech startup (see above).
- Coursework includes: mathematics of deep learning, generalized linear models, convex optimization, probability, stochastic processes, high dimensional statistics, mathematical statistics, time series, dynamic programming.

Washington University in St. Louis   *St. Louis, MO*
B.A. IN PHYSICS   *Aug. 2006 - Dec. 2009*

- Graduated Phi Beta Kappa and with College Honors in Arts and Sciences. Dean's List every semester.

Online Coursework (MOOCs)
                                                                        Jan. 2016 - Dec. 2017

- **Coursera:** "Machine Learning" (Andrew Ng/Stanford), "Econometrics: Methods and Applications" (Erasmus University Rotterdam).
- **Stanford Lagunita:** "Statistical Learning" (Trevor Hastie and Rob Tibshirani), "Databases: SQL," "Databases: Relational Algebra."
- **Datacamp:** "Intro to Python for Data Science," "Data Visualization with ggplot2," "Intermediate R," "Introduction to R."

Data Science Competitions

City of Los Angeles and RMDS
2020 COVID-19 COMPUTATIONAL CHALLENGE   *Jun. 2020*

- Team won second place. Interpreted general goals without specific directions.
- Used daily neighborhood test results across Los Angeles to estimate numbers of susceptible, infected, and recovered citizens in each neighborhood on each day. Forecasted new infections in each neighborhood a day ahead based on currently available information using a GLM panel data model (Poisson autoregressive count model with time- and neighborhood-based covariates). Finally, calculated per capita predicted new infections in each neighborhood and converted to risk scores.
- Full solution writeup and code available at grmds.org/2020challenge.

Orange County R Users Group
ORANGE COUNTY R USER GROUP HACKATHON 2019   *May 2019*

- Team won "Best Model." Over the course of two days, developed a model associating health outcomes in California counties with levels of water pollutants. More information at gregoryfaletto.com/2019/05/19/our-entry-in-the-ocrug-hackathon-2019/.
- Presented models at May 2019 meeting of Orange County R Users Group.

Publications

Proceedings of the National Academy of Science of the United States of America
"MEASURING THE PREDICTABILITY OF LIFE OUTCOMES WITH A SCIENTIFIC MASS COLLABORATION"   *May - Aug. 2017*

- Participated in Princeton University Fragile Families Challenge. Cleaned data and created models to predict three outcome variables in a training set with thousands more predictors than observations and a great deal of missing data. Co-author of article on project. Article: pnas.org/content/117/15/8398. Code and details of my contribution: github.com/gregfaletto/fragilefamilies.