

# PEYTON J. POLITEWICZ

Irvine, California | (814) 954-2250 | [peyton.james.pz@gmail.com](mailto:peyton.james.pz@gmail.com) | <https://pjamespz.github.io> | [linkedin.com/in/pjpmgmt](https://www.linkedin.com/in/pjpmgmt)

## TECHNICAL SKILLS

---

ML/AI/Statistics:	Statistical analysis, machine learning, regression, classification, predictive modeling/forecasting, supervised & unsupervised learning, NLP, descriptive statistics, ANOVA, experimental design, A/B testing, data mining, exploratory data analysis, data visualization
Python:	spaCy, scikit-learn, Keras, TensorFlow, gensim, PyTorch, BayesianOpt, XGBoost
R:	STAN, glm, lmer, caret, survival, clustMixType, tidyverse, ggplot2
Databases:	SQL, PySpark/Spark, Hadoop, Arrow, DuckDB
Workflow/Viz:	Git, Docker, PowerBI, Tableau, Amazon Web Services (S3, EC2, Redshift, Sagemaker, Textract)

## EDUCATION

---

UNIVERSITY OF CALIFORNIA, IRVINE, Irvine, California | **Master of Data Science**, Expected December 2024  
PENN STATE UNIVERSITY, State College, Pennsylvania | **B.S., Economics**, 2015

## EXPERIENCE

---

SOCAL CENTER FOR NONPROFIT MANAGEMENT | Los Angeles, California | 05/2024 - present  
**Graduate Data Engineering Researcher**, Research and Social Impact Team

- Mined, cleaned and collated datasets from the IRS, the Employment Development Department of California, and various municipalities into a single-source overview of the current state of the nonprofit sector in California.
- Automated data intake and refinement processes, ultimately feeding into a series of Tableau dashboards immediately deployed for novel internal research purposes by social impact analysts.
- Differentiated 'true' direct aid nonprofits from other 501(c)(3) organizations using k-prototypes clustering methods for mixed data.
- Performed survival analysis to compare longevity of direct aid nonprofits based on their self-reported Exempt Entity category designation.

UC IRVINE – DONALD BREN SCHOOL OF INFORMATION & COMPUTER SCIENCES | Irvine, California | 11/2023 - present  
**Data Science Research Analyst**, Center for Statistical Consulting & Biostatistics, Epidemiology and Research Design Unit

- Provided expert statistical & machine learning consultation, supporting other statisticians, MDs, biologists, clinical practitioners, geneticists, public health experts and epidemiologists in both academic and private settings.
- Designed an ETL pipeline around DuckDB to efficiently condense large datasets, reducing storage need by over 80%.
- Research engagements included clinical trials, public health studies, clinical operations analysis, longitudinal biomedical studies, electronic health record (EHR) data processing and analysis, grant writing and publication support.
- Supported drug development research by implementing novel imputation methods drawn from domain knowledge, overcoming challenges presented by underpowered studies.
- Compared hierarchical Bayesian modeling to mixed effects modeling in the study of socioeconomic determinants of childhood health across California, accounting for dependence at the county, school district, and school level.
- Studied surgical outcomes and timelines for diabetics with nephrological comorbidities, merging massive datasets from TriNetX encompassing longitudinal electronic health records of hundreds of thousands of subjects.
- Taught recurring seminars on statistical methods and database management for research staff, emphasizing the importance of statistical support and building relationships across units.
- Built out the Center's cloud knowledge base and architecture, leveraging AWS and Spark for big data analysis and secure storage of sensitive health data.
- Performed predictive modeling (linear regression, logistic regression, mixed models, survival analysis, time series, GLMs, random forest, deep learning), clustering, A/B testing, NLP, and data mining using Python, R, SAS & Stata.
- Statistical work included exploratory data analysis, estimation and inference (incl. Bayesian methods), power and sample size analysis, significance testing, ANOVA and experimental design.

PENN STATE UNIVERSITY | State College, Pennsylvania | 06/2019 – 12/2020

**Data Science Lead, AI Associates**

- Led program strategy, team recruitment, budget administration and project design, proposal, and management.
- Streamlined Undergraduate Admissions with AWS Textract and spaCy, saving each applicant over 40 minutes.
- Forecasted medical inventory using AWS SageMaker, predicting time-sensitive 'outs' with 93% accuracy.
- Synthesized data from the Department of Labor, census tract demographics, and GIS outputs to craft a dashboard for City of Philadelphia leadership using XGBoost modeling, predicting the impact of COVID office closures on public transit need.
- Created a bespoke search engine using gensim and the Wikipedia corpus for the university's 1700 donation fund opportunities, dramatically improving functionality compared to its predecessor per expert review.
- Studied sensor data from heavy machinery to model catastrophic failure and identify optimal emergency stoppage points.

SAINT FRANCIS UNIVERSITY (VIA TEKSYSTEMS) | Loretto, Pennsylvania | 02/2019 – 06/2019

**Senior Business Analyst, Office of Information Technology**

- Reviewed the University's IT practices while reporting directly to the CIO and VP of Academic Affairs.
- Identified inefficiencies and collated IT ticket response data into reports easily understood by non-technical leadership.
- Proposed process changes conserving approximately 40% of IT's resources and earning high level investment into transforming project management and ticket response approach, minimizing impact of staff skill gaps.
- Architected implementation of JIRA for effort tracking and designed initial dashboard design and dispatch schedule.

INTERACTIVE SERVICES | Dublin, Ireland | 08/2015 – 04/2018

**Business Analyst & Learning Manager**

- Led an internal process improvement team to build a bespoke development tool for eLearning software, demonstrating a 50% reduction in labor cost and production time with negligible delta in deliverable quality.
- Designed training software for clients including Deloitte, Facebook, Edwards Lifesciences, Walmart, and Sam's Club while directing a cross-functional team of 12 staff: developers, graphic designers, instructional designers, and testers.
- Quantified effects of upskilling and lowering employee turnover, providing detailed financial reports, proposing training rollout schedules, and forecasting business impact to earn project support from client leadership.

## PAPERS IN PROGRESS

---

Yiu A., Politewicz P., Guo Y., Cooper D., *Neighborhood Opportunity Associated with Physical Fitness in California Elementary Schools* (Pending submission, manuscript available on request)

Yiu A., Politewicz P., Salehi S., O'connell R., Chow E., *Socioeconomic Deprivation and Opportunity's Impact on Vaccine Decisions in Southern California* (Pending submission, manuscript available on request)

## ACHIEVEMENTS, COURSEWORK, PROFESSIONAL DEVELOPMENT

---

### SOCAL R USERS GROUP DATA SCIENCE HACKATHON

**First Prize: "Best Insight"**, for random forest modeling of English language integration of households per California microdata area

### STATISTICS COURSEWORK

Probability and Statistical Theory I & II  
Statistical Methods I & II (Generalized Linear Models)  
Bayesian Data Analysis

### COMPUTER SCIENCE COURSEWORK

Databases & Big Data Management  
Artificial Intelligence  
Machine Learning & Data Mining  
Graphical Models & Statistical Learning

### PROFESSIONAL DEVELOPMENT

All of Us Training and Certification  
REDCap Training  
TriNetX Training  
**SoCal RUG**: GitHub for Data Science  
**SoCal RUG**: Sleek Data Management with DuckDB

**JSM**: Statistical Genetics and Genomics: Fundamentals and Advanced Topics  
**JSM**: Practical Considerations for Bayesian and Frequentist Adaptive Clinical Trials  
**JSM**: EHR Data Processing and Analytics for Research and Real-World Evidence Discovery  
**JSM**: Introduction to Bayesian Nonparametric Methods for Missing Data