

STACY LEE

(646) 481 8838 📞
stacy-lee.github.io 🌐
linkedin.com/in/lee-stacy 📄
stacy.lee013@gmail.com ✉️

Languages	Proficient: Python, R, SQL	Basic: HTML5/CSS/JavaScript, C++, SAS
Libraries	Python: NumPy, Pandas, Sklearn, SciPy, StatsModels, Matplotlib	R: Tidyverse, Caret
Data Tools	Google Cloud Platform, Google Data Studio, Tableau, Trifacta, Azure, AWS, UNIX/Bash, Git	
Experience	Carat, a Dentsu Company New York, NY Analyst, Data Analytics <i>Data Science Impact Team</i> Sept 2019 - Present <ul style="list-style-type: none">Lead the strategy for creating a new composite scoring metric using principal component analysis, selecting test & control market pairs to measure advertising impact, & creating cost estimates to generate baselines of significant sample sizes for Fortune 100 financial clientImplement additive time series model using over 50 variables for causal impact analysis that received positive client feedback with approval & became an official team productImprove online ad performance by over 20% through optimal frequency & creative analysesOptimized Python data automation scripts & data flows in Trifacta joining over 5 disparate data sources to support Tableau reporting dashboards for Fortune 100 clientsCreate data visuals to investigate discrepancies and gain deeper insights on data trendsCollaborate with account team to ensure proper tagging taxonomy for reporting accuracyMentor team members on data strategies and methodologies that I introduced to client Ameren Champaign, IL Data Science Innovation Intern <i>Analytics Team</i> Jan 2018 - Dec 2018 <ul style="list-style-type: none">Led a 5-person team project for identifying individuals out of 1.4 million customers with high propensities of enrolling & saving in energy savings programs using demographic data to create a target list of recipients who will receive the promotional bill insert<ul style="list-style-type: none">Wrote custom scripts for exploratory data analysis to gain insights & formulate a data strategyImplemented random forest for 1% imbalanced class ratio and improved recall by 90%Applied Bayesian statistics with logistic regression for targeted marketing deliverableSurfaced and presented insights on customer trends through storytelling to stakeholdersResearched methods to replace the traditional utility pole health assessment process<ul style="list-style-type: none">Converted & reformatted raw data collected from sensors to facilitate analyticsUtilized clustering methods to find patterns in pole health based on movementAssisted with recommender system to estimate solar PV adoption for distribution planningMentored team members on machine learning, statistics, and programming in Python or RExecuted & documented ETL processes in Amazon Web Services EC2 for efficiency	
Education	University of Illinois at Urbana-Champaign M.S. in Statistics <i>Analytics Concentration</i> Dec 2018 B.S. in Civil Engineering <i>Systems Engineering Concentration</i> May 2017	
Competitions	Synchrony Financial Datathon Top 5 of 25 Teams with Best MSE Score April 2018 <ul style="list-style-type: none">Researched the impact of federal interest rates on home improvement spendingImplemented Elastic Net with Five-Fold Cross Validation for the final submission	
Projects	Machine Learning & Computational Statistics Projects [stacy-lee.github.io/ds/projects.html] <ul style="list-style-type: none">Monte Carlo method to identify Weibull PDF in Traffic Volume Counts 2012 NYC Open DataTime series with fourier seasonality on Walmart sales data achieved WMAE less than 1630Logistic regression for movie review analysis with NLP achieved AUC greater than 0.95 Algorithms (Written From Scratch) Random Forest, k-NN, Lasso Regression, Apriori	