NATHALIE JONES

(770) 508-4908
nathalie.jones@gmail.com Kennesaw, GA
linkedin.com/in/nathalie-jones711 github.com/njones738

Graduate Student at Kennesaw State University

Knowledge in numerical methods, parametric and nonparametric methods, binary classification, and logistical modeling, multilevel/hierarchical statistical modeling, machine learning, data cleaning and variable reduction. Proven ability in Python, R & SAS

EDUCATION

Kennesaw State University **Dec 2023**
Master of Science *(Expected)*
Applied Statistics & Analytics

Kennesaw State University **Dec 2021**
Bachelor of Science
Applied Mathematics
Applied Statistics & Analytics Minor

SKILLS & CODING LANGUAGES

- BigQuery
- Python
- R
- SAS
- SQL
- JAVA
- HTML
- Machine Learning
- Text Analytics
- Parametric Methods
- Multilevel Models
- Longitudinal Analysis
- SEM, EFA, CFA
- Feature Engineering
- Numerical Analysis
- Logistic Regression
- Nonparametric Methods
- Statistical Learning
- Microsoft 365
- Number Witch
- Git/Github

RESEARCH PROJECTS & WORK EXPERIENCE

CARES Research Lab – *Graduate Research Assistant* **Aug 2022 – Dec 2023**

- Conducted research related to foster and unaccompanied homeless youth in higher education
- Designed research hypotheses, surveys, data collection methods to aid in the labs research goals
- Worked under Dr. Sarah Young who guided me through the IRB application and research development

IHG Hotels & Resorts – *Data Science Internship* **Jul 2023 – Aug 2023**

- Collaborated with the Advanced Analytics team to report on the ROI of promotional campaign
- Measured ROI via incrementality and registration rate
- Networked with coworkers within and outside of the GIAD team

"Mapping Post-Secondary Transfer Rates" – *Coded in R, Python* **May 2023**

- Applied Graph Theory Learning Outcomes such as the clustering coefficient and network transitivity in Python
- Conducted research into the association between institutional transfer rates the number of schools within 100 miles of each other using the 2020-'21 CollegeScorecard dataset
- Used R to visualize findings with ggplot for use in presentation
- Awarded 3rd place Graduate Research Project at KSU's 2023 Analytics Day

"Predicting Email Click" – *Team Project coded in R, Python* **May 2023**

- Collaborated with a team to predict whether a member would click on an email
- Analyzed member behavior to build, test, and compare several models
- Presented insightful recommendations to IHG Hotel & Resorts

"Why Withdraw?" – *Coded in R, Python, SAS* **Dec 2022**

- Continued research into the CollegeScorecard with a longitudinal analysis
- Comments and Ratings left by reviewers were scraped from RateMyProfessor
- Conducted a sentiment analysis of the comments for use in the model

"Modelling User Chatbot Experience" – *Coded in R, Python* **Dec 2022**

- Analyzed user interactions with a chatbot using modern text analysis methods to create topic models
- Presented insightful recommendations to Southern Company for chatbot improvements
- Used Python libraries like SpaCy and Transformers for processing user inputs and creating topic models

"Does the Pell Grant Come with a Price?" – *Coded in R, Python* **Dec 2021**

- Continued my research into the CollegeScorecard with a spatial look at the institutions
- Spatially joined together the Census shapefiles with the CollegeScorecard dataset to analyze the association between debts accumulated by either an independent or dependent student
- Awarded 3rd place at KSU's 2021 Analytics Day and selected to present at: Posters on the Hill 2022 (1 of 88 posters), Harvard National Collegiate Research Conference 2022, and Posters at the GA capitol 2022

"Classification of Pell Institutions" – *Coded in R, Python* **Dec 2021**

- Continued in the pursuit of understanding the stewardship of low-income students by classifying Post-secondary institutions using a binary indicator created in "Access to Higher Education"
- Created several models and compared including XGBoost, PCA, Random Forest, and Logistical Regression
- Used R and Python to clean and structure the data to build models that classify Pell institutions
- Used the following packages: Tidyverse, magrittr, feather, and ggplot2 in R and sklearn, scipy, scikitplot, xgboost, category encoders, feather, matplotlib, plotnine, numpy and pandas

"Access to Higher Education" – *Coded in R* **May 2021**

- Conducted research on the CollegeScorecard dataset through parametric and nonparametric methods
- Studied the differences in US schools with either a majority or minority proportion of their student popluation receiving a Pell grant
- Used RStudio with tidyverse functions to clean the dataset of missing values and manipulate the data into usable information. Visualizations were created with ggplot methods

"Two-Layer Neural Network" – *Coded in Python* **May 2021**

- Created a Multilayer Perceptron in Python as a class object to predict whether an individual in 1994 earned $50,000 or more from Census data
- Used NumPy functions to define a class that creates a 0, 1, or 2-layer neural net from user input
- Used Pandas and NumPy to define functions, as well as the shuffle method from sklearn's utilities module

"Using Logistic Regression to Build Credit Scores" – *Coded in Python, SAS* **May 2021**

- Created a model to predict a customer's credit score by a binary predictor that indicated whether a customer was considered a credit risk
- Used SAS procedures to conduct a logistic regression analysis on the profitability of models created
- Used Python to decide for which parameters' missing values should be imputed and which should be excluded

"Get that Number" – *Coded in Python, SAS* **May 2021**

- Data from 'Using Logistic Regression to Build Credit Scores was used to deciding which parameters should have their missing values imputed and which should be dropped from the dataset
- Used Least-Squares Approximation, the first and second derivative test, and the bisection method to approximate the inflection point of an interpolated equation
- Visualized finding and methodology in R with ggplot for use in presentation