# Sameen Salam

sameen73.github.io | 336-259-3591 | ssalam@ncsu.edu

## EDUCATION

**INSTITUTE FOR ADVANCED ANALYTICS, NC STATE UNIVERSITY**
M.S. Analytics
Jun 2019 - May 2020
Raleigh, NC

**UNC CHAPEL HILL**
B.S. Biology, B.A. Music
Aug 2015 - May 2019
Chapel Hill, NC

## LINKS

Github:// **sameen73**
LinkedIn:// **Sameen-Salam**
Kaggle:// **sameensalam**

## COURSEWORK

**GRADUATE**
Statistics
Time Series Analysis
Text Analytics
Machine Learning
Linear Algebra
Cloud Computing Fundamentals
Optimization
Financial Analytics

## SKILLS

**Programming**
Python • R • SQL • MATLAB • Git

**Visualization**
Tableau • Excel • ggplot2 • matplotlib/seaborn • Jupyter

**Techniques**
Linear Reg. • Logistic Reg. • Time Series • PCA • Clustering • Bayesian Statistics • SVM • Survival Analysis • Random Forest • XGBoost • Hypothesis Testing • Latent Dirichlet Allocation • Sentiment Analysis • Feature Engineering

## EXPERIENCE

### CAPE FEAR COLLECTIVE | Data Science Intern
May 2020 - Present | Wilmington, NC | R

- Sourced, cleaned, and postured publicly available demographic, population, and economic data to create over 100 county-level socioeconomic indices
- Created cumulative performance metrics summarizing said indices for over 3000 counties
- Utilized K-means methodology to determine county-level peer groups
- Summarized insights in a comprehensive dashboard to inform future work on a regional scorecard

### PRIMROSE SCHOOLS | Practicum Project Tech Lead
Sep 2019 – May 2020 | Atlanta, GA | R + Python

- Enabled early differentiated support by creating linear and tree-based models to forecast school enrollment twelve months and three years after opening
- Identified areas of strength, weakness, and consumer interest for over 400 schools nation-wide by performing sentiment analysis of online school reviews
- Presented interactive Tableau dashboard to senior executives
- Spearheaded software development in an AGILE team as Technical Lead

### UNIVERSITY OF NORTH CAROLINA | Research Assistant
Aug 2017 – May 2019 | Chapel Hill, NC | R + Python + MATLAB

- Analyzed hundreds of image data sets for different strains of yeast
- Created a script that takes pixel data from images and constructs a noise floor histogram as part of a larger software package to take on deeper analyses
- Wrote technical reports and presented a research poster for dozens of fellow students and faculty in the biology department
- Edited lab protocols for image analysis to guide incoming undergraduates
- Collaborated with other members of the lab to develop and implement analytical action items

## PROJECTS

### STOCKX 2019 DATA CONTEST | Machine Learning Project
Feb 2020 – Mar 2020 | Raleigh, NC | Python

Utilized linear and random forest regression to predict how much customers would pay for a sneaker over its retail price. Created a dashboard containing visualizations demonstrating differences between sneaker brand performance in the market. Wrote a blog post summarizing process and findings.

### UCI BANK MARKETING DATASET | Machine Learning Project
Mar 2020 | Raleigh, NC | Python

Predicted client subscription to term deposit product using logistic regression, random forest, SVM, and XGBoost. Explored various strategies to handle target class imbalance such as SMOTE and up/down sampling.

### COVID-19 OPEN RESEARCH CHALLENGE | NLP Project
May 2020- Present | Raleigh, NC | Python

Cleaned and postured 75,000 research paper abstracts pertaining to COVID-19. Used LDA and K-means algorithms to sort abstracts into topics based on similarity. Algorithms will inform the creation of topic-level summaries and a COVID-19 search engine for research (in progress). Created visualizations of clusters and topics to better understand distribution of subject matter.