Trey Capps

(919) 306-8508 | trey.capps@gmail.com | LinkedIn | Github | treycapps.com

Education

North Carolina State University — Raleigh, NC
Bachelor of Science in Statistics, Minor in Mathematics — *Aug. 2020 – May 2022*

Sandhills Community College — Southern Pines, NC
Associate of Science — *Aug. 2018 – May 2020*
Committed 30+ hours per week to Sandhills Community College Golf Team

Experience

Statistical Analyst Intern — Aug. 2021 – May 2022
Institute for Transportation Research and Education — *Raleigh, NC*

- Designed experiments that leveraged regression techniques and hypothesis testing to quantify intervention effectiveness from over 10 years of time series data spanning multiple tables
- Implemented various time series techniques to determine high-impact violations which provided insights into ways to increase the effectiveness of future interventions
- Collaborated with a group of 3 GIS analysts to introduce and present program effectiveness to 20+ clients in a conference setting

Projects

Spot It | *Python, Google Cloud Platform, MongoDB, Airflow, Terraform* — May 2022 – Present

- Developing a web application to allow Spotify users to find similar songs based on various Reddit communities providing users with an integration of the two platforms
- Utilized Apache Airflow to create a DAG to schedule and monitor a data pipeline that ingests data from multiple APIs and loads the data to cloud storage
- Experimented with clustering and linear algebra techniques to generate similarity metrics that will provide users with song recommendations

Undergraduate Research | *SAS, R, ggplot2* — Jan. 2022 – May 2022

- Utilized logistic regression to identify key factors contributing to student success providing stakeholders with insights to make department specific policy changes
- Collaborated with a group of 4 students to present results at the research symposium

Patient Safety Topic Modeling | *Python, R, pandas, scikit-learn, nltk, matplotlib* — Oct. 2021 – Dec. 2021

- Utilized unsupervised learning to extract topics from device reports reducing the manual review time by more than one-half
- Suggested future device problem areas using time series analysis and forecasting methods

Technical Skills

Languages: Python (Pandas, Matplotlib, NumPy, Scikit-learn), SQL, R
Tools: Tableau, Google Cloud Platform, Git, MongoDB
Statistics: Time series analysis/forecasting, Regression, Classification, Experimental design, Clustering

Awards

2019 NJCAA Division III Men's Golf National Champion June 2019