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*

- Extracted and analyzed over a decade of time series data to create reports providing the GIS team with detailed insights into trends, seasonality, and correlation metrics surrounding both crashes and inspections
- Worked with minimal oversight to research and design experiments resulting in the derivation of key performance metrics for our client
- Implemented intervention models using time series forecasting and regression techniques to quantify the intervention performance of our client
- Analyzed numerous features of the inspection process to provide actionable insights to improve future program effectiveness
- Collaborated with the GIS team to introduce and present effectiveness to our clients using visualizations and dashboards

Projects

Spot It | *Python, Google Cloud, pandas, MongoDB, Airflow, Terraform, scikit-learn* — *In Progress*

- 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

Patient Safety Topic Modeling | *Python, pandas, R, scikit-learn, nltk, matplotlib*

- Extracted and preprocessed over 15 million device malfunction reports from multiple tables in the MAUDE database
- Utilized unsupervised learning to extract topics from device reports reducing the manual review time by more than one-half
- Provided future device problem areas using time series analysis and forecasting methods

Undergraduate Research | *SAS, R, ggplot2*

- 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

Technical Skills

Languages: Python (pandas, matplotlib, numpy, scikit-learn), SQL, R (tidyverse, caret, shiny)
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