

# Trey Capps

(919) 306-8508 | [trey.capps@gmail.com](mailto:trey.capps@gmail.com) | [LinkedIn](#) | [Github](#) | [treycapps.com](http://treycapps.com)

## EDUCATION

---

### North Carolina State University

*Bachelor of Science in Statistics, Minor in Mathematics*

Raleigh, NC

Aug. 2020 – May 2022

### Sandhills Community College

*Associate of Science*

Southern Pines, NC

Aug. 2018 – May 2020

Committed 30+ hours per week to Sandhills Community College Golf Team

## EXPERIENCE

---

### Statistical Analyst Intern

*Institute for Transportation Research and Education*

Aug. 2021 – May 2022

Raleigh, NC

- 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 and suggest ways to further improve future performance
- 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 Airflow 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, PySpark, pandas, R, Google Cloud, scikit-learn, nltk, matplotlib*

- Extracted and cleaned over 15 million device malfunction report cases using Spark SQL from the FDA's MAUDE database
- Utilized topic modeling to extract topics from the cleaned device reports reducing the manual review time by more than one-half
- Provided insights into 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