# Benjamin Cox

bjcbencox97@gmail.com | (931) 675-4534 | Petersburg, TN | LinkedIn | Github

## Education

**MS: Analytics** — *Huntsville, AL*
University of Alabama in Huntsville — *2023*

**BS: Mathematics** — *Birmingham, AL*
University of Alabama at Birmingham — *2021*

**Google Data Analytics Certification** — *Online*
Coursera — *2021*

## Experience

**Process Improvement Analyst Intern** — *Memphis, TN*
Frontdoor — *Summer 2022*

- Conducted preliminary research using the DMAIC method to evaluate the potential for standardization for the cash-in-lieu business process. The research resulted in the start of a pilot with an estimated **$4.8M** cost reduction.
- Performed complex queries(Joins, With statements) on large sets of data using Snowflake and Looker. Analyzed those queries to bring meaningful insights in Excel using Pivot Tables and Excel formulas. Developed those insights into clear visualizations using Power BI. Integrated those visualizations into the presentation with PowerPoint.

## Skills

**Data Science Skills:** Machine learning(Unsupervised and Supervised), Statistical analysis, Optimization(Linear, Nonlinear, Integer), Modeling, Exponential Smoothing, Database Management, Regression, Classification, Network Analysis, Data Visualizations

**Data Science Tools:** Snowflake, Looker, PowerBI, Visual Studio Code, Excel, Oracle, Tableau

**Programming Languages:** Python, R, Java, C, SQL

## Projects

### Potential Customer Prediction using Supervised Machine Learning

Training Data | Classification & Regression | Balancing Data

- Predicted the number of claims and the cost of claims a potential customer test dataset would have based on the accuracy scores (F1 and mean absolute error used) of a 75/25 independent test split of the training dataset for machine learning models (k-nearest neighbors, decision trees) including both classification and regression. Balancing techniques such as undersampling, oversampling, and SMOTE were used to improve F1 scores.
- Training dataset had ~60000 rows, but data binning reduced the number of deleted rows significantly. Data usage was maximized with hot encoding. Test dataset had ~7500 rows.

### Cyclistic Bike Share Case Study

Cleaning Data in R | Analyzing Data in R | Visualizing Data in R

- Cleaned, processed, and analyzed over .5GB of data in R, using the tidyverse, janitor, ggplot2, lubridate packages, and basic machine learning.
- Generated, in R, a line graph comparing the number of riders to the time of day, and generated bar graphs comparing the member riders to casual riders' behavior based on day of the week.

### Movie Rental Relational Database

Building a Relational Database | Performing Complex Queries | Scripting with PL/SQL

- Built a relational database from scratch using a figurative movie rental company's records.
- Performed complex queries on the data including join statements to display members who have not paid their rental fees, converting date data from a standard m-d-y format to numbers of weeks passed, and creating a PL/SQL block statement to display the top n movies in numbers of sales.