Blake Goehman

Austin, Texas • blakegoehman@gmail.com • 573-382-1917 • www.linkedin.com/in/blake-goehman/ • www.github.com/BlakeBigG

Software Engineer with specialization in Python, Java, machine learning, and web development. Looking for a role within tech that will challenge me and engage my problem solving, time management, and organizational skills.

EDUCATION

Master of Science in Data Science & Analytics – High Performance Computing Track
University of Missouri • Columbia, MO • May 2022 • 4.00 GPA
Bachelor of Science in Physics & Economics – Minor in Mathematics
University of Missouri • Columbia, MO • May 2020 • 3.47 GPA

EXPERIENCE

INFOSYS, LTD. **July 2022 – Present**
Associate Software Engineer *Austin, Texas*
• First three months consists of intensive training in Java, DBMS, consulting, and software development using agile methodology.
• Completed Java training, which was equivalent to a university object-oriented programming class using Java.
• Currently training in JavaScript, React, and Spring.

CHIPOTLE MEXICAN GRILL **May 2022 – June 2022**
Crew Member *Cedar Park, Texas*
• Worked in a fast-paced environment where it was necessary to balance multiple responsibilities concurrently.
• Worked on a team to achieve daily, weekly, and monthly goals.
• Gained experience in customer service.

SKILLS

Technical Skills: Python, Java, PostgreSQL, AWS, Google Cloud, R, Excel, git, Windows; some exposure to TensorFlow, PySpark, Linux, HTML, CSS, and JavaScript
Functional Skills: Software development, Agile, machine learning, data science, distributed systems, cloud computing, data mining, databases, data visualization, data structures, algorithms, mathematics, statistics, teamwork, leadership, communication, technical writing

ACADEMIC PROJECTS

Understanding World Population Dynamics (MS capstone project) • github.com/BlakeBigG/Masters-Capstone
- Technologies used: Python, R, SQL, PostgreSQL, Shiny, Jupyter, Linux, git
- Used Python, R, and panel data modeling to both predict changes in countries' birth rates and find variables that affect birth rate.
- Started with multiple, disorganized datasets and applied the entire data science process (data carpentry, exploratory analysis, visualizations, modeling, and presenting findings).
- Found that fixed effects models on groups of 3-5 countries produced best results, and that the most important factor in predicting birth rate is the age/sex breakdown of a country.

Automated Reddit Analysis using Google Cloud • github.com/BlakeBigG/GCP-Reddit-Analysis
- Technologies used: Google Cloud Platform, Python, Jupyter, Linux, git
- Designed and built a pipeline for collecting posts from Reddit's RSS feed, putting the text content through Google's natural language API, and loading the information into a SQL (BigQuery) database.
- Collected thousands of posts and found insights about the text content of posts.

Using Tweet Sentiment to Predict Price of Bitcoin • github.com/BlakeBigG/Tweets-Bitcoin-Case-Study
- Technologies used: Python, Jupyter, Linux, git
- Tried to use tweet count, popularity, and sentiment to predict the price of Bitcoin.
- Pulled hundreds of thousands of tweets containing the word Bitcoin and related terms using Twitter API, analyzed metadata and text sentiment, and trained a time series model.
- Failed to find a causality between tweets and the price of Bitcoin, and unfortunately did not become a Bitcoin millionaire.

Chicago Crime Database • github.com/BlakeBigG/Chicago-Crime-Database
- Technologies used: SQL, PostgreSQL, Python, SQLAlchemy, Jupyter
- Designed and created a database to store all the crime data in Chicago in a particular month.
- Drew an entity relationship diagram (ERD), loaded data into the database, and wrote somewhat complex queries (using joins, aggregation, nested queries, etc.).

Article Popularity Regression • github.com/BlakeBigG/Article-Regression
- Technologies used: Python, R, Anaconda, Jupyter
- Used various machine learning methods and models such as random forest and XGBoost to predict the number of times an article will be shared based on numerical data about the text, such as character count, sentiment score, and title character count.
- Used Gini importance to find the variables which most contributed to article popularity, which could be given to writers and editors to optimize their articles.