

# SOPHIA STEWART

## Data Scientist

sophia.marcell.stewart@gmail.com | 210-902-7028 | github.com/sophia-stewart | in/sophia-m-stewart

### TECHNICAL SKILLS

Python - SQL - Tableau - Applied Statistics - Pandas  
- Matplotlib - Seaborn - Machine Learning -  
Natural Language Processing - Data Storytelling -  
Git - Jupyter Notebooks - Spark - Anaconda

### EDUCATIONAL BACKGROUND

#### Codeup Data Science

*Certificate of Completion, March 2022*

Fully-immersive, project-based 22-week career accelerator that provides students with 670+ hours of expert instruction in applied data science. Students develop expertise across the full data science pipeline (planning, acquisition, preparation, exploration, modeling, delivery), and become comfortable working with data to deliver actionable insights to diverse stakeholders.

**San Antonio College** August 2018-August 2020

*Accounting Technician Level I Certificate, May 2019*

#### The University of Texas at San Antonio

46 Credit Hours August 2013-December 2014

### WORK EXPERIENCE

**Delivery Specialist, Deliv** May 2018 - July 2020

Provided delivery services while collaborating with a virtual team to streamline tasks from clients.

**Gaming Support Agent, VMC** Sep 2017 - Feb 2018

Provided technical support for video games.

**Member Service, Sam's Club** Aug 2016 - Aug 2017

Provided comprehensive support services for club members, worked with a team to achieve goals.

**Teacher Assistant, PreK 4 SA** Oct 2015 - June 2016

Supervised classroom of children and worked with fellow teachers to develop weekly lesson plans.

### ABOUT ME

I am an emerging data scientist with a commitment to excellence and constant improvement. My creativity in approaching problems leads me to success and my superior communication skills allow me to present my findings in a way that is clear and concise.

### DATA PROJECTS

#### Sound the Alarm - Wildfire Analysis March 2022

We analyzed features of wildfires in the U.S. from 1992-2018 to determine causes of wildfires and how wildfires have changed over time. With this data, we provided insights and recommendations for steps we can take to reduce the destruction caused by wildfires. We also gave focused attention to particularly large wildfires to provide insight on a more personal scale.

#### GitHub Language Classification February 2022

Utilizing web scraping and natural language processing techniques, my team and I created a model to predict the primary coding language of GitHub repositories. Our top model outperformed the baseline by 30%.

#### Job Listing Classification January 2022

Using data from Kaggle, I analyzed features of job postings in order to classify them as authentic or fraudulent. With a Random Forest model I created using available features as well as feature engineering, I was able to achieve 95% accuracy in my predictions.

#### Zillow Clustering January 2022

With a partner, I acquired, prepared, and explored data from the Zillow database within Codeup's SQL server with a focus on reducing error in home value predictions. As part of our exploratory data analysis, we created several clusters using various features contained within the data. We then used those features to build a linear regression model which performed better than the baseline.