

WENLIANG SUN

Data Scientist

✉ Wenliangsonny@gmail.com

☎ 804-412-5432

in [in/wenliang-sun/](https://www.linkedin.com/in/wenliang-sun/)

🐙 [sunwenliang14](https://github.com/sunwenliang14)

With a solid background in data science and business, I enjoy analyzing and discovering relationships within data. My appetite for providing insightful findings to colleagues and stakeholders that can solve real-world problems grows with each new data set that I encounter.

TECHNICAL SKILLS

Applied Statistics - SQL - Python - Pandas - Matplotlib - Seaborn - Scikit-learn - SciPy - Machine Learning - Natural Language Processing - Data Storytelling - Git - Spark - Jupyter Notebook - Anaconda - Tableau - SAS

EDUCATION

Codeup Jul 2020

Certificate of Completion, Data Science

Fully-immersive, project-based 20-week career accelerator that provides students with 670+ hours of expert instruction. I developed expertise across the full data science pipeline (planning, acquisition, preparation, exploration, modeling, delivery).

University of the Incarnate Word 2019 –2021
MBA, Data Analytics (GPA 3.9)

American Military University 2016 – 2018
Bachelor of Arts in Political Science
Cum Laude (GPA 3.6)

PROFESSIONAL EXPERIENCE

United States Army 2014 – 2018

During my active duty service in the Army, I worked extensively on patients' health and nutrition data, classified and analyzed patients' health data to deliver personalized nutritional treatments; analyzed inventory data and assisted superiors for the daily running of cafeterias in a military hospital, documented and generated daily reports.

SECURITY CLEARANCE

Active Secret Clearance Granted on 2016

LANGUAGE SKILL

Chinese-Mandarin

DLPT Test Score (Reading: 3, Listening:3)

DEVELOPMENT PROJECTS

Texas COVID-19 Analysis

Jul 2020

On a team of four, analyzed which counties have been most impacted by COVID-19 across the State of Texas, looked for common characteristics of vulnerability among these counties, highlighted what areas are vulnerable and why, understood social determinants of what makes an area a hotspot

Natural Language Processing

Jun 2020

Scraped GitHub repositories and their readme files using Beautiful Soup library and GitHub API. Contents inside the readme files were used to predict the primary programming language used for each repo. The best prediction was achieved with a decision tree model.

Zestimate Error Prediction

May 2020

Partnered with a team member, improved the log error of original Zestimate by using clustering methodologies. I dealt outliers, filled null values, scaled the data with MinMaxScaler in Sklearn, applied various models such as linear regression and random forest to compare whether our best model can outperform Zillow's baseline model.

Telco Churn Classification Analysis

Apr 2020

Applied classification methodologies to analyze drivers of churn for Telco customers. Performed feature engineering and created predictive models with various algorithms, selected the best model with domain knowledge to predict the probability of a customer would churn or not. Presented findings to senior leadership and made recommendations on how to reduce customer's churn

Zillow Property Value Prediction

Apr 2020

Utilized SQL and Pandas to acquire data from Zillow database, cleaned and prepared data to identify the leading drivers of properties values. Visualized relationships among variables with matplotlib and seaborn. Applied hypothesis testing such as T-Test and Pearson's correlation test to explore relationships between variables. Used various combinations of features to build up linear regression models, and selected the best model to predict properties values. Presented findings and recommendations to colleagues and senior leadership.