JASON LU

Data Scientist • 626 354 3033 • jlu7922@gmail.com • linkedin.com/in/jasonflu/ • github.com/jlu7922

PROFESSIONAL PROFILE

As a data scientist with a background in computational neuroscience, I am passionate about leveraging data to uncover insights and drive innovation. Through my expertise in data analysis, statistical modeling, and machine learning, I am dedicated to solving complex problems and empowering teams to make informed decisions. My experience in computational neuroscience provides a unique perspective and skill set that enables me to understand and model complex biological systems, enabling me to create innovative solutions that optimize outcomes. With a focus on collaboration, communication, and continuous learning, I strive to transform data into valuable insights for businesses and organizations.

SKILLS

Programming: Python, C++, R, NumPy, SciPy, Pandas, PySpark, Data Structures and Algorithms, Jupyter
Data Analytics: SQL, Hadoop, MapReduce, Scikit-learn, AWS, MongoDB, Tableau, SAS, SAP BI, Database systems
Statistics: Statistical Inference and Data Analysis, Structural Bioinformatics

WORK EXPERIENCE

Lead Data Scientist 05/2023 – 09/2023
U.S. Census - The Opportunity Project

- Spearheaded the integration and analysis of the CDC's COVID-19 data, uncovering key geographical trends that facilitated the creation of a website aimed at standardizing the reporting of at-home COVID-19 testing patterns.
- Designed and implemented a dynamic visualization tool, offering an enhanced version of the MakeMyTestCount approach, allowing users to view localized COVID case distributions based on their zip code, promoting broader public health insights.
- Engineered an intuitive tool that seamlessly translates federal government FIPS codes into user-friendly zip codes, amplifying the tool's utility for end-users by bridging the gap between detailed geographic datasets and localized, user-relevant insights.

Data Science Immersive Fellow 03/2023 – 06/2023
General Assembly

Successfully completed 500+ hours of expert-led instruction in Python, Data Visualization, and statistics, with hands-on learning of data manipulation fundamentals and the industry's most in demand technologies. Developed projects, including:

- **NLP Classification on Reddit Posts** - Utilized web scraping through the Reddit Pushshift API to collect a comprehensive dataset of over 10,000 subreddit posts. Applied Natural Language Processing techniques to process and analyze the text. Build a classification model for accurate prediction of a subreddit post's origin.
- **Predicting Housing Price** - Developed multiple linear regression models to accurately predict sale prices for homes in Ames, IA within a $30,000 margin of error.
- **Standardized Test Analysis** - Analyzed how location and intended college major of high school students majorly correlates with projected SAT/ACT performance, utilizing the Pandas and Matplotlib data analysis libraries.

Machine Learning Engineer Intern 11/2020 – 01/2021
Siemens PLM (Los Angeles, CA)

- Successfully designed and implemented an API for accessing and converting XML file data into CSV files to aid root cause data mining and to prepare datasets for training machine learning models.
- Improved an internal configuration tool to help users solve validation failures within Siemens NX software
- Optimized model to better predict error-prone features and provide insight for automating solutions

EDUCATION

Bachelor's Degree in Computational Neuroscience, Minor Degree in Data Science May 2024
University of Southern California, Los Angeles, CA