A.J. Goldsman

linkedin.com/in/aj-goldsman | ajgoldsman@gmail.com | 818-800-2809 | ajgoldsman.github.io/My-Portfolio/

Education & Certifications

Thinkful, Online

- Data Science Flex Program, Certificate of Completion — January 2020
  - **Relevant Coursework:** Programming in Python, Data Visualization & Experimentation, Supervised/Unsupervised Machine Learning, SQL Fundamentals, Statistics & Probability

University of California, Los Angeles, Los Angeles, CA

- B.S. Chemical & Biomolecular Engineering, Minor in Theatre — June 2019
  - **Relevant Coursework:** Computer-Aided Design and Analysis with Pro/II, Computer Programming with C++, Computer Programming with MATLAB, Introduction to Geographic Information Systems

Work Experience

Spatial Front Inc., *Data Scientist* — February 2020 – Present

U.S. Census Bureau, Economy-Wide Statistics Division (Contract)

- Performed general statistical analysis, data science, and Python programming work required by the various teams within the Census' Economy-Wide Statistics Division (EWD).
- Reviewed old Python code written by members of the team that have since left the Census. Translated the Python scripts to more easily understandable pseudocode and generated a report summarizing the program's functionality, thus enabling current and future team members to better understand and use sections of the code in related projects.
- Currently developing and testing a Python application that will accurately address over 70% of the million non-useful entries reported with each Economic Census cycle. By integrating this Natural Language Processing (NLP) based tool, the Census will be able to quantify up to 20% more revenue data than in previous cycles, without investing analysts' time and resources.

Dr. Playlist, *Co-Founder & CTO* — May 2019 – January 2020

- Developed a machine learning pipeline with the goal of improving music classification, with a focus on music genre/style, via analysis of instrument type, chord progressions, meter, etc.
- Built and compared more than a dozen supervised machine learning models to classify over 100,000 tracks from the Free Music Archive within 16 genre labels. Achieved a model testing accuracy of over 90%.
- Developed a Heroku-hosted, browser-based Flask app that recommends Disney songs to a user based on a comparison of various features extracted via SpotiPy API and NLP analysis of the songs' lyrics.

NovellusDX, *Data Science Intern* — Summer 2018

- Designed and implemented Python functions utilizing NumPy and Pandas to compare feature data collected by the image recognition software searching plates of cells for specific, introduced mutations.
- Utilized logistic regression, principal component analysis, and t-SNE analysis, as well as other statistical models, to analyze the relationship between the dozens of features from the cells being analyzed and the algorithm's probability of successfully predicting whether a specified mutation was present.
- Built a Python program to automate the data analysis process by determining the most significant experimental features and their appropriate ranges, improving the algorithm's accuracy while reducing the amount of employee time required to tune it.

Projects

Thinkful, *Data Science Fellow* — August 2019 – January 2020

- Utilized advanced programing skills and algorithms to efficiently sort and filter datasets for the purpose of removing outliers, filling missing data, and normalizing skewed data distributions.
- Processed structured and unstructured data within large datasets (scale of over 1,000,000 samples) to build predictive models, optimizing the collection of valuable intelligence to aid in the solution of business issues.
- Created meaningful data visualizations using Matplotlib and Seaborn that clearly communicated findings to non-technical audiences and related the findings to business impact, leading to proposed solutions that provided a foundation for future data-driven decision making.

Skills

Programming Languages: Python, MATLAB, C++, SQL
Libraries and Frameworks: NumPy, Pandas, Matplotlib, Seaborn, Scikit-learn, SpaCy, ScraPy, NLTK
Software Tools: SimSci Pro/II, QGIS, Microsoft Office Suite, Adobe Creative Suite