A.J. Goldsman

linkedin.com/in/aj-goldsman | ajgoldsman@gmail.com | 202-430-6786 | ajgoldsman.github.io/My-Portfolio/

Education & Certifications

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

Thinkful

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

Work Experience

Capital One - U.S. Card, *Sr. Associate - Data Scientist* — October 2022 - Present

- Developed a pipeline of Python helper functionality that uses the SQLGlot library architecture as a base for efficiently finding relationships and extracting core calculation logic from within inter-related segments of complex SQL queries, which encompassed many thousands of lines of code, often with intricately nested sub queries.
- Designed and implemented Python functionality for normalizing comparison populations by a given base population filter, by which more accurate analysis can be done across populations without improperly weighted biases.
- Improved upon existing Python infrastructure that allows analysts to quickly and accurately query large databases and aggregate commonly requested metrics. Work involved both core Python logic and formatting of SQL queries via Jinja templates to reduce memory allocation and time complexity problems.

Deloitte Risk & Financial Advisory, *Consultant - Data Scientist* — June 2021 – September 2022

Federal Law Enforcement Agency (Contract)

- Created and improved upon Python applications, Flask APIs, NiFi flows, and Kibana dashboards, focused on using Natural Language Processing to assist the client in identifying criminal activities, especially those related to money laundering.
- Developed a Python module with a Command Line Interface to allow forensic accountants to process large CSV files of financial transaction data. Compared to their old Excel template, this new tool allowed them to process orders of magnitude larger datasets and used regex to greatly improve parsing of text into entity name, address, and country of origin.
- Improved upon architecture that used HuggingFace transformer models to create a dashboard user interface into which the client could feed text data and relevant questions, which increased efficiency exponentially as the amount of text being processed grew linearly.

Spatial Front Inc., *Data Scientist* — February 2020 – June 2021

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

- Developed a Python application to effectively address over 70% of the million non-useful entries reported with each Economic Census cycle. With the planned integration of this tool into the 2022 Economic Census, we will be able to quantify up to 20% more revenue data than in previous cycles, accounting for billions of dollars in additional revenue.
- Developed a Tableau-esque Python application that will allow Census analysts to pull relevant data and produce general visualizations for pinpointing areas of focus. The tool runs as a Flask app and produces visualizations via Plotly Dash.
- Translated outdated Python scripts to more easily understandable pseudocode and generated reports summarizing the programs' functionality, enabling team members to better understand and use sections of the code in related projects.
- Worked on several projects in which Supervised Machine Learning was used to develop classification models and interpolation pipelines for matching publicly available data and imputing missing records in Census and IRS databases.

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 cancerous mutations.
- Utilized logistic regression, principal component analysis, and t-SNE analysis, as well as other statistical strategies, 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 program to automate the data analysis process by determining the most significant experimental features and their appropriate ranges, improving model accuracy while reducing the amount of employee time required to tune it.

Skills

Programming Languages: Python, SQL, Javascript, HTML/CSS, MATLAB, C++
Libraries: NumPy, Pandas, Matplotlib, Seaborn, Sklearn, SpaCy, NLTK, Bokeh, Flask, Transformers, Haystack, SQLGlot
Software Tools: QGIS, Linux, Jenkins, Vim, Elasticsearch, Kibana, NiFi, Git, Snowflake