Weston Odend'hal

+1-415-539-6833 | wodend@tuta.io | Seattle, WA
github.com/wodend | linkedin.com/in/wodend

TECHNICAL SKILLS

- **Proficient**: Spark, Python, Presto, NumPy, SciPy, Pandas, Matplotlib, BeautifulSoup, PHP/Hack, React, R, C/C++, Bash, Linux, Git, CI/CD, Docker, Java, Maven, Jakarta EE, Selenium, Pandoc, LaTeX, HTML/CSS, LDAP
- **Exposure**: Django, bqplot, C#, .NET, CSS, Gradle, LXD, Scala, Lua, Javascript/Typescript, SAS, Airflow

WORK EXPERIENCE

Facebook | Data Engineer | Seattle, WA   Feb. 2022 - Present

- Automated processing for hundreds of tables and samples with Jupyter Notebook + Python.
- Enriched billion-scale datasets with custom transforms using Hack and Python.
- Designed pipelines with DAG execution frameworks to ETL big data using SQL (Presto, Spark) on Hive.
- Enforced privacy & security on sensitive data through tagging and automated filtering.

Facebook | Data Engineer (Contract) | San Francisco, CA   Oct. 2020 - Dec. 2021

- Designed and implemented ETL pipelines to produce billion-scale training data pool tables for research scientists to sample from for annotation projects.
- Wrote libraries to facilitate the stratified sampling and annotation of data for responsible AI at the hundreds-of-millions scale.
- Designed React GUIs to enable the annotation of datasets for multimodal supervised learning.
- Implemented libraries for calculating common computer vision metrics such as AUC and MAP at scale.
- Created a library to convert datasets from Hive to JSON at scale to support mobile training.

Deloitte | Consultant (Cybersecurity) | San Francisco, CA   Jul. 2019 - Jul. 2020

- Created data pipeline to aggregate and clean user login data across a distributed environment with PowerShell scripts and LDAP database queries, reducing time spent 50% and including previously unavailable data.
- Automated analysis of user login data with a Java program using the Tablesaw library to generate daily and weekly login statistics, reducing time spent by 90%.
- Introduced a new process to generate security scan crawls using Selenium, saving the client $12,000 per month.
- Generated sanitized test data using IBM Optim to obfuscate Personally Identifying and Protected Health information from the production SQL Server database with over 100,000,000 records for 450,000+ users.

Cal Poly | Teaching Assistant (Research) | San Luis Obispo, CA   Jun. 2017 - Jun. 2018

- Cleaned and curated SAS datasets downloaded from the Panel Study of Income Dynamics, allowing researchers to access 100+ of GB's of previously unavailable data.
- Automated determination of optimal office hours for professors with a tool written in R using matrix algebra, saving 90% of time spent scheduling each quarter.
- Created a custom PDF scraping tool in C using the Poppler library to facilitate research, allowing us to include critical factors that were necessary for our model.

EDUCATION

B.S. Economics; Minor Computer Science, *California Polytechnic State University*   Jun. 2019

PROJECT WORK

PySimplePlotter | Software Engineer | github.com/wodend/pysimpleplotter   Apr. 2021 - Present
Plot output from scientific instruments quickly and easily with a simple GUI.

- Implemented file reader to automatically detect, clean, and read various delimited file formats.
- Wrote a tool using pandas and matplotlib which allows the user to visualize plots of multiple relationships and export the output.
- Designed a GUI interface to matplotlib's API to allow the user to customize the plot.