Weston Odend'hal

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

TECHNICAL SKILLS

- **Languages**: Hive SQL, Python, Presto, C/C++, PHP/Hack, Rust, Javascript, Java, PowerShell, Scala
- **Libraries**: React, OpenCV, NumPy, SciPy, Pandas, Matplotlib, requests, BeautifulSoup, Jakarta EE, PyTorch
- **Tools/Platforms**: Linux, Git, JupyterLab, DigitalOcean, Docker, Buck, Maven, Pandoc, HTML/CSS, LaTeX

WORK EXPERIENCE

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

- Optimized ETL pipelines processing billions of records using hourly denormalized scraping techniques, reducing critical table landing time by several hours.
- Automated processing for hundreds of tables and samples with Jupyter Notebook + Python for grain prediction and tagging for privacy & security.
- Enriched billion-scale datasets with custom transforms using Hack and Python, making new features available and reducing compute time for large Hive joins by 80%.
- Designed data models and ETL pipelines for core datasets supporting analytics, finance, and regulatory teams for payments data at scale, enabling millions of dollars of incremental revenue through new product launches.

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.
- 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.
- 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.

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.

EDUCATION

B.S. Economics; Minor Computer Science, *Cal Poly San Luis Obispo*   Jun. 2019