

GRAHAM WATERS

Austin, Texas • 512.568.2351 • gtxdatascientist@gmail.com

OBJECTIVE

Seeking a position at a company with the resources to provide opportunities for growth, social good, using data and AI for ethical transformation in data science, analytics, engineering, or other related roles.

Experience

Tools in Tech Stack Related to Position on the right.

*PowerBI, Excel, VBA,
SharePoint, Pandas,
Python*

*VSCoDe, ChatGPT,
OpenAI, GitHub Copilot,
Plotly, Seaborn*

*GitHub
Version Control*

*Sklearn, Machine
Learning, OpenCV*

Data Analyst Intern, engage2Learn, Inc.

**March 2023 –
Present**

- Worked with coaches and education leaders to determine use-cases for large language models in our codebase without compromising security.
- Analyzed and interpreted data from multiple sources to inform decision-making for school districts and their coaches.
- Conducted statistical analyses using Excel, Sheets, R, and Python to identify trends and correlations in coaching and student performance data.
- Developed and implemented data visualization dashboards using Domo and AWS QuickSight to communicate insights to stakeholders.
- Collaborated with cross-functional teams to identify areas for process improvement and to drive service improvements.
- Utilized strong problem-solving and critical thinking skills to develop creative solutions to complex data challenges.
- Demonstrated excellent communication skills in presenting data to stakeholders, active listening, and conflict resolution.
- Designed and implemented data warehousing and reporting structures to increase efficiency and scalability of data systems.
- Prioritized data privacy and security in all data-related processes and decisions.

While at engage2Learn served on the ai-Ops Tiger Team:

On a multidisciplinary team of ten specialists to integrate modern ML & AIOps Tools into our product offerings: Algorithmic prediction, GPT-4, chatGPT, and ML for ed-tech data, coaching effectiveness, and promoting educator well-being.

Skills: Large Language Model Prompting · Artificial Intelligence (AI) · Python (Programming Language) · Writing · Data Science

Lead Data Analyst for SCEM Division, Entegris, Inc. Austin, Texas

Nov 2020 – Jan 2021

- Partnered directly with the executive team as the first data expert hired to attempt a redesign of the existing data warehousing and reporting apparatus and **completed a 20% transformation.**
- Revised processes, data storage formats, and structures to increase reporting systems' efficiency to improve future ML applications' scalability.
- Created statistical information for monthly reports detailing processes for the SCEM (*specialty chemicals and engineered materials*) division, which **directly impacted decision-making for over 100 supervisors, product planners, and factory managers.**
- Utilized PowerBI, Excel, Python, and VBA to automate the two most challenging, time-intensive supply chain and logistics reports from their legacy Excel formats into PowerBI Dashboards and performed cursory analyses in Pandas to identify weak points and inefficiencies.
- Delivered status updates to supervisors and recommended changes in design based on challenges faced by various stakeholders.

Regex, NLTK, SpaCy

PandasProfiling

- Drove service improvements by working cross-functionally to close process gaps.
- Reviewed large amounts of data across multiple platforms and sources to determine operational impacts and needed actions.
- Compiled weekly data in SharePoint and Excel for production reports communicating metrics to various cross-functional decision-makers.

HTML, markdown

Technical Author on Medium

May 2020 – Present

Austin, Texas

Current articles received
1,848 views in the last
30 days.

- Actively contributing as a Top Writer for MLearning.AI with over ten featured papers and/or articles for the publication.
- Generating content consistently while maintaining over 200 personal followers and 130+ published articles.
- Key points relating to this role:
 - My top article has 6.1k views.
 - Eight articles have over 1,000 views.
- Participated in local meetups.
- Developed three Open-Source tools that are useful for data scientists and contributed to the Open-Source Ecosystem by producing over thirty personal repositories and collaborating with The *FlatIronSchool* and National Renewable Energy Laboratories.
- Used Scikit-Learn models to predict home prices with RMSE scores near 0.12 and R2 scores on Test and train sets above 90%.
- Actively cleaned and wrangled unstructured data into analysis-ready formats for Pandas and ML Techniques.

EDUCATION

- **Master of Science in Strategic Analytics (4.0 GPA)** August 2021
Brandeis University, Waltham MA
- **Bachelor of Science in University Studies in Mathematics, Agricultural Leadership, and Plant and Soil Science** May 2018
Texas Tech University, Lubbock, TX
- Completed **300 hours of post-graduate level hands-on data science/analytics coursework.** Sept. 2022

SKILLS AND TECH STACK

Technical Skills

- Proficient in **Python programming**
- Experienced in data analysis using **Pandas**, **NumPy**, Tableau, PowerBI, and Excel
- Skilled in data visualization and storytelling with **Matplotlib**, **Seaborn**, PowerBI, and Excel
- Proficient in Microsoft Office and Google Suites
- Knowledge of machine learning techniques such as clustering, regression, and boosting
- Familiarity with the process of academic research and conformance to APA standards
- Web scraping with Selenium, Requests, and BeautifulSoup in compliance to site limits.
- Natural language processing using both nltk and SpaCy (lemmatization, stemming, etc.)
- Critical Thinking and Evaluation of NLP responses and prompting.

Soft Skills

- Excellent cross-functional **communication** and **collaboration** skills
- Strong problem-solving, solution-minded, and cross functional thinker
- Highly **adaptable** and eager to learn.
- **Diligent trouble-shooting ethic**
- Active Listener with attention to detail
- Prompt Engineering for LLMs
- Collaborative sprint project planning
- SCRUM

Evolving Tech Stack

- **Python** and libraries **Pandas** and **NumPy**.
- **Scikit-Learn** machine learning library.
- **Matplotlib**, Bokeh and **Seaborn**
- Knowledge of database management and querying using **Pandas**, CSV, JSON, and MySQL.
- Selenium, Requests, and BeautifulSoup
- **GitHub Copilot**
- AWS Cloud
- **OpenAI ChatGPT, Codex**
- Regex, Figma, Canva, Miro
- NLP Tools: **SpaCy**, NLTK
- Pillow
- AWS QuickSight Dashboard generation
- AWS Redshift DW
- Domo
- GPT-4