

# William Medwid

Redwood City, CA 94062 | [william.medwid@gmail.com](mailto:william.medwid@gmail.com) | [LinkedIn](#) | [williammedwid.com](http://williammedwid.com)

## Product Data Scientist

Data scientist with 4+ years of experience building scalable data systems, predictive models, and analytical tooling across advertising, logistics, content moderation, and healthcare. Expert in scalable distributed environments, from data ingestion to ETLs and dashboards. Patient and engaged communicator, putting team consensus building first and technical implementation as the follow-through. Quick to demonstrate self-led prototypes and to amend or pivot to new priorities.

M.S. in Statistics with a deep foundation in experimental design, statistical modeling, and machine learning, including LLM architectures. Strong advocate for practical AI automation supported by careful review processes and concise technical communication.

## Areas of Expertise

Statistical Communication | Cross-functional Collaboration | Data requirements | Strategic Prioritization | Presentation KPIs | Experimental Design | A/B Testing | AI Workflow Management | Data Governance | Data Warehousing | Semantic Models | Technical Troubleshooting | Dashboard Management | Data Visualization | Data Reliability Assurance | ETL | Compute/Storage Minimization | Data Modeling | Data Analysis | Analytics Engineering | Business Analytics

## Technical Skills

**Programming Languages and Packages:** Python, SQL, R, Javascript, Pandas, Scikit-Learn, TensorFlow, Tidyverse  
**Cloud Infrastructure & Distributed Systems:** Dataswarm, Amazon Redshift, Apache Hive, Apache Spark, JDBC  
**Dashboarding Tooling:** Amazon Quicksight, Tableau, Unidash (Meta proprietary), Nest, Metabase

## Languages

Spanish - Fluent | Mandarin - Basic

## Work Experience

**Meta** **July 2025 - May 2026**

Data Engineer, Integrity - Menlo Park, CA

- Developed and maintained **Dataswarm SQL ETL pipelines** and data monitoring for finance trend analysis and **AI efficiency** tracking
  - Created an estimation system, tuned hyperparameters, reducing errors in projected invoices by 75%.
- Implemented automated data quality monitoring using internal agentic AI tools, reducing resolution time from days to hours and ensuring data reliability for data scientists and leadership.
- Developed user-friendly executive **dashboards** with Unidash and **Nest**, providing technical guidance on budget status and identifying specific workflows for automation.
- Automated routine workflows, eliminating over 30% of manual data engineering touchpoints such as initial code drafts and stale data quality checks.

**Amazon**

June 2023 - July 2025

Business Intelligence Engineer, Logistics - Bellevue, WA

June 2024 - July 2025

- Established and maintained **SQL ETL** pipelines using **Amazon Redshift** central warehouse infrastructure, ensuring timely daily updates to data and quick resolutions to data quality issues.
- Streamlined supply chain network flow updates from an hour-plus process to a 5-minute task using **Python (Pandas, Xpress optimizer)** to accelerate responses to changes nationwide demand distribution.
  - Provided detailed impact analysis to logistics leadership for pre-launch approval, maintaining cross-functional alignment and trust
- Built **Tableau** and **Quicksight** dashboards to expose prior data blind spots and improve operational visibility for cross-functional stakeholders.

Jr. Business Intelligence Engineer, Ads - San Luis Obispo, CA

June 2023 - June 2024

- Built **Tableau** dashboards exploring user experience and engagement within the advertising partnership system.
- Developed **ETL** pipelines and data modeling for aggregated data analysis to support product feature tracking.
- Performed detailed WBR data reviews and identified internal metric inaccuracies of up to 20%, improving overall data quality and reliability.

**Applied Orthopedic Design**

May 2022 – June 2023

Lead Data Engineer - Arroyo Grande, CA

- Maintained and grew the central full-stack data pipeline for knee and hip implant outcomes using **JDBC, SQL, Google Sheets, and JavaScript** to centralize clinical measurements and outcomes data.
- Developed and evaluated **ML pipelines** for pre-op planning using Python, scikit-learn, and TensorFlow, while identifying data quality and volume constraints impacting model performance.
- Delivered Metabase dashboards highlighting systematic and individual improvement opportunities to Biomedical Engineering teams, focusing on intuitive **data visualization** and **user-friendly** reporting.
- Launched a new data collection initiative with follow-up surveys to track patient recovery and experience, to validate marketing claims and improve future results.
- Automated administrative record keeping and patient report generation with Google Suite integrations and JavaScript, reducing manual effort and improving data consistency.
  - Reduced patient report generation time by 90% to a single click and double-check process.

**Undergraduate Data Science Researcher**

Sep 2020 – June 2022

Cal Poly - San Luis Obispo, CA

- With a Frost research grant, using R to investigate how infants communicate with their mothers while feeding.
- Examined a rich time series of events during feeding, as well as longitudinal data on individuals' outcomes.
- Created onboarding documents, clarifying the mechanisms behind the R package Flair.

**Education****M.S. Statistics**, Cal Poly San Luis Obispo

June 2024

**B.S. Statistics, Minor in Data Science**, Cal Poly San Luis Obispo

June 2023

Key coursework: Distributed Computing (**Spark and Hadoop**), Advanced Topics in Data Science, Data Science Process Ethics, **Mathematical Foundations of Data Science**, Estimation Sampling Theory, Probability Theory, **Knowledge Discovery from Data**, Applied Multivariate Statistics, Game Theory, Design and Analysis of Experiments, **Design and Analysis of Algorithms**, Generalized Linear Models.

Thesis (450+ downloads): [Recursive Matrix Game Analysis: Optimal, Simplified, and Human Strategies in Brave Rats](#)