Connor Day

contact@connor.day | linkedin.com/in/con-day | github.com/dayconnor | connor.day

Education

University of California, San Diego — San Diego, CA

Bachelor of Science in Data Science — *Expected June 2028*

- **Relevant Coursework:** Exploratory Data Analysis and Inference, Systems for Scalable Analytics, Recommender Systems and Web Mining, Data Visualization, Data Management (SQL), Probability, Practice and Application of Data Science, Theoretical Foundations of Data Science I & II

Projects

Polymarket Calibration | *Python, SQL, DuckDB, PyArrow, Parquet* — Oct. 2026 – Present

- Querying 1.27 billion on-chain trades (45.8 GB) in place with DuckDB, reading only the needed columns and row groups, and profiling data issues that changed the design, such as end dates that are not resolution times
- Built a rate-limited, resumable API client and a staging layer that cross-checks two outcome sources, keeping 2.21 million of 2.22 million resolved markets after excluding voided, disputed, and conflicting ones
- Pre-registered three hypotheses before comparing any price with an outcome, including a crypto vs. non-crypto test of calibration slopes using logistic regression, series-clustered errors, and Holm correction

Garment Search | *Python, PyTorch, Hugging Face Transformers, FAISS, Pandas, NumPy* — Sep. 2026 – Oct. 2026

- Built a text-to-image search engine over 45,622 Fashionpedia clothing photos, comparing fashion-specific FashionCLIP against general-purpose CLIP using exact FAISS cosine-similarity indexes
- Pre-registered the evaluation before embedding any images: 20 queries across four difficulty tiers, written relevance criteria, pass thresholds, and six predicted failure modes
- FashionCLIP outperformed CLIP on capped recall@10 (0.612 vs. 0.488) and hand-judged precision@10 (0.755 vs. 0.540), but both failed the compositional tier and neither returned a clearly oversized blazer (0 of 10)
- Hand-judged 391 query-image pairs blind to model and rank, finding that 28% of results the dataset's labels scored as irrelevant were clearly relevant, so label-based recall understated retrieval quality

Thyroid and Metabolic Health | *Python, Pandas, Polars, statsmodels, scikit-learn, Matplotlib* — Aug. 2026

- Pre-registered four falsifiable hypotheses linking thyroid function to metabolic health, then tested them on three pooled NHANES cycles (2007–2012) covering 3,459 U.S. adults
- Fit survey-weighted regression models using NHANES strata, PSUs, and sample weights, adjusting for age, sex, and BMI, and built a metabolic syndrome flag from the 2009 Harmonized clinical criteria
- Applied FDR and Bonferroni correction: three of four hypotheses were unsupported, and the one significant result ran opposite to the prediction (+14.9 mg/dL triglycerides per unit FT3/FT4 ratio, $p < 0.001$)
- Built a reproducible pipeline from raw NHANES survey files to per-hypothesis analytic samples

Experience

Ralph Lauren Factory Store — Livermore, CA

Front & Back of House Associate — *June 2022 – Sep. 2025*

- Selected by management to collaborate with top-performing team members on styling store sections for corporate lookbook photography, adopted company-wide as the seasonal floor standard for all Ralph Lauren factory stores
- Trained 10+ new team members on store procedures, POS operations, and customer service standards, explaining processes clearly to people with no prior retail experience

Technical Skills

Languages: Python, SQL, R, Java, C++
Libraries: Pandas, NumPy, Polars, SciPy, statsmodels, scikit-learn, PyTorch, Hugging Face, FAISS, PyArrow, Matplotlib, Seaborn, Plotly, BeautifulSoup, Gradio
Data & Cloud: PostgreSQL, MySQL, SQLite, DuckDB, Parquet, AWS, Apache Spark
Data Skills: Data profiling, hypothesis testing, survey-weighted regression, multiple-comparison correction, retrieval evaluation
Tools: Git, GitHub, GitHub Actions, Docker, Linux, Conda, Jupyter, VS Code, Excel, REST APIs, LaTeX, Claude Code, Codex