

Trevor White

Trevor White resume (shippable)

Default 2-page version for cold-email attachments and portal applications without JD-specific tailoring. Trimmed from resume-canonical.md to lead with the strongest cross-track signal for C (AE/BIE) primary, E (AI Eng / FDE) secondary, A (Applied DS) tertiary. When a JD is in hand, tailor from canonical instead.

Last refreshed: 2026-05-25

CONTACT

- **Trevor White**, Seattle, WA
- trevwhite1329@gmail.com · 206-697-1263
- linkedin.com/in/trevor-w-white · github.com/tdub29

SUMMARY

Sr. Data Scientist (BI Lead) running BI for **Forbes' #2-ranked US charity**, the nonprofit that moved **\$3.9B in donated product in FY25**. Built and governed a hub-and-spoke Power BI platform on Azure Synapse serving 69 actively shipped report pages viewed daily across 20 departments, with 1,000+ average monthly views. Took the function over from a managed-service vendor while still in college; owned the data behind the 2024 IRS Form 990 filing and the in-flight 2025 financial audit.

EXPERIENCE

GOOD360, Sr. Data Scientist (BI Lead)

Jan 2026 to Present · Remote · Reports to Director of BI

- **Owned the data behind the 2024 IRS Form 990** (the org's most public regulatory filing) and currently own the same role for the 2025 financial audit, including the CBIZ FMV audit, which validated **\$16.29M GEM FMV** end-to-end alongside FMV by donor and saved the finance team weeks. I own the data behind **Schedule B (contributors)** and **Schedule M (non-cash contributions)**, the schedules carrying fair-market-value reporting on product donations, plus the **Amazon FMV** data work.
- Cleaned and reclassified roughly two years of the departmental income statement (**~\$130M** in revenue and expense) via SQL + Excel forensic workflows; delivered to Finance with documented SOPs to keep it from drifting back, enabling accurate budgeting and internal departmental analysis.
- Uncovered a **~\$1M cross-system inventory-valuation gap** and hundreds of thousands of dollars in FMV-vs-cost miscoding across multiple years of NetSuite + Salesforce records via Python + SQL forensics; drove reconciliation back into alignment and built a repeatable detection process so the gaps don't reaccumulate.
- Authored a **four-scenario warehouse-network expansion model**, a linear decision model driven by NPO-level demand prediction and great-circle distance routing, with capacity-cutoff outbound NPO logic, three revenue scenarios, and 107-row documented methodology. Clarified a **multi-million-dollar expansion opportunity** and a **30%+ lift in \$/mile shipping efficiency** for the recommended scenario; paired it with a MILP facility-siting model over a scraped national demand surface to independently validate the recommended sites.
- Built and own two production apps on top of the platform. The first is a **Power BI Documentation Agent (DocPBI)**: FastAPI + Celery + multi-agent Claude orchestration + Playwright visual capture + custom pbixray extraction, run against the governed production model and 69 actively shipped report pages, compressing weeks of manual documentation into a minutes-long run. The second is a **Web Donor Impact Portal** (FastAPI + Azure AD service principal + Power BI embed SDK + per-donor RLS + JWT) that gives external donors authenticated access to filtered impact reports without a Power BI license each.
- **Enabled 34 new metrics** previously clouded by ambiguity by combining warehouse cleanup, definitions work, and the Metric Dictionary co-built with the Director of BI; onboarded the incoming Director onto warehouse history, governance gates, and vendor relationships through the same window.

GOOD360, Data Scientist & Power BI Specialist

May 2023 to Dec 2025 · Remote · Part-time until Jun 2024, leading the BI function throughout; full-time from Jun 2024 after graduating USD. Data Intern Jun 2022 to Apr 2023.

- Built and governed a hub-and-spoke Power BI platform serving **69 actively shipped report pages viewed daily across 20 departments, with 1,000+ average monthly views**. Absorbed the prior managed-service vendor engagement through 2023-

2024 and formally offloaded it in Feb 2025, replacing **~\$72k/yr in high-rate vendor spend** with internal capacity.

- Integrated a previously missing donation channel in Jan 2025 (the missing **60% of total org volume**) into the warehouse, adding **\$1.42B of clean recipient-level attribution YoY** and lifting 2024 Schedule I coverage to **94%**.
- Own a **four-layer dimensional warehouse on Azure Synapse Dedicated SQL Pool** (staging → ODS → core → data-model → curated views) feeding Power BI through **346 curated production views (32k lines)**. Snowflake-schema dimensional modeling, idempotent metadata-driven delta loads (ods_pipeline_run_id tracking), multi-source UNION ALL fact tables at one grain, deterministic HASHBYTES synthetic keys, fuzzy entity resolution via multi-prefix ROW_NUMBER ranking. Integrates **nine source systems** through 65+ SQL Server tables and 30+ REST endpoints: NetSuite (ERP), Salesforce (CRM), Magento (e-commerce), Amazon partner manifest / BOL / LPN, Jotform, **FEMA National Risk Index** (465-column county-level), EM-DAT, Census + Social Vulnerability Index, and CPI. Roughly **107 of those views feed the Power BI semantic layer**, and **86% of the curated view code was last authored by me**.
- Operate the platform underneath it: **431 orchestration pipelines running 2,652 nested activities**, a **429-procedure metadata-driven load framework totalling 76,436 lines**, and **2,521 tables across 52,459 columns**. The load framework is vendor-built and generated per source table; I run, maintain and extend it. The curated view layer on top is the part I authored.
- Governed the production surface through the BI Asset Catalog Hierarchy I authored (Active / Beta / Planned / Archived state machine, 16 governance fields per page, 5-level Super-User Training rubric), with field parameters consolidating user-driven selectors instead of letting the model drift into measure-per-question duplication.
- Authored GOOD360's **three-pillar BI strategy** (Data Foundation & Governance / Impact Measurement & Metric Integrity / Analytics & Decision Support) and **led the Data Governance Steering Committee**: established the gate that "Salesforce reporting must be approved by BI before using," authored the escalation framework + decision-rights structure, replaced a predecessor info-sharing committee with one that ships decisions.
- Shipped **production Python automation** across the org (Jotform-to-warehouse ingestion, ~400-LOC openpyxl per-state donation generator, donation / donor classification pipelines, foreign-currency joins, and FMV reconciliation across NetSuite + Magento + Salesforce); operated production **gradient-boosted (XGBoost)** models for compliance risk + disposition decision-modeling; built a **natural-language Asana agent** (intent parser + name-to-ID resolver + executor with confirmation gates) that collapses hours of manual task setup and triage into single plain-English commands.
- Quantified 2024 business impact: **15% reduction** in estimated CO₂ per truckload delivery; **4% decrease** in time to place a donation; **10% increase** in FMV donated to High Need areas. Backed **~\$17.8M in grant funding secured since 2023**, including a **\$500k corporate-sponsored grant** that funded BI team expansion.

University of San Diego Baseball, Volunteer Analyst (de facto lead)

Jan 2023 to Present · San Diego, CA · Weekly opponent-driven scouting cadence, refreshes tagged to upcoming series so the coaching staff binder is current Tuesday-before-series.

- **10 production models** across NCAA TrackMan and MLB Statcast (Stuff+, expected whiff, expected slugging, run value with plate location), plus a **top-25 finish** in the NWDS batted-ball prediction competition (tdub29/WanBattedBall).
- Modeled **pitcher arm angle** from release-point coordinates, body measurements, and pitch kinematics: engineered interaction features and tuned a gradient-boosting regressor against a linear baseline, reaching **84% explained variance (6.4° RMSE)** with residual-based prediction intervals; built the training set by joining MLB StatsAPI body data to pitch-tracking feeds.
- Built a **Stuff+ pitch-quality model** on **~1.6M NCAA Trackman pitches**, trained on MLB-derived run-value labels (count × event → delta-run-exp) over velocity, spin, extension, release point, and fastball-vs-secondary differentials, normalized to the standard 100-mean / 10-stdev scale. Scores every pitch against an NCAA-wide reference set by pitch type.
- Shipped a **Streamlit pitcher app** (tdub29/streamlit-app-1, "TORERO PITCHING") the USD coaching staff uses for game prep: TrackMan ingestion, plate-location plots split by batter handedness, break plots, release-point clustering with command-ellipse overlays, Stuff+ rolling trends, polar tilt plots with model-driven target locations.
- Generated automated **PDF post-game scouting reports** with run-value-coded heatmaps and count-stratified tables benchmarked against NCAA percentile distributions; integrated a **multi-source data layer** stitching Trumedia game logs, TrackMan In-Stadium, MLB Statcast (bat-tracking), and a custom master CSV with pitcher-name resolution and unified feature engineering.

EDUCATION

- **University of San Diego, Knauss School of Business: B.B.A., Business Analytics** · May 2024 · **Magna Cum Laude** · GPA 3.7 / 4.0
- **MITx MicroMasters Program in Statistics and Data Science**, estimated completion Oct 2026

SKILLS

- **Languages:** Python, T-SQL (Azure Synapse Dedicated SQL Pool), R, DAX, Power Query M

- **BI / warehouse:** Power BI hub-and-spoke serving 69 actively shipped report pages across 20 departments, 1,000+ average monthly views, BI Asset Catalog governance, field-parameter consolidation, RLS, embed SDK; Azure Synapse Dedicated SQL Pool, snowflake-schema dimensional modeling, idempotent metadata-driven delta loads, HASHBYTES synthetic keys, fuzzy entity resolution; Tableau (familiar); ArcGIS in Power BI
 - **ML / modeling:** XGBoost, scikit-learn, pandas, numpy, statsmodels, LightGBM, Streamlit, gradient-boosting / RandomForest / Ridge; Stuff+ on ~1.6M NCAA Trackman pitches; expected-outcome modeling on launch-speed / launch-angle inputs
 - **AI / agents:** Claude API, multi-agent orchestration (intent parsing → resolver → executor with confirmation gates), prompt + context engineering, AI-assisted dev workflows (Cursor); Playwright (browser automation / visual capture)
 - **Web / app stack:** FastAPI, Next.js, React, TypeScript, Tailwind, Celery + Redis, Docker, PostgreSQL + Prisma + SQLAlchemy, Cloudflare Workers (OpenNext), Vercel, JWT, OAuth2 / Azure AD, Expo + React Native
 - **Source systems:** Salesforce, NetSuite, Magento, JotForm
-

SELECTED PROJECTS

PRTL: natural-language image generation for in-home displays · prt1home.com

Multi-platform monorepo (web app, mobile app, serverless API, marketing site), 208 commits Dec 2025 to Feb 2026. Integrated **Google Gemini 3 Pro Image Preview** behind a Vercel serverless function: natural-language prompt → image generation pipeline with 17 mode-configuration components composing user inputs + mode context into the prompt before generation. React Router 7 + Bun + TypeScript + Tailwind + Tanstack Query + Stripe + Neon Postgres + Hono + Auth.js on web; Expo 54 + React Native 0.76 + expo-router 6 on mobile. Marketing site is live; product is in development.