

Will Charles

Boise, ID | 775-771-6460 | wjcc91@gmail.com

SUMMARY

Data engineer with an M.S. in Computer Science who has owned a data stack end to end: schema design, API ingestion, dbt modeling and testing, CI, and the dashboards at the end. Five years building the analytics backend of an eCommerce company I co-founded, now writing T-SQL against operational court databases and building the reusable datasets and data-quality checks the Idaho judiciary relies on.

WORK EXPERIENCE

Data Analyst

2025 – Present

Idaho Supreme Court

Boise, ID

- Build reusable, maintained datasets in T-SQL over large operational court databases, replacing one-off requests with data products that court leadership and the Idaho Legislature rely on
- Validate and reconcile reporting datasets against source systems, catching upstream data-quality defects before they reach published reports
- Own data dictionaries, metric definitions, and reporting standards (data governance) so metrics stay consistent across teams and over time
- Design and publish Power BI semantic models – Power Query transformations, table relationships, DAX measures – managing security and distribution through the Power BI Service

Co-Founder, Data & Analytics

2020 – 2025

TidyBoard

San Francisco, CA

- Designed and ran a PostgreSQL warehouse consolidating sales, advertising, inventory, and customer data from multiple vendor APIs into a single queryable source
- Automated daily ingestion and transformation from those APIs, replacing manual exports and keeping the warehouse consistent enough to run the business on
- Forecasted regional inventory demand, reducing both stockouts and excess inventory across distribution centers
- Co-founded and scaled the brand from concept to eight figures in revenue with a lean team, owning the technical and analytics side end to end

PROJECTS

Job Market Analytics Platform | *Python, SQL, dbt, Airflow, PostgreSQL, GitHub Actions*

2026

- Built a daily batch pipeline ingesting thousands of company job boards into PostgreSQL, using pydantic contracts to reject malformed records and natural-key upserts so reruns never double count
- Derived posting lifecycle history that no source API exposes by snapshotting board state daily with dbt SCD2, producing a time-to-fill proxy by company, location, and title
- Modeled a two-source star schema with a skill bridge and presentation marts, covered by 136 dbt tests and source freshness checks, with CI running the full build on fixtures instead of live APIs
- Orchestrated the pipeline as an Apache Airflow DAG (ingest → dbt build → retention) with task retries and a sensor gating the transform on data availability; validated backfill and catchup replay against the DAG's logical dates
- Delivered an Evidence.dev dashboard and a version-controlled Power BI semantic model from the same warehouse, rebuilt unattended each morning by scheduled GitHub Actions

Long-Form Audio RAG System | *Python, FastAPI, Qdrant, Whisper, Vue.js*

2025

- Built a retrieval-augmented search system over long-form audio, transcribing podcast episodes with Whisper and serving semantic queries through a FastAPI backend and Vue.js frontend
- Chunked and embedded transcripts into a Qdrant vector database, with an LLM answer layer that retrieves the relevant passages at query time

EDUCATION

University of Illinois Urbana-Champaign

May 2025

M.S., Computer Science in Data Science – 4.0 GPA

Champaign, IL

- Built a distributed key-value store in C++ using the SWIM protocol for membership and failure detection

University of Nevada, Reno

April 2022

B.S., Discrete Mathematics, Minor in Computer Science

Reno, NV

TECHNICAL SKILLS

Languages: Python, SQL (T-SQL, PostgreSQL), C/C++, Bash, JavaScript, R

Data Engineering: dbt, Apache Airflow, PostgreSQL, SQL Server, dimensional modeling (star schema, SCD2), data quality testing, pydantic contracts, Docker, Google Cloud, Git, GitHub Actions (CI and scheduled builds)

BI & Visualization: Power BI (Power Query, DAX, semantic models, Power BI Service), Evidence.dev, Tableau, D3.js

ML & Backend: pandas, NumPy, scikit-learn, XGBoost, PyTorch, Hugging Face, LangChain, FastAPI, Qdrant

Concepts: ELT/ETL, data warehousing, dimensional modeling, slowly changing dimensions, incremental and idempotent loads, data contracts, data quality testing, source freshness monitoring, CI/CD, data governance, API integration, workflow orchestration, time-series forecasting, retrieval-augmented generation, distributed systems