# OZAN OZBEKER

Pittsburgh, PA · 304-651-6856 · contact@ozanozbeker.com · linkedin.com/in/ozanozbeker · ozanozbeker.com

*Industrial Engineer turned Analytics Engineer running solo on Fortune 500 client engagements.*

## Experience

### Data Science Consultant

**OneMagnify (formerly Splash Analytics, acquired 2023)**, Remote, January 2023 to Present

- **Run the day-to-day client relationship for two flagship Fortune 500 engagements.** Lead the standing cadence (bi-weekly and monthly), scope new work directly with manager-level stakeholders on the call, and own delivery end to end across data engineering, analytics engineering, modeling, and reporting. Stack is Python, SQL, and Quarto; infrastructure choices, package boundaries, and orchestration patterns have been mine to make.
- **Own the recurring reporting cadence for two engagements: monthly data refreshes, weekly and monthly Excel reports for stakeholders who want raw data, Quarto summary-over-time reports for stakeholders who want charts, and ad-hoc analyses on request.** Pipelines that originate with client-sent files have a manual kickoff (clients send human-generated CSVs and Excel files at irregular intervals, with quality issues every cycle); everything downstream of that is automated, idempotent, and reproducible via `uv run`. Pipelines that work off existing data or my own scrapes run fully unattended on Windows Task Scheduler. Production code is scripts, never notebooks.
- **Took ownership of a flagship customer-engagement-scoring pipeline (26M US and 8M Canada customers, every Sunday) when the IBM DB2 driver broke under R/dbplyr and nobody else on the team knew R.** Rebuilt the pipeline twice under production pressure, first to Python/Polars to unblock, then to native DB2 SQL once the intermediate Python steps proved unnecessary. Final architecture: Python orchestrates table rotation and run scheduling, all transformation logic runs as DB2 SQL on the client database. Cut runtime from 16 hours to 4 by eliminating the data-transfer tax; data quality checks run at every stage.
- **Designed and built `oxy`, a Python CLI tool (Typer-based) that replaced a fragile Selenium-on-Helium-10 setup with an async `httpx` wrapper around the Oxylabs HTTP API.** Workloads that previously took ~5 days of unattended browser automation now finish in roughly 5 minutes; ~20,000 scrape calls a month run unattended across 5 internal projects for 2 client engagements. Two CLI namespaces: `oxy scrape` for ingestion, `oxy build` for datalake operations. Raw JSON preserved verbatim; processed layer materialized to Parquet; downstream reports query the model layer directly.
- **Built `omutil`, an internal Python package distributed via GitLab with optional dependency groups (`omutil[database]`, `omutil[email]`, `omutil[brand]`).** Standardized client-database connections (SQLAlchemy connection objects keyed to each client database; ADBC as the default transport for substantial speedups over ODBC), Quarto report branding, and structured logging across the team. Used by every project I own and adopted by colleagues for the email submodule specifically.
- **Replaced the team's Helium 10 sales-estimate dependency with `amazon-atlas`, an in-house LightGBM model that predicts rolling 30-day Amazon unit sales for any ASIN.** Distributed as both an installable Python package and a FastAPI service; trained on 2.5 years of historical data to keep client reporting continuous. Per-client calibration where ground-truth sales data is available; documented "this is an estimate" framing in client deliverables where it is not. Eliminated a recurring SaaS subscription and an operational dependency on a brittle browser-extension product.

### Adjunct Instructor

**West Virginia University**, Morgantown, WV, January 2025 to Present

- **Sole instructor of record for ~150 students across two academic years for IENG 331 (Computer Applications in Industrial Engineering).** Redesigned the curriculum from basic Python alone into a more complete computing course covering Python, modern Python tooling (`uv`, Ruff, basedpyright), terminals and CLIs, SQL with DuckDB, and reproducible reporting with Quarto. Authoring a free public Quarto book at ozanozbeker.com/cw4e as the course's textbook and a working artifact for documentation discipline.

### Operations Leadership Development Program (Business Data Analyst, Manufacturing Engineer)

**Xylem, Inc.**, Uniontown, PA and Morton Grove, IL, July 2021 to December 2022

- **Standardized the annual physical-inventory-counting process across three manufacturing facilities producing different products under different VPs.** Ran discovery interviews with site leads, shipping managers, and floor team leads; designed a single workflow adopted by all three facilities; recovered 0.5 to 1 day of full production capacity per facility per year.
- **Automated quarterly obsolete-and-excess inventory reporting for three site materials managers.** Replaced 3 days of manual ERP work per manager per quarter with a SQL pipeline against the ERP-mirror Postgres database and a Tableau dashboard updating weekly.

## Education

**Bachelor of Science, Industrial Engineering**, West Virginia University, August 2017 to May 2021

## Skills | Tech Stack

- **Daily:** Python (Polars, DuckDB, Typer, marimo, httpx), SQL (DB2, Postgres, SQL Server, DuckDB), Quarto, uv, Ruff, basedpyright, Git, JSON, Parquet, NDJSON, ETL/ELT, REST APIs
- **Comfortable:** SQLAlchemy, FastAPI, pydantic, ADBC, Selenium, LightGBM, HTML/CSS, R (Tidyverse), Excel, Power BI, Shiny, Linux, Bash
- **Learning:** Dagster, dbt, dlt