Gregory Joshua

Data Analyst | Business Intelligence & Operations Analytics

gregjoshua11@gmail.com | github.com/0gregoryj | linkedin.com/in/gregjoshua | gregoryjoshua.com | (650) 924 – 8937 | San Mateo, CA | Open to Relocation

Business Analytics master's candidate graduating December 2026 with experience translating operational data into decision-ready KPIs, dashboards, and recurring reports using SQL, Python, Tableau, Excel, PostgreSQL, and dbt. Very passionate about data, and eager to learn and contribute wherever I can.

Professional Experience

Data Analyst & Lead Instructor

Mathnasium | San Francisco, CA | August 2024 – Present

- Built a Python, PostgreSQL, and dbt reporting workflow for student, session, and topic data, producing standardized KPIs and visualizations that surfaced performance gaps across 5–10 topics and dozens of students.
- Equipped center directors with performance views used several times per week to monitor overall center performance, inform email and in-person stakeholder communications, and support decisions about instructional priorities, curriculum progression, and scheduling for ~250 students.
- Supported and maintained existing workflows for tracking individual students' curricula and progress, made decisions about instructional content and course progression, scheduled students and instructors, instructed students on concepts from numerical fluency to pre-calculus, and managed front desk and customer success operations.

Logistics Associate

MedGenome, Inc. | Foster City, CA | June 2021 – June 2022

- Extracted and analyzed LIMS data in Excel to track project duration and client distribution, communicating findings through scheduled presentations and in response to ad-hoc requests; analyzed project lifecycle data to optimize sample archiving timing, increasing storage efficiency by ~10%.
- Served as a liaison among clients, research teams, and operations; coordinated cross-functional workflows, maintained compliance logs, tracked samples and supplies, and supported timely project completion and data delivery.

Projects

Retail Rocket ELT Pipeline and Data Warehouse

July 2026 – August 2026 | Tools: Python (PySpark, pytest), PostgreSQL, dbt/MetricFlow, Lightdash, Docker

- Built a containerized ELT pipeline for 23M+ rows of e-commerce data, using PySpark, PostgreSQL, and dbt to create a BI-ready galaxy schema for reporting and self-service analysis; implemented ingestion validation, pytest coverage, and dbt model tests to ensure data reliability.
- Derived 1.76M visitor sessions using 30-minute inactivity logic and resolved category hierarchies with recursive SQL; defined standardized conversion, engagement, and cart-abandonment metrics in MetricFlow. Used Lightdash to analyze session conversion rate and found that most abandoned-cart sessions lasted under three minutes, highlighting session engagement as a potential conversion lever.

Bay Wheels Operational Data Analysis

June 2026 – July 2026 | Tools: Python (pandas), Parquet, Tableau, Jupyter

- Analyzed and validated 10.6M bike-share trips from 30 monthly CSVs using pandas, producing a cleaned Parquet dataset and four reusable summary tables to evaluate station rebalancing, hourly demand, rider segments, and growth trends.
- Built a Tableau dashboard and executive summary that identified two stations with approximately 1,000+ net trip outflows and two with approximately 600 net inflows, compared member-versus-casual demand by location and time, and recommended targeted bike redistribution.

Education

M.S, Business Analytics - San Francisco State University | San Francisco, CA | August 2024 - Expected December 2026

B.S, Mathematics - University of California, Santa Cruz | Santa Cruz, CA | June 2022 – August 2024

A.T, Economics - Cañada College | Redwood City, CA | June 2020 – June 2022

Technical Skills

Business Analysis & BI: Microsoft Excel, Tableau, Lightdash, Streamlit, SAP S/4HANA, Odoo ERP, dashboard development, data visualization, KPI design, ad-hoc analysis and reporting, executive summaries, process improvement, stakeholder communication

Analytics Engineering: Python (pandas, PySpark, pytest, Jupyter), SQL, PostgreSQL, dbt, dimensional modeling, ELT/ETL pipelines, data quality testing, semantic modeling, metric definition, Apache Airflow, Docker, Git/GitHub