David Do

858-905-4168 | and013@ucsd.edu | www.linkedin.com/in/davidanhdo

Education

University of California San Diego San Diego, CA

Bachelor of Science in Applied Mathematics, Bachelor of Science in Economics, GPA: 3.7 *Sep. 2023 – June 2027*

Relevant Coursework: Honors Linear Algebra, Honors Multivariable Calculus, Probability, Programming and Computational Problem-Solving, Numerical Analysis

Technical Skills

Languages/Frameworks: Python, Java, SQL, C#, HTML/CSS, JavaScript, DAX
Developer Tools/Libraries: Pandas, NumPy, Matplotlib, Git, Power BI, Excel, SQLite, Vega-Lite, API, Jupyter

Experience

Data Engineer Intern June. 2025 – Present

AllianceChicago *Illinois, Chicago*

- Supported the cloud migration of Alliance's enterprise data warehouse from SQL Server to Snowflake; developed ETL scripts using SQL and SnowConvert to automate data transfers and optimize future platform scalability.
- Refactored legacy SQL queries and stored procedures for Snowflake compatibility; implemented Git workflows to version control all changes, improve reproducibility, and streamline collaboration across distributed data teams.
- Verified data consistency through cross-platform reconciliation and automated checks; improved analytics pipeline reliability, ensuring accurate reporting and dashboard performance for executives and downstream business intelligence users.

AI & Data Science Intern July, 2025 – Present

University of California, San Diego *San Diego, California*

- Gained hands-on experience in relational databases and enterprise-scale SQL development; wrote and optimized complex structured queries to support backend data workflows across large-scale application systems.
- Explored RESTful API architecture and system-level application design; analyzed internal data flow, modularity, and authentication protocols to understand how distributed systems communicate and scale in production.
- Assisted in LLM-based agent development for internal AI initiatives; structured dynamic prompt chains, integrated external tool use, and supported orchestration of autonomous workflows for advanced business automation.

Economics Research Assistant Mar. 2025 – Present

University of California, San Diego *San Diego, California*

- Collected and digitized historical datasets on gender-based violence and missionary activity in China, ensuring data accuracy and completeness for econometric analysis of religion, women's labor, and cultural development over time.
- Used Python for parsing and tokenizing textual transcripts, applying stopword filtering and frequency-based keyword extraction to analyze religious, economic, and cultural themes with interpretive rigor.
- Assisted in statistical analysis by cleaning and transforming datasets; applied Stata and R to conduct regression analysis, contributing to modeling of religious broadcasting, labor outcomes, and cultural change dynamics.

Projects

End-to-End Amazon Product Data Pipeline July 2025

- Built a scalable Python ETL pipeline using pandas, os, and re to clean 1.1M+ Amazon product records, applying null filtering and multi-stage deduplication to generate a structured dataset of 268K unique products.
- Designed a normalized SQLite database with products and amazon data tables, using SQL JOINs and aggregation to calculate discount percentages, review counts, and export a distinct, analysis-ready dataset for Excel.
- Developed an interactive Excel dashboard using PivotTables and PivotCharts, transforming disorganized product data into a clear, filterable interface for exploring trends in price, ratings, and category segmentation

Spotify Engagement & Metadata Analytics July 2025

- Built a Python data pipeline using requests, pandas, and Spotify's API to extract track metadata and album cover URLs for 1,000+ songs. Exported clean dataset for dashboard with full API integration and error handling.
- Designed an interactive Power BI dashboard featuring KPI cards, Deneb heatmaps, DAX measures, and HTML-based image rendering to highlight top-streamed tracks and reveal patterns through interactive slicers.
- Delivered a full-stack music analytics solution that turned raw Spotify data into an insightful, engaging tool. Enabled pattern recognition across time, artists, and stream counts - showcasing data storytelling skills.