

Stephanie Patricia Anshell

sanshell@ucsd.edu | (925) 348-7120 | [linkedin.com/in/stephaniepatriciaanshell](https://www.linkedin.com/in/stephaniepatriciaanshell) | github.com/stephaniepatriciaans | stephaniepatriciaans.github.io/portfolio

Data Science student at UC San Diego with a Finance minor and experience applying machine learning, analytics, automation, and business analysis across real estate, supply chain, and financial operations. Built end-to-end data pipelines, predictive models, dashboards, and AI-powered tools that translate complex problems into measurable business outcomes. Proficient in Python, SQL, Excel, Tableau, and machine learning. Fluent in English and Indonesian.

EDUCATION

UNIVERSITY OF CALIFORNIA SAN DIEGO

Bachelor of Science in Data Science and Minor in Finance | Cumulative GPA: 3.6

La Jolla, California, USA

Expected 2026

UNIVERSITY OF CALIFORNIA BERKELEY EXTENSION

Certificate Data Analytics, Data Analytics Bootcamp

Berkeley, California, USA

November 2023

SKILLS

AI / LLM: GenAI, Claude API, Prompt Engineering, Agentic Pipelines, RAG, Tool Use / Function Calling, Multi-step LLM Workflows

Languages: Python, SQL, JavaScript, MATLAB, C++

ML / Data: scikit-learn, TensorFlow, pandas, NumPy, SciPy, GeoPandas, PySAL, Matplotlib, Seaborn, PyTorch, Spark

Methods: Hypothesis-Driven Problem Solving, ML, Deep Learning, NLP, Predictive Modeling, Time-Series Forecasting, A/B Testing, Feature Engineering, ETL, Data Visualization, Business Strategy, Business Transformation

Tools: Tableau, Streamlit, Jupyter, Git, Linux, ROS, AWS, Microsoft 365 (Excel, SharePoint, Teams)

Automation: API Integration (REST), Data Pipeline Automation, Workflow Automation, Cross-Functional Stakeholder Collaboration

TECHNICAL EXPERIENCE

UC SAN DIEGO TECHNICAL TEAMS (Triton AI Member & Autonomous Racing Team)

La Jolla, California

ROS, Linux, Python, Robotics, Autonomous Systems, Simulation | Autonomous Systems / Robotics Contributor

January 2026 - Present

- Integrated and tested the ROS 2 autonomous go-kart stack on NVIDIA Jetson using Linux and Docker; configured launch workflows and debugged sensor and vehicle-control nodes
- Resolved recurring sensor and control-node faults across repeated track-test sessions, cutting pre-race integration downtime and contributing to a 2nd-place finish among 9 teams at the 2026 Autonomous Karting Series at Purdue University.

EXPERIENCE

IDX Exchange (Real Estate Data Science Platform)

Remote, USA

Data Scientist Intern

September 2025 - December 2025

- Engineered a client-facing ML pipeline (XGBoost, Random Forest, Linear Regression) integrated via Cotality's Trestle API, achieving $R^2=0.89$ and ~6% MDAPE across 5+ home-price segments, then translated results into executive decision briefs that shaped pricing strategy.
- Designed a standardized validation workflow across 5+ client experiments (EDA → feature engineering → tuning → evaluation), defining test criteria at each stage to ensure accuracy before delivery, cutting turnaround time 40%.

GAF (Authorized distributor of GAF Roofing materials in Indonesia, focused on supply chain operations)

Jakarta, Indonesia

Data Scientist Intern

January 2024 - September 2024

- Structured an ambiguous "reduce stock-outs" business question into a measurable forecasting problem; built ARIMA and Decision Tree models on 9 months of procurement data, lifting forecast accuracy 15% and cutting stock-outs 20% for purchasing stakeholders.
- Built and maintained automated data ingestion pipelines feeding Tableau dashboards across 8+ procurement and inventory metrics, eliminating manual reporting errors and saving 3 hrs/week while improving data quality and definition consistency across sources.

STAL Corporation (A printing supplier for Indonesian hospitals and clinics)

Jakarta, Indonesia

Accounting & Data Analyst Intern

Jun 2023- Aug 2023 · Hybrid/Remote

- Drove a 15% cut in procurement costs by identifying hidden over-ordering patterns through time-series analysis, informing corrective purchasing policy across finance and operations.
- Recovered 4 hrs/week in manual reconciliation time by designing automated Excel-based expense-tracking and budget-planning workflows, strengthening financial data modeling precision and reporting consistency across monthly cycles.

PROJECTS

Menu Engineering Bot | Python, pandas, NumPy, Streamlit, Claude API | [GitHub : https://github.com/stephaniepatriciaans/menu-engineering-bot](https://github.com/stephaniepatriciaans/menu-engineering-bot)

- Built a Streamlit menu analytics tool analyzing 180 days of cafe sales across 15 items, classifying products into Stars, Plowhorses, Puzzles, and Dogs and simulating -10% to +15% price changes
- Integrated Claude API to translate quantitative results into owner recommendations, with the sample analysis projecting \$47.07/day in additional profit.

Loan Charge Offs | Python, scikit-learn, TensorFlow | [GitHub : github.com/stephaniepatriciaans/Project-Loan-Charge-Offs](https://github.com/stephaniepatriciaans/Project-Loan-Charge-Offs)

- Built an ETL pipeline on FRED data (390 rows, 1991-present) and compared 4 ML methods
- Random Forest hit 93.91% accuracy, with threshold analysis for business precision-recall tradeoffs.
- Constructed an end-to-end ETL pipeline transforming monthly/quarterly FRED data into a unified, quality-checked modeling dataset, ensuring consistent data definitions across economic cycles and enabling reproducible credit risk analysis.

Urban Transit & Crime | GeoPandas, PySAL | [GitHub : github.com/stephaniepatriciaans/Urban-Transit-and-Crime](https://github.com/stephaniepatriciaans/Urban-Transit-and-Crime)

- Led a team of 5 to integrate 1M+ records across 4 public datasets (CTA, MTA, NYPD, CPD), directing the build of spatial regression models (GeoPandas, PySAL, spreg) that detected statistically significant transit-crime clusters across Chicago and NYC.

TrueFork | Python, scikit-learn, Pandas, JavaScript | [GitHub : github.com/stephaniepatriciaans/TrueFork](https://github.com/stephaniepatriciaans/TrueFork) | [Live Site :](https://stephaniepatriciaans.github.io/TrueFork/)

- Built and tested an NLP pipeline on 80K+ Food.com review records, validating sentiment classification accuracy via A/B testing ($p<0.05$) before shipping as a live, user-facing public site, reducing simulated decision time ~50%.

Adult Obesity Prevalence | Python, scikit-learn, pandas | github.com/stephaniepatriciaans/adult-obesity-prevalence

- Independently built and validated a regression pipeline (OLS, Ridge, Lasso) on CDC PLACES data across 2,299 counties and 7 predictors, achieving $R^2=0.628$ via 5-fold CV, coordinating with teammates on data sourcing and reporting.