

Christopher Peterson

✉ chris.will.peterson@gmail.com [in](https://www.linkedin.com/in/chriswillpeterson) [chriswillpeterson](https://www.linkedin.com/in/chriswillpeterson) [ptechr](https://github.com/ptechr) [📍](#) Redlands, CA

Summary

Business Intelligence Engineer with 7+ years of experience translating large-scale operational data into actionable insights and scalable analytics systems at Amazon. Designed real-time data platforms, automated ETL pipelines, and forecasting models supporting \$1B+ business decisions and 165+ stakeholder teams. Combines strong data engineering fundamentals (Spark, SQL, AWS, distributed systems) with advanced statistical modeling and stakeholder-facing tools to deliver measurable outcomes.

Experience

Senior Business Intelligence Engineer

[Amazon](#)

📅 July 2024 – Present [📍](#) Virtual Location California (Remote)

- Designed and scaled a real-time distributed data ingestion platform processing 1.3B+ events per day with zero data loss, reducing inventory data latency from 24 hours to 10 minutes through a hand-crafted ETL pipeline using SNS, SQS, ECS, Kinesis, DynamoDB, & S3.
- Scaled team-owned Parquet-based data lake by replacing legacy snapshot pipelines with fully stream-derived datasets featuring optimized partitioning, deduplication, enrichment, and schema design—eliminating race conditions, boosting query performance, and driving adoption to 178 production tables (4,350% growth from 2024), generating 1,200 access requests across 165 teams in 2025.
- Developed a Spark-based distributed life-cycle engine processing 900M records/day, constructing end-to-end item lineage with deterministic hashing and FIFO-based indexing using a dual-tiered maximum bi-partite matching algorithm. This generated life-cycle completeness to 97.8% and reduced lineage computation from multiple weeks of developer effort to minutes, enabling scalable inventory tracing across Amazon as a generalized inventory-tracing-as-a-service model.
- Translated life-cycle and defect prediction signals into data-driven operational decisions using automated validation, confidence-interval sampling, and prioritization logic. Delivered \$21.8M in realized savings in 2025 (\$34M annualized) and improved network defect remediation efficiency by 334%, validating model-driven interventions.
- Implemented CI/CD governance and version-controlled deployments for over 1,700 production data pipelines, enforcing code review verification and locked UI edits. Reduced deployment risk and improved reliability across a highly interdependent data ecosystem serving hundreds of teams.

Business Analyst

[Amazon](#)

📅 March 2022 – July 2024 [📍](#) Virtual Location California (Remote)

- Built and deployed a North America inventory labor forecasting system managing a \$1.3B budget, delivering 2.8M metric forecasts across 320 fulfillment centers over a 104-week horizon. Engineered fully automated ETL pipelines for 22k KPIs and designed dynamic Sagemaker time series models supporting multiple prediction types (Linear Regression, SARIMAX, KNN, Random Forest, XGBoost) with hierarchical reconciliation, hyperparameter optimization, and residual course correction. Developed custom EDA and visualization notebooks analyzing seasonality, correlations, and network-FC dynamics. Delivered a stakeholder-facing labor planning web portal with override

capabilities, algebraic optimization recalculation, and network-level reporting, enabling actionable, constraint-aware workforce planning.

- Delivered a rapid-response tooling solution that eliminated schema deployment friction in Amazon's centralized data lake, which automated transformation from Redshift to SDL schemas with full support for constraint keying and data type translations. This reduced dataset onboarding time $6\times$ and scaled organically to 2,000+ engineers across the Software Development, Data Engineering, Business Intelligence, and Data Science job families.

Data Analyst

[Amazon](#)

📅 August 2018 – March 2022 📍 Redlands, CA

- Built a quarterly logistic regression model to detect shift false pick shorts in real time, achieving 79% precision and enabling targeted coaching. Drove site to become the top-performing building in the region for pick short DPMO for all of Q4 2021.
- Developed a supervised multinomial classification model to identify root cause work reasons and associate responsibility for bin inventory adjustments, increasing regional classification accuracy from 3.5% (manual) to 52% via automated workflow.
- Designed a bias-free quality coaching application that measures pre- and post-coaching performance, enabling data-driven manager-associate conversations. In 2021, 75.6% of coachings led to improvement, with median DPMO reductions of 49.55%.

Skills

- **Programming/Scripting:** Python, TypeScript, HTML, CSS
- **Version Control:** Git, ECR, Docker
- **Infrastructure as Code (IaC):** AWS CDK
- **Data Engineering/ETL:** PySpark, SQL, SQS, SNS, Kinesis, DynamoDB, Lambda, S3, ECS, Redshift, EMR, Partition Optimization, Schema Design
- **Statistics/Model Selection:** VIF, AIC/BIC, Adjusted R-Squared, SHAP, PCA, Confidence Intervals, Hypothesis Testing, Paired t-tests, A/B Pilot Selection/Validation, Autocorrelation, Correlation, Probability Theory, Combinatorics, Augmented Dickey-Fuller Test, Data Interpolation, Hyperparameter Grid Searching, Recursive Feature Elimination, Train/Test Splits
- **Predictive Modeling:** Sagemaker, Timeseries, Sklearn, Statsmodels, PyTorch, SARIMAX, XGBoost, Random Forest Regression, Linear Regression, Logistic Regression, Lasso Regression, K-nearest Neighbor, Artificial Neural Network
- **Visualization:** Quicksight, Matplotlib, AG Charts, Jupyter Notebook

Education

Bachelor's of Business Administration, Business Management (BBA)

[Marywood University](#)

📅 2012 – 2015 📍 Scranton, PA

- Placed 2nd in Business Accreditation Assessment
- 3.76 Business GPA
- Dean's List Recognition
- NCAA Baseball Player