

CHRISTOPHER TON

christopher.ton@sjsu.edu — (669) 254-6967 — linkedin.com/in/chriztopherton — github.com/chriztopherton — US Citizen

Professional Experience

Genentech

San Diego County, CA

Senior Data Engineer

Apr 2024–Present

- **Technical Leadership:** Lead end-to-end architecture design for complex healthcare data solutions, mentoring junior engineers and collaborating with cross-functional teams to deliver robust, scalable applications that support critical clinical decision-making processes.
- **Advanced Data Engineering:** Engineered high-performance data models in Snowflake using PySpark, SQL, and dbt across 7+ global healthcare data sources; implemented comprehensive testing frameworks and CI/CD pipelines via GitLab to ensure data quality and compliance with healthcare regulations.
- **Complex Problem Solving:** Designed and deployed scalable ETL pipelines processing terabytes of clinical documentation into Azure AI Search, enabling semantic search through RAG workflows; troubleshooted and optimized performance bottlenecks, reducing query latency by 40%.
- **Cloud Infrastructure & DevOps:** Developed serverless APIs using AWS Lambda, S3, Bedrock, and OpenSearch supporting multiple RAG applications; implemented automated testing, monitoring, and deployment strategies ensuring 99.9% uptime for critical healthcare applications.

Data Scientist

Jun 2022–Apr 2024

- **Healthcare Analytics:** Reduced clinical resolution times by 20% through development of predictive anomaly detection models using XGBoostRegressor, partnering with C3.AI to deliver insights that directly impacted patient care workflows.
- **Data Quality & Validation:** Automated clinical reporting validation processes by building sliding-window search modules in R, ensuring data accuracy and compliance with healthcare data standards while reducing manual effort by 60%.

Deloitte

Remote

Data Engineer via Brooksource

Feb 2022–May 2022

- **Data Security & Compliance:** Deployed data pipelines using Informatica TDM to automate sensitive data masking, ensuring compliance with data protection regulations (GDPR, HIPAA) and minimizing security risks for enterprise clients.
- **Advanced Analytics:** Developed Python-based solutions for data distribution modeling and population parameter analysis, enabling comprehensive risk assessment for data re-identification through quasi-identifier and pseudo-anonymization techniques.

Shell Recharge (formerly Volta Charging)

San Francisco Bay Area County, CA

Data Engineer

Nov 2021–Mar 2022

- **ETL Pipeline Development:** Automated and optimized data ingestion workflows by developing Python scripts to extract, transform, and curate datasets from multiple vendor sources, achieving 30% improvement in data quality and observability.

Data Marketing & Insights Intern

Jul 2021–Sep 2021

- **Business Intelligence:** Enhanced EV network planning efficiency by conducting in-depth exploratory analysis with KMeans clustering, delivering actionable insights that informed strategic growth decisions and real-time performance monitoring.

Technical Publications

arXiv: A Comparison of LLM Finetuning Methods & Evaluation Metrics with Travel Chatbot Use Case

Implemented and evaluated various fine-tuning techniques including Supervised Fine-tuning (SFT), Retrieval Augmented Fine-tuning, and Reinforcement Learning with Human Feedback (RLHF), demonstrating advanced machine learning expertise.

Education

San Jose State University
Master of Science, Data Analytics

San Jose, CA
Graduated May 2024

University of California, Davis
Bachelor of Science, Statistics

Davis, CA
Graduated Jun 2020

Technical Skills

Programming Languages: Python, Scala, SQL, R

Big Data & ETL: Apache Spark, PySpark, Pandas, NumPy, dbt, Talend, Informatica

Cloud Platforms: AWS (S3, Lambda, Redshift, Bedrock), Azure AI Search, GCP BigQuery, Snowflake

Data Formats: Parquet, Avro, Delta Lake, JSON, CSV

DevOps & CI/CD: GitLab, GitHub, Docker, Kubernetes, Automated Testing

ML & AI: SageMaker, LangChain, PyTorch, scikit-learn, XGBoost

Analytics & Visualization: Tableau, Jupyter Notebooks, R Shiny

Project Management: JIRA, Confluence, Agile Methodologies