

ANDY YANG

+1-647-461-6466 • syyyang003@gmail.com • Toronto, ON / Vancouver, BC • [LinkedIn](#) • [Portfolio](#) • [GitHub](#)

PROFESSIONAL SUMMARY

Data Scientist & Analytics Engineer with 5 years of experience designing automated ML pipelines, building cloud-native analytics solutions, and validating high-impact models in financial services and tech. Adept at automating manual workflows using Python, SQL, and cloud platforms (Hadoop, Azure, AWS). Proven track record driving decision-making through predictive modeling, LLM automation, and stakeholder engagement.

CORE COMPETENCIES

Coding Languages: Python (Plotly, Pandas, PyTorch), R (Rshiny), Jupyter Notebooks, PySpark, SAS, SQL
Machine Learning: Predictive Modeling, Classification, Feature Engineering, A/B Testing, Hyperparameter Tuning, LLMs
Tools and Platforms: Azure (Databricks), AWS (S3, Bedrock, Lambda, SageMaker), Tableau, Power BI, Docker, Jupyter
Soft Skills: Stakeholder Communication, Critical Thinking, Data Storytelling, Presentation, Agile Collaboration

WORK EXPERIENCE

TD Bank Group - Toronto, ON

Senior Model Validation Analyst – Data Science

Apr 2023 - Present

- Designed scalable Python and SQL-based ETL pipelines to automate ingestion of loan portfolio data into Tableau dashboards, improving reporting cadence and model governance for \$10B+ in assets.
- Built PySpark pipelines on Azure Databricks to evaluate model performance (AUROC, WAPE, calibration, Stress Testing), cutting runtime by up to 50% and increasing evaluation coverage.
- Developed benchmark supervised learning models (XGBoost, logistic regression) to forecast consumer default risk, and exposure at default, maintaining mean error under 10% and supporting compliance reviews.

TD Bank Group - Toronto, ON

Model Validation Analyst – Data Science

Sep 2022 – Apr 2023

- Collaborated with data scientists and risk analysts to validate production-grade ML models on millions of records, uncovering fraud trends and behavioral insights critical for real-time risk response systems.
- Standardized analytics toolkits using Python, RShiny, and SAS, and implemented CI/CD workflows via GitHub Actions, reducing deployment lead time by 50% and enhancing onboarding for new users.
- Translated complex analytics results into actionable business insights and presented analytics code demos at executive town halls, aligning stakeholder decisions with model performance metrics for non-technical audiences.

Royal Distributing - Guelph, ON

API Implementation Specialist

Apr 2021 – Sep 2021

- Deployed cloud-based internal APIs on Microsoft Azure to automate manual data sync tasks, boosting availability for supply chain risk dashboards and expediting data transfer from legacy FoxPro databases.
- Rebuilt ETL pipelines and redesigned SQL schemas, delivering a 200%+ reduction in query latency.
- Acted as primary liaison between internal dev teams and third-party vendors to troubleshoot and resolve data inconsistencies and integrity issues, leading to a reduction in data-related issues by 80% in the following quarter.

PROJECTS

LiDAR Object Detection and Classification for Cities

PyTorch, AWS (S3, SageMaker), Deep Learning

Reliance Foundry Co. Ltd., University of British Columbia

- Built a deep learning model using PyTorch CNN architecture to classify urban objects in 3D LiDAR point clouds, achieving >93% precision in city planning use cases; deployed training and inference pipelines to AWS SageMaker.

AWS Flask Portfolio with LLM Chat Bot

AWS (Bedrock, Lambda, EC2), Python Flask, Docker

- Designed and deployed a full-stack Flask web portfolio app on AWS EC2 with integrated LLM chatbot powered by Meta's Llama 3 via Bedrock API.

EDUCATION

Master's of Data Science

Sep 2021 – Jul 2022

University of British Columbia, Vancouver, BC

Bachelor's of Applied Science in Chemical Engineering

Sep 2016 – Apr 2021

University of Waterloo, Waterloo, ON