

Andy Yang

Data Scientist

+1 647-461-6466 • syyyang003@gmail.com • www.linkedin.com/in/andyyang80 • andyyang.ca • Vancouver, BC

Professional Summary

Data Scientist & Analytics Engineer with 5 years of experience designing automated ML pipelines and building high-impact AI 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. Master's in Data Science from UBC.

Skills

Coding Languages: Python (Plotly, Pandas, PyTorch), R (Rshiny), Jupyter Notebooks, PySpark, SAS (certified in 9.4), 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

Data Scientist

TFG Financial | Vancouver, BC

Jan 2026 - Jun 2026

- Engineered data quality ETL pipelines (SQL, Python, Tableau) to detect Scorecard and PD anomalies, flagging ~10% of the portfolio in a weekly dashboard used by 10+ stakeholders across US and Canadian credit leadership.
- Developed Cox Proportional-Hazards PD survival models (scikit-learn, lifelines) with engineered macroeconomic features to generate early default warnings at the account level for multiple portfolios, achieving a C-index of 0.81.

Senior Quantitative Analyst, Model Validation

TD Bank Group | Toronto, ON

Apr 2023 - Dec 2025

- Developed a custom Python GUI application to automate data ingestion for model performance reporting across \$5B+ portfolios, reducing quarterly reporting cycle time by 2+ weeks for senior stakeholders.
- Built PySpark pipelines to supersede SAS scripts on Azure Databricks that evaluate model performance (AUROC, WAPE, calibration, Stress Testing), expanding testing coverage and cutting runtime by up to 2 hrs per validation.
- Trained benchmark supervised learning models (XGBoost, logistic regression) on Azure ML to forecast default risk in multiple credit card portfolios, achieving MAPE metrics under 10% and supporting OSFI audit compliance.

Quantitative Analyst, Model Validation

TD Bank Group | Toronto, ON

Sep 2022 - Apr 2023

- Validated over 15 production-grade ML models on millions of records, uncovering behavioral insights critical for real-time risk response systems, collaborated with data scientists to implement adjustments to PD models.
- Standardized model analytics toolkits using Python, RShiny, and SAS, and implemented CI/CD workflows via GitHub Actions, reducing QA and deployment lead time by 50% and enhancing onboarding for new users.
- Lead department town halls, aligning model performance with business decisions for non-technical audiences.

API Implementation Specialist

Royal Distributing | Guelph, ON

Apr 2021 - Sep 2021

- Deployed cloud-based internal REST APIs on Microsoft Azure to automate manual data sync tasks, enabling automatic supply chain risk dashboards and automatic data migration from legacy FoxPro databases.
- Refactored schema of SQL Server Databases and revised query criteria, cutting query latency in half (~30 mins).

Projects

LiDAR Object Detection for Cities (UBC Capstone)

Technologies: AWS (SageMaker, Lambda, EC2), PyTorch

- Trained deep learning model using PyTorch CNN architecture to classify urban objects in 3D LiDAR point clouds, achieving 93% precision within parking lot environments. Deployed to AWS SageMaker for mobile inferencing.

Education

Master's of Science in 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