

Aamodit Acharya

📞 416-617-5955 | ✉ aamodit.acharya@uwaterloo.ca | 🏠 aamodita.com | 📱 ammo-byte | 🌐 aamodit-acharya

Education

University of Waterloo

Waterloo, ON

Bachelor of Statistics with Computer Science Minor (Co-op)

Expected April 2026

- **Courses:** Data Analysis, Stochastic Processes, Sampling and Experimental Design, Applied Linear Models, Investment Science
- **Certifications:** Supervised Machine Learning (Deeplearning.ai), SQL for Data Analytics (Coursera), SOA Exam FM and SOA Exam P

Work Experience

Statsig

Bellevue, WA

Data Science Intern

May 2025 - August 2025

- Working on implementing the HYperLogLog++ algorithm using **PySpark (Python)** and **BigQuery (SQL)** to build a scalable count-distinct metric, and leveraging KLL sketches for approximate percentile estimation within Statsig's data warehouse platform.

TD Bank

Montreal, QC

Data Science Intern

Jan 2024 - Apr 2024

- Engineered a predictive tool using **Impala (SQL on Hadoop)** with automated data pipelines to restructure and update contract data for real-time monitoring of client-contract cancellations, enabling the residential pricing team to boost customer retention by 22%.
- Optimized data extraction by converting legacy **R code** into **SQL scripts** to retrieve latest website data, reducing processing by 97%.
- Integrated **Excel PivotTables** using **Power Query**, improving the **ETL** process, automating data aggregation and enabling real-time competitor quoted-premium analysis to enhance market positioning, driving a 16.5% increase in Q2 residential insurance premiums.
- Deployed a **Python** web scraping tool on **AWS EC2**, leveraging **Selenium**, **BeautifulSoup** and **Pandas** to aid in pricing decisions.

TD Bank

Toronto, ON

Data Science Intern

May 2023 - Aug 2023

- Designed a **Python framework** to forecast customer and call center metrics using **Pandas** and **Numpy**, with **SQL** for raw data extraction and organization; utilized **Docker** and **Kubernetes** for deployment on **AWS EC2**, reducing overall forecasting time by 67%.
- Created a **modular framework** to forecast call center staffing needs for the next year, using historical trends; projected to save \$5mm.
- Implemented **RidgeCV** and **linear regression** using **Scikit-learn** to forecast agent closing rate, improving model performance by 25%.
- Aggregated legacy Cisco data with Genesys data using **Hive SQL on Hadoop** and partitioned tables, cutting report generation by 50%

Desjardins

Toronto, ON

Data Science Intern

Sep 2022 - Dec 2022

- Developed **SAS code** for the new-business-progress report, which identified discrepancies in insurance quotes caused by incorrect information from brokers or customers, resulting in a 30% improvement in fraud detection accuracy.
- Streamlined premium data extraction for the Nova Scotia Pre-NeXT initiative, an online insurance premium quoting tool, by extracting profile data with **WTW-Radar** and rating it using a **Microsoft Excel** macro to train the risk scoring algorithm.
- Conducted a review of the New Brunswick Segmentation model, ensuring alignment between **WTW-Radar** and statistical **R models**, resulting in a 10% improvement in accuracy across 10 risk factors.

Projects

- **What Game?** 🔄 : Developed a **Python** and **Streamlit** app using **cosine similarity** with **scikit-learn** to recommend games based on IGDB API data, enhanced with **seaborn** visualizations.
- **What Image?** 🔄 : Built an **image captioning model** in **Python** using **CNNs** for image feature extraction and **RNNs (LSTMs)** for generating text captions, leveraging **PyTorch** to construct and train this multimodal architecture.
- **Drake vs Kendrick** 🔄 : Conducted a comparison of Kendrick Lamar and Drake's song popularity in **R** using **k-Nearest Neighbors (kNN)** classification and **exploratory data analysis** to analyze play distributions and distinct artist features.
- **GameStop Stock** 🔄 : Modeled GameStop (GME) stock volatility in **R** with **linear regression**, **influence metrics**, and **robust regression** (using **gradient descent**) to address outliers and highlight influential trading days.
- **Edmonton Oilers** 🔄 : Assessed Edmonton Oilers' goal-scoring patterns in **R** with **Poisson-based MLE** and **MDE** for goal distribution, applying the **Horvitz-Thompson estimator** to evaluate **bias**, **variance**, and **MSE** in home vs. away goal averages.
- **MartMetrics**: Currently working on a sales forecasting model using the Walmart Kaggle dataset and **XGBoost** in **Python**, leveraging **time-series decomposition**, **hyperparameter tuning**, and **cross-validation** to predict sales trends.

Skills & Interests

Languages: Python, SQL, R, GraphQL, VBA, HTML/CSS

Infra: Git, GCP, Linux, Docker, Kubernetes, Gitlab

Frameworks: NumPy, Pandas, Scikit-learn, Matplotlib, XGBoost, Pytorch

Databases: MySQL, PostgreSQL, SQLite, Snowflake

Interests: Gym Goer, F1 Racing Enthusiast, Fan of Liverpool Football Club and the Toronto Raptors