

Ajun Jo

4th Year Statistics student in University of Waterloo

Waterloo, Canada | 905 651 5768 | joajune03@gmail.com | [Snapshots.Me](https://snapshots.me) | [GitHub](https://github.com) | [LinkedIn](https://www.linkedin.com/in/ajunjo)

EDUCATION

University of Waterloo | Bachelor of Mathematics Honours Statistics Major | Computing Minor | Sep.2021 – Apr. 2026

Relevant Coursework : Probability, Statistics, Software Development, Databases, Statistical Learning

Extracurriculars : Semi-competitive Basketball league, Soccer Intramural (Placed first in league)

EXPERIENCE

Data Science Research Assistant (Incoming) | University of Waterloo | Waterloo, Ontario, Canada

- Tech-Stacks: Python, PySpark, Bash, Deep Learning, Time Series Forecasting
- Developing predictive models discovering the relationships between variables

Software QA Engineer | People Inc. | New York, United States

Sep. 2025 – Present

- Built a pull-request CD pipeline using Jenkins to generate deployable build artifacts, enabling Squadron-based preview environments for UI, API and database testing prior to merge.
- Executed database API tests for a FastAPI service generator, validating PGVector and Milvus integrations across health, metadata, text, and vector search endpoints.
- Identified and fixed a silent failure that allowed 400/500 errors to pass without warning messages, **reducing UI/UX errors by upto 80%**.
- Validated PostgreSQL schema by testing connectivity, schema creation and data retrieval flows via service APIs.
- Streamlined the deployment pipeline by executing release cutting procedures, ensuring all approved pull requests were captured in the correct version artifact before production rollout.

Data Analyst | Scotiabank | Toronto, Canada

Jan. 2025 – Apr. 2025

- Developed a time study tool that automated the collection of branch activity data from emails, **consolidated 100k+** records into a centralized dataset, and analyzed them to **identify the top 30 activities**, presenting the results through interactive dashboards in **Power BI and Tableau**.
- Utilized **MySQL** aggregation queries to **extract KPIs** like frequency of branch tasks, time per activity and productivity by regions
- Improved team performance by **up to 25%** by leading cross-functional team projects such as customizing time study tool for other internal teams enabling them to capture and assess daily productivity efficiency and patterns
- Led multiple cross-functional time study projects, managing scope, timelines, and stakeholder coordination to deliver all project phases on schedule and support data-driven operational decisions

LLM Data Engineer | Vosyn | Etobicoke, Canada

May 2024 – Aug. 2024

- Reduced **data processing time by 30%** by automating ETL pipelines, such as collecting multilingual datasets for TTS models, cleaning its dataset, transforming audial data into texts for model input suitability.

PROJECTS

California Real Estate Data Analytics | [Github](https://github.com)

Jan. 2024

Project Summary: Python data analysis which provides visualizations and insights of real estate information in California.

- Utilized NumPy, Pandas, Matplotlib and Seaborn to analyze with **over 20,000 real estate listings**
- Data cleaned then visualized variable relationships using a correlation table to uncover key drivers of house prices

Sephora Product Intelligence & Machine Learning Analysis | [Github](https://github.com)

Aug. 2025 – Sep. 2025

Project Summary: Performed ETL process on product data, transforming raw datasets into BI dashboards and ML models

- Built a **PostgreSQL database (8k+ products)** to support scalable analysis and report a Power BI dashboard
- Wrote SQL analyses to identify ROI drivers, hidden gems, and category-level pricing and promotion trends
- Trained XGBoost models to classify high-potential products and predict ratings, likes, and launch popularity

SKILLS

Data Science: Python, MySQL, R, Excel, Power BI, PySpark

Software Engineering: Bash, Git, CI/CD, Jira