

# Joseph Lim

☎ 647-929-6726 | ✉ j67lim@uwaterloo.ca | 📄 jhlim0921 | 🌐 josephlim0921 | 🌐 josephhyunjinlim.com

## TECHNICAL SKILLS

---

**Programming Languages:** Python (Pandas, NumPy, PySpark Matplotlib, Seaborn), SQL (MySQL, PostgreSQL, MS SQL)

**Machine/Deep Learning:** Scikit-Learn, XGBoost, TensorFlow, Keras, PyTorch

**Data Engineering/Analytics:** Azure Databricks, BigQuery, Prefect, Airflow, MLflow, UiPath, Tableau, Power BI, Looker

## EXPERIENCE

---

### Data Scientist | Gore Mutual Insurance Company

May 2024 – Present | Toronto, ON

- Collaborating with the Data & Analytics team to develop and refine ML models, addressing key pricing and underwriting challenges to enhance business solutions
- Developing Python scripts for data collection from multiple sources and conducting exploratory data analysis to identify trends that guide strategic decisions

### Data Scientist | PepsiCo

Sept 2023 – Dec 2023 | Mississauga, ON

- Spearheaded a national store segmentation project for Quaker, employing PCA and K-Means Clustering on demographics data to effectively cluster 3000+ Canadian stores, identifying opportunities to optimize retail operations across 7 product categories
- Commercialized a ML project with senior data scientists by building 10+ interactive Power BI dashboards linked to model outputs in Delta Lake, providing real-time shopper insights to business stakeholders
- Conducted comprehensive data analysis on over 1 billion rows of POS sales and demographics data using SQL, Pandas, and PySpark, driving strategic execution recommendations for the field team in preparation for a new Frito-Lay product launch
- Developed Ridge Regression models to forecast the sales performance of non-existing store-product combinations across 4 competitor product lines, thereby generating a prioritized list of 1000+ high-potential stores to target for competitive market entry
- Identified and resolved over 5 critical data errors across the store master, sales and demographics datasets, significantly improving their quality for future ML projects

### Associate Producer | Zynga

Jan 2023 – Apr 2023 | Toronto, ON

- Developed SQL queries in MS SQL and used Python libraries (Pandas, NumPy) to streamline data collection and analysis on team KPIs, increasing the efficiency of processes by more than 80%
- Built interactive reports and dashboards in Looker to equip 10+ cross-functional agile teams with valuable insights for data-driven improvements to their sprint performances
- Analyzed project data using SQL by generating relevant statistics on resource availabilities and project durations to create project roadmaps, resulting in a 50% increase in project/OKR tracking efficiency for teams
- Collaborated with the FinOps team to create 10+ interactive Tableau dashboards that visualize resource management trends and establish resource forecasting, enabling data-driven decisions in resource allocation across the organization

### Junior Product Manager | Front Rush, NCSA, Zcruit

May 2022 – Aug 2022 | Chicago, IL

- Utilized Heap to analyze customer data across 3 products by defining KPIs and usage metrics that generated insights on over 10,000 daily users, ultimately guiding future product decisions and feature enhancements
- Conducted a customer retention analysis for the Zcruit portal, identifying critical improvement areas that informed the development of a strategic product plan to enhance user satisfaction
- Leveraged insights from analyzing 300+ feedback tickets submitted by NCSA customers to identify key trends and plan for future OKRs, fostering a more targeted approach to NCSA's product development

## PROJECTS

---

### NBA Game Winner Predictor | Pandas, NumPy, Scikit-Learn, XGBoost

Mar 2024 | Toronto, ON

- Led a comprehensive ML project to predict NBA game outcomes by overseeing and executing all stages from data collection to modeling, achieving a test accuracy of 70%

### K-pop Song Recommender | Pandas, NumPy, Scikit-Learn, Spotipy

Dec 2023 | Toronto, ON

- Constructed a data pipeline using a Spotify API to extract and transform features of songs from multiple K-pop artists, thereby creating a well-structured dataset for efficient downstream analysis
- Developed a content-based recommendation system for K-pop songs, utilizing cosine similarity to calculate similarity scores and suggest top song recommendations to users

## EDUCATION

---

### University of Waterloo | B.A.Sc. Systems Design Engineering

Sept 2020 – Present | Waterloo, ON

- Courses: Data Structures and Algorithms, Probability and Statistics, Machine Learning, Applied Linear Algebra
- Cumulative GPA: 85.8%

### DeepLearning.AI | Machine Learning Specialization

Mar 2023 – Apr 2023 | Toronto, ON

- Courses: Supervised Machine Learning – Regression and Classification, Advanced Learning Algorithms, Unsupervised Learning – Recommenders – Reinforcement Learning
- Projects: Linear/Logistic Regression, Neural Networks, Decision Trees, K-Means Clustering, Anomaly Detection