

# 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 & Visualization:** Azure Databricks, BigQuery, dbt, Prefect, Apache Airflow, MLflow, UiPath, Docker, Tableau, Power BI, Looker

## EXPERIENCE

### Data Scientist | PepsiCo

Sept 2023 – Present | *Mississauga, ON*

- Initiated a nationwide store segmentation project, employing PCA and K-Means Clustering to effectively categorize 3000+ Canadian stores into distinct groups, allowing the organization to optimize retail operations tailored to similar shopper demographics
- Spearheaded an end-to-end supervised machine learning project to predict the most valuable stores for various competitor products, developing a pipeline that handles data extraction, preparation, modeling and visualization of insights
- Enhanced model performance by experimenting with machine learning algorithms such as XGBoost and Ridge/Lasso Regression, and fine-tuned hyperparameters, achieving a 20% increase in  $R^2$  score
- Conducted in-depth exploratory data analysis on over 1 billion rows of competitors' sales and demographics data using SQL, Pandas, and PySpark, driving strategic marketing and execution recommendations for the field sales team
- Contributed to the development and commercialization of 5+ analytical reports in Power BI, providing teams with crucial insights into the sales performance, demographics and shopper behaviours of distinct clusters

### Data Analyst/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 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*

- Performed data visualization and analysis across 3 products using Heap by defining KPIs and usage metrics, generating insights on 10,000+ daily users to aid in product decisions and drive improvements in features/functionality
- Successfully led a feature improvement project for Zcruit by effectively translating a product plan into user stories for engineers and designers, resulting in a 25% increase in UX measures
- 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

### Junior Product Manager | Front Rush

Sept 2021 – Dec 2021 | *Chicago, IL*

- Co-led the development of 2 products at Front Rush by working closely with cross-functional agile teams to define, plan, and prioritize product requirements, resulting in successful product launches
- Composed 50+ user stories in JIRA related to the development and implementation of new features for Front Rush's products, resulting in increased usage by 9,500+ teams and 30,000+ coaches
- Conducted 20+ interviews with coaches to gain in-depth knowledge of athletic department processes and identify specific needs across various schools, helping the product team to optimize user experience and improve product functionality

## PROJECTS

### ASA DataFest Hackathon | Python (Pandas, Matplotlib, Seaborn), SQL

Apr 2023 | *Waterloo, ON*

- Analyzed patterns/trends in a raw dataset related to an online Q&A platform using SQL and Python to help devise effective business strategies and suggest data-driven improvements in lacking areas

### NBA Data Analysis | Python (PyTorch, Pandas, NumPy), SQL, Looker

Feb 2023 – Apr 2023 | *Toronto, ON*

- Built an LSTM model using PyTorch that predicts future NBA season averages using historical averages from the past 40 years
- Developed SQL queries in MS SQL to explore player statistics extracted from an NBA API and created a player dashboard on Looker, highlighting key performance metrics that can be used for game predictions

## EDUCATION

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

Sept 2020 – Present | *Waterloo, ON*

- Relevant Coursework: Data Structures & Algorithms, Probability and Statistics, Machine Learning, Applied Linear Algebra
- Cumulative GPA: 3.90/4.00

### 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