

YONG SHIN

Cambridge, MA

(713) 876-5192 ◊ shin.yongwhan23@gmail.com ◊ [Portfolio](#) ◊ [LinkedIn](#) ◊ [GitHub](#)

EDUCATION

Master of Statistics, Rice University

Houston, TX (2021-2022)

GPA: 3.70/4.00

Coursework: Machine Learning, Inference, Time Series and Forecasting, Probabilistic Algorithms and Data Structures

BA in Computer Science, Rice University

Houston, TX (2018-2021)

GPA: 3.93/4.00, magna cum laude

Louis J. Walsh Engineering Scholarship, President's Honor Roll

Coursework: Object-Oriented Programming, Program Design, Parallel Programming, Operating Systems, Compilers

SKILLS

Languages/Technologies: Python, R, SQL, Apache Airflow, Pandas, NumPy, Scikit-Learn, PyTorch, Bash & High-Performance Computing, JavaScript, React, Flask, HTML & CSS, Java, C++, C, Git

EXPERIENCE

Harvard University

Data Scientist

Boston, MA (2023-)

- Developing and launching data analysis pipelines on **High-Performance Computing** clusters for physician-scientists and creating visualizations of dataset summary statistics using **Python**, **R**, and **Bash**.

Meta

Data Scientist, Instagram Trust - Security

Menlo Park, CA (2022-2023)

- Built out Security's first **ML**-based detection system for compromised accounts by **hierarchically clustering** account events capturing **20K+** compromised users with **>80%** precision.
- Led **Apache Airflow** pipeline implementation to populate relational database with tables containing users' security checkpoint statuses and conversion histories.
- Conducted an **A/B test** resulting in the launch of a security checkpoint which applies to **1B+** monthly active users.
- Created and maintained a **BI dashboard** powered by **SQL** and **Python** to measure work progress in account security space for cross-functional collaborating teams across Instagram Trust.

Facebook

Data Scientist, Oculus Content - Beat Saber

Menlo Park, CA (2021)

- Graphed in-app purchasing data from log pipelines using **Python** and **SQL** to present on speed of decay of advertising-based purchases after music pack release.
- Led the **A/B test** for an alternative music pack recommendation system based on decay of advertising-based purchases for engineering team as one of two data scientists leading virtual reality game Beat Saber, resulting in launch.
- Investigated differences in in-app purchasing habits between demographics and prototyped a **BI dashboard** displaying insights on in-app purchasing statistics to deliver to stakeholders.
- Powered **2 Apache Airflow** pipelines to generate tables linking music pack purchasing histories to user demographic information for product analytics team members.

PROJECTS

LSH Search Engine. Developed a search engine in **Python** using **20M+** web queries from **650k+** users using **Locality-Sensitive Hashing** algorithms to measure similarity between multiple URLs.

Eatsy. Designed and released web app built on **Flask** backed by a **logistic regression** model that generates restaurant recommendations based on friends' Yelp ratings. **3rd Place** at Texas A&M University Datathon.