

# HONGCHAO HU

Mobile : (224) 600-6852 | Email : Jason.Hu168@outlook.com | Location : Worcester, MA 01604  
LinkedIn: [www.linkedin.com/in/hohu](https://www.linkedin.com/in/hohu) | Website: <https://hongchaohu.github.io/>

## SKILLS

- **Programming Tools:** Python (Pandas, Torch, SageMath, Seaborn, Scikit-learn), Java, JavaScript
- **Data Management Tools:** MySQL, SQL Server, MapReduce infrastructure, Spark, HBase, NoSQL Databases
- **Data Science & Modeling:** Exploratory Data Analysis (EDA), Feature Engineering, Clustering, Predictive Modeling, A/B Testing, Statistical Analysis, Model Development
- **Data Visualization:** Tableau, Power BI, Interactive Dashboards, d3.js, Datawrapper

## EDUCATION

### Worcester Polytechnic Institute

Aug 2025 - May 2027

- *Master of Science in Data Science*

### The University of California San Diego

Sep 2020 - Jun 2024

- *Bachelor of Science in Mathematics and Computer Science | Minor in Business Economics*

## WORK EXPERIENCE

### Tanbii

Oct 2025 - Present

#### CRM Data Analyst Intern | Part-Time | Nashville, TN | Remote

- Processed Terminal and in-game behavioral data using **Statbot** and **MySQL** to uncover retention patterns; applied predictive analytics and clustering to segment user cohorts, supporting lifecycle strategies.
- Translated data analytical insights into weekly and sprint new activity release for optimizing onboarding flow, enhanced in-game sustainability engagement, user segmentation, and in-app conversion rates.
- Built interactive **Tableau** and **Jupyter** dashboards to visualize user adoption metrics with development decisions.

### Mega IT

Jun 2024 - Jan 2025

#### Data Engineer | Full-Time | San Jose, CA

- Performed exploration data analysis using **Python** (Pandas, SageMath) to identify product usage inefficiencies with 94% accuracy and optimize feature rollout timing, decreased developing-wide operational costs by 23%.
- Conducted **ETL** Process, extracted and modeled telemetry data from 5 disparate vendor sources.
- Through **Aris**, monitored feature-level tracking metrics for platform usage; authored actionable roadmap insights.

### Mobalytics

May 2023 - Sep 2023

#### Data Scientist Intern | Santa Clara, CA

- Built a client attrition prediction model using **Random Forest** in **JupyterLab**, improving quarterly retention by 12 basis points through targeted intervention strategies.
- Led geospatial analysis and developed forecasting models with **Scikit-learn** and **SciPy** to project market trends and optimize monetization; presented insights via **Power BI**, reducing erroneous deliveries by 12%.
- Applied **Python** and **Excel** to segment users and evaluate **A/B test** performance on feature launches and platform updates, informing product iteration decisions.

## PROJECT

### ACL Injury Risk Prediction | WPI | Dr. Bahman Moraffah's group

Sep 2025 - Dec 2025

- Developed ML pipeline in **Python** (**scikit-learn**, **pandas**) evaluating 7 models with **GridSearchCV**; addressed class imbalance via **SMOTE** and oversampling across 16 preprocessing configurations integrating **PCA**.
- Validated using **Stratified K-Fold** with threshold optimization across **ROC-AUC**, **F1**, and recall to balance sensitivity and specificity for clinical decision support.

### Microservice Data Base API Development | UC San Diego

Dec 2022 - May 2024

- Authored and designed microservice architecture, streamlined data retrieval and reduce third-party dependency.
- Built RESTful services and pipelines using **MongoDB** and **ARIS Express**, standardized interaction flows.

## CERTIFICATIONS

- Microsoft Power BI Data Analyst Professional Certificate

