

# Yuhao (Bill) Liu

(617)710-8647 | New Jersey | yl5308@nyu.edu | <https://github.com/LYH36> | [linkedin.com/in/yuhao-bill-liu-088519177](https://www.linkedin.com/in/yuhao-bill-liu-088519177)

## EDUCATION

### New York University

*Master of Science in Data Science*

New York, NY

Sep 2022 – May 2024

- **Coursework:** Optimization and Computational Linear Algebra, Programming for Data Science

### Boston College

*Undergraduate Degree, B.A. in Economics, B.A. in Mathematics / GPA: 3.8/4.0*

Chestnut Hill, MA

Aug 2018 – May 2022

- **Honors:** Dean's List First Honors
- **Coursework:** Statistics, Probability Theory, Linear Algebra, Algorithm, Machine Learning, Data Science, Optimization

## SKILLS

**Tools:** Python, Microsoft Excel, STATA, SPSS, SQL, Java, Tableau, JMP, R, BigQuery, HTML, MongoDB, JavaScript

**Techniques:** Data Mining, Machine Learning, Statistics, Data Analytics, Data Visualization, Mathematical Modeling.

## PROFESSIONAL EXPERIENCE

### P&G

*Data Science Analyst – Intern*

Beijing, China

Jul 2021 – Aug 2021

- Gathered ~**20k** records of diaper packaging data, captured **15** key business metrics based on consumer tests in Japan, cleansed & transformed the data in **Excel**, summarized the insights for decision-making by baby care department.
- Applied linear regression in **JMP** to quantify the impact of compactness on length of package with **R-squared of 0.93**.
- Envisioned & created the **1<sup>st</sup>** globally used business model for diaper packages, maximized diaper's softness and quantity, improved consumer perception scores by **20%** and adopted as the standard for global package design.

### MaiMai

*Data Analyst – Intern (User Growth Team)*

Beijing, China

Feb 2021 – Apr 2021

- Cleansed & processed **100k+** user transactions using **Google BigQuery** and conducted exploratory **Big Data** analysis, implemented user's retention rate regression model using **Scikit-learn in Python**, achieved an **RMSE of 0.067**.
- Designed daily refreshed **Power BI** Dashboard with 8 key metrics & integrated automatic alerts for real-time decision-making.
- Analyzed spurts in user growth to identify root causes, developed user personas applied on user activity & demographic features and provided data-driven recommendations for marketing & product decisions.

### Gerson Lehrman Group

*Financial and Data Analyst – Intern*

Beijing, China

Sep 2020 – Jan 2021

- Extracted relevant information from **10M+** historical transactions using **SQL** scripts, analyzed the data with summary statistics & pivot charts in **Excel**, identified anomalies & patterns in the data & presented insights to manager.
- Developed & monitored data quality in the database with **Python Pandas**, visualized data quality report with **Seaborn** plots & achieved **98%** data quality compliance.
- Defined **12+** employee performance metrics and computation logic, transitioned to a data-driven appraisal system & reduced Chinese section's labor costs by **10%**.

### KPMG

*Business Analytics Consulting Virtual Internship*

Online

Jul 2020 – Jul 2020

- Identified **15+** key business metrics for Sprocket Central, presented quantitative analysis in **Tableau** dashboard.
- Assessed data quality for dataset of **20,000** customers with **30+** fields, generated summary statistics, imputed missing data and treated outliers, conducted data deduplication & normalization, improved overall data quality by **24%**.
- Applied **k-means clustering** for customer segmentation & targeted high-value customers based on 8 customer demographic & transaction attributes, optimized current transactional data model.

## PROJECTS

### Boston College COVID-19 Prediction

*Project leader*

Chestnut Hill, MA

Apr 2020 – May 2020

- Conducted COVID-19 prediction using **Python Pandas**, visualized prediction model with **Matplotlib plots**.
- Applied **TensorFlow** and **Scikit-learn in Python** to predict the peak death rate date with **R-squared of 0.87**.