

# Sean Yang

212 West 91<sup>st</sup> Street Apt.311, New York, New York, 10024  
(332)248-8199| [yangxuehan7@gmail.com](mailto:yangxuehan7@gmail.com)  
Linkedin: [www.linkedin.com/in/SeanYangDS](https://www.linkedin.com/in/SeanYangDS)

## EDUCATION

### Columbia University Mailman School of Public Health

New York

**Master of Science (MS)**, Biostatistics

Sept 2021-Present

#### Relevant Courses:

Probability, Data Science, Statistical Inference, Biostatistical Methods, Principles of Epidemiology

### Nankai University

Tianjin

**Bachelor of Economics**

Sept 2016-Jun 2020

**Relevant Courses:** Multivariable Calculus, Data Structures and Algorithms, Linear Algebra

## PROFESSIONAL EXPERIENCE

### Amazon

Beijing

Business Intelligence Intern | Global Store Selection Team

Sept 2019-Feb 2020

- Constructed 30+ SQL query templates with Oracle to automatically extracted business data every week, thus highly improved business intelligence.
- Independently conducted Black Friday troubleshooting project, filtering unbuyable items, figuring out the mistaken limitations from legal affairs, and making products with high GMV pass through the pipeline. Made 300+ items buyable on Global Store which sold 50,000+ units during black Friday.
- Used Web scraping to combine products' ASIN with their legal limitations into one Excel chart with Python, for the reference of colleagues in Marketing Team, thus turned a week-long manual work into several minutes.
- Built and maintained a well-structured weekly report with Excel to explain reasons for items' inflow and outflow and other sales fluctuation causes.

### Ping An Property & Casualty Insurance Co of China Ltd

Shenzhen

Actuarial Intern | Product Pricing Group

Jun 2019-Sept 2019

- Set premium rates for 14 insurance products and wrote 14 actuarial reports based on actuarial methods, all of which got approval from the China Banking Regulatory Commission into the insurance market.
- Constructed Excel templates with VBA to display monthly income statements and wrote 5 Quality Analysis reports based on those data to analyze the performance of different insurance types for the general manager's meeting.

### China Merchants Bank

Tianjin

Project Intern | Wealth Management Group

Sept 2017-Jan 2018

- Performed natural language preprocessing work like noise cleaning, data labeling, and rules extraction to help construct core corpus for machine learning. Extracted 4K pieces of keywords with an error rate under 1%.
- Helped solve the test set underfitting problem and improved training accuracy through manual screening.
- Through 400 hours of high-intensive work, accelerated the progress of our program and helped the team made a breakthrough from 0 to 1 in life insurance AI visualization system in China.

## RESEARCH EXPERIENCE

### Anti-fraud Methods on Big Data in Health Insurance

Tianjin

Research Assistant

Jun 2020-Mar 2021

- Summarized 4 methods of detecting insurance fraudulence in physicians with a combination of big data and machine learning algorithms like SVM, Logistic Regression, and Random Forest.
- Wrote a research report on two experiments in detecting fraudulent physicians based on three datasets from The Centers for Medicare and Medicaid Services (CMS) and explored whether the methods could be used in the Chinese insurance industry.

### National College Student Innovation & Entrepreneurship Program

Tianjin

Research Assistant

Mar 2018-Mar 2020

- Selected 13K sample companies and expected features from the Wind database as variables.
- Constructed six regression models based on modified Two Tiers Stochastic Frontier model with SPSS to quantify the influence of information asymmetry in corporate acquisitions. And used variance decomposition to explain model interpretation ability.

## SKILLS&INTEREST

- **Programming and Software:** SQL (Expert), R (Advanced), Python (Advanced), C++ (Hands-on)
- **Certificate:** Linear Regression and Modeling, Inferential Statistics, Introduction to Data Science in Python