

# Zheng Bai

+86-13021191066 | [zeheng@hgc.jp](mailto:zeheng@hgc.jp) | [celestial-bai.github.io](https://github.com/celestial-bai)

## EDUCATION

---

|  |                                       |
|--|---------------------------------------|
| <b>The University of Tokyo</b><br><i>Master of Engineering in Computer Science, Information Science and Technology</i> | Tokyo, Japan<br>Apr 2021 – Apr 2023   |
| <b>The University of Tokyo</b><br><i>Research Student in Institute of Medical Science</i>                              | Tokyo, Japan<br>Oct 2020 – Apr 2021   |
| <b>Renmin University of China</b><br><i>Bachelor of Science in Mathematical Statistics, School of Statistics</i>       | Beijing, China<br>Sep 2016 – Jul 2020 |

## RESEARCH EXPERIENCE

---

|  |                                       |
|--|---------------------------------------|
| <b>Test and improvements for VIBRANT</b><br><i>The University of Tokyo</i> <ul style="list-style-type: none"><li>Research assistant, Advisor:<br/>Prof. Seiya Imoto, Professor of Human Genome Center, Institute of Medical Science, The University of Tokyo<br/>Prof. Yaozhong Zhang, Associate Professor of Human Genome Center, Institute of Medical Science, The University of Tokyo</li><li>Made deep understanding of whole pipeline of VIBRANT by using Docker and Ubuntu</li><li>Made familiar with popular tools or models for recognizing the category of genome, such as VirSorter, VirFinder, MARVEL, VIBRANT</li><li>Ran some example to test VIBRANT by using supercomputer</li><li>Made some assumptions try to improve VIBRANT, like using XGBoost or XLNet</li></ul>  | Oct 2020 – Present<br>Tokyo, Japan    |
| <b>The Relationship between Air Quality and Epidemic Diseases in Taiwan</b><br><i>Renmin University of China</i> <ul style="list-style-type: none"><li>Research assistant, Advisor:<br/>Prof. Yang Li, Vice Dean of School of Statistics, Renmin University of China<br/>Prof. Ben-Chang Shia, Vice President, Fu Jen Catholic University<br/>Prof. Feifei Wang, Lecturer of School of Statistics, Renmin University of China</li><li>Got the each-hour data of the climate of Taiwan from 2010 to 2018 by web crawler using Python and joined the data of the climate and air quality of Taiwan according to the same watch points using R</li><li>Got the everyday and every-month data of each district of Taiwan from 2010 to 2018 calculated by the each-hour data of the climate of Taiwan based on Kriging using R</li><li>Found outliers by the rules we had made before, change them to K-means value by close-time data</li><li>Made heat maps of the everyday and every-month data of each district of Taiwan, density plots of data of the climate of each district of Taiwan, and correlation scatter plots between the climate and air quality</li><li>Found the relationship between air quality and respiratory diseases using disease mapping and variable coefficient regression</li><li>Taught the new members about data visualization make some improvements of the codes and make great progress of the coding ability; We could finish the whole task of data cleaning and data visualization in only one day</li></ul> | Nov 2018 – Jul 2020<br>Beijing, China |
| <b>Report: Traditional Chinese Medicine Big Data 2019</b><br><i>Peking University</i> <ul style="list-style-type: none"><li>Research assistant, Supervisor:<br/>Prof. Xiaohua Zhou, Director of Department of Biostatistics, Peking University, and Director of TCM Big Data Center<br/>Dr. Xinting Huang, Assistant Director of TCM Big Data Center</li><li>Got the data of the programs about Traditional Chinese Medicine Big Data carried on by Chinese companies by web crawler using Python</li><li>Improved and corrected them, and made further understanding about the situation of Traditional Chinese Medicine Big Data by surveys using email</li><li>Finished the report “Traditional Chinese Medical Science Big Data 2019: Chapter Companies</li></ul>  | Jul 2019 – Nov 2019<br>Beijing, China |
| <b>Active Healthy Living and the Aging Population in China</b><br><i>Renmin University of China</i>  | Jun 2019 – Jul 2019<br>Beijing, China |

- Research assistant, Supervisor: Prof. Xiaojun Wang, Dean of School of Statistics, Renmin University of China  
Prof. Feifei Wang, Lecturer of School of Statistics, Renmin University of China
- Got the data about people's feeling towards the Active Healthy Living and the Aging Population in China by Distributed Web Crawler using Python
- Proposed an improved emotion analysis model which was more suitable for our project based on LDA Model

## INTERNSHIPS

---

### **Meituan Co. LTD** | *Data Product Manager Intern*

Oct 2020 – Feb 2021

- Standardized the naming and definition of metrics
- Provided data sheets for the company leaders
- Completed iteration of data products
- Guaranteed the timeliness of the output of daily newspapers and standardized the procedures about the process of guarantee

### **Xiaomi Technology Co. LTD** | *Data Analyst Intern*

Nov 2019 – Feb 2020

- To be in charge of the data about the Xiaomi Overseas App Store (GetApps)
- To save those data by Hive, whose codes are similar to SQL
- To make the reports on the strategies this can increase our income from Mi App Store
- I also make some attempts on other approaches, like LSTM, and I also improved some codes on Mi Data Factory System

### **China Railway Construction Co. LTD** | *Software Engineering Intern*

Jan 2019 - Feb 2019

- Got familiar with the functions of the APP (China Railway Construction Financial Services)
- Helped them to find the errors about the APP by running it on several kinds of smart phones

## OVERSEAS EXCHANGE

---

### **Summer Session 2018**

Jul 2018 – Aug 2018

*University of California, Berkeley*

*Berkeley, the United States*

- Did great work during the exchange program and got high scores: Probability and Statistics in Biology and Public Health (A+) Introduction to Neuroscience (A+)

## SKILLS

---

**Programming Skills and Software:** C(Mastered), Python(Mastered), Visual Basic(Mastered), SparkSQL(Proficient), R(Proficient), HiveSQL(Proficient), SPSS(Proficient), Stata(Proficient), Office(Proficient)

**Developer Tools:** Docker, Singularity, VS Code, Notepad++, Xshell, Xftp, Visual Studio, PyCharm, Rstudio, Ubuntu

**Operation System:** Linux, Windows **Languages:** Mandarin Chinese(Native), English (Fluent), Japanese(Basic)