

## Zhihao (Johnson) Du

+1 (510) 833-4417

zhihao617@berkeley.edu

### Education

<b>University of California, Berkeley</b>	Current GPA: 3.6/4.0	<i>Fall 2019 - Now</i>
<ul style="list-style-type: none"><li>Statistics, BA – Modern Statistical Prediction and Machine Learning; Linear Modelling: Theory and Applications; Concepts in Computing with Data.</li><li>Computer Science, BA – Discrete Mathematics and Probability Theory; Data Structures and Programming Methodology; Designing Information Devices and Systems II.</li></ul>		
<b>YKPao School, Shanghai</b>	Final Grade: 40(+2)/42(+3)	<i>Fall 2007 – Summer 2019</i>
<ul style="list-style-type: none"><li>International Baccalaureate bilingual diploma – Physics HL, Mathematics HL</li></ul>		

### Personal Experience

<b>HOYO Lab</b>	Co-Founder	<i>March 2021 – Aug 2021</i>
<ul style="list-style-type: none"><li>Self-initiated, interest-driven, data-centered project on the RPG game Genshin Impact:</li><li>Subproject I “Damage Calculator”: Predicted the game’s underlying damage mechanism using linear regression with stochastic regressors and feature mapping (16 features); collected data through 500+ in-game simulations; initialized the model with naïve feature map functions and finalized the feature map function through cross validation; achieved an R-squared approximately 1; displayed the final model through web app using the shinyapp package in R<sup>[1]</sup>; utilized the model and published tutorial articles that received 400,000 views and 50,000+ likes.</li><li>Subproject II “Celestial Database”: Lead a group of 3; aimed to build an in-game character power rank based on character usage in combat challenge; crawled 1000+ user profile in game associated mobile app using Python requests; overcame protection barrier using virtual headers; decoded JSON files to extract data of interest; stored and updated data in MySQL database (every 3 weeks); showcased final power rank on website<sup>[2]</sup>.</li></ul>		
<b>Beijing PM2.5/PM10 Prediction</b>	Project Leader	<i>Nov 2021 – Dec 2021</i>
<ul style="list-style-type: none"><li>Lead and coordinated group of 3; aimed to predict the real-time measurement of PM2.5/PM10 in Beijing; imported data (30,000 rows, 12 features) from Beijing Environmental Monitoring Center between 2013 and 2017;</li><li>We performed robust EDA through visualizing univariate distribution, bivariate relation, and interaction significance; employed AIC with backward stepwise selection for feature engineering; arrived with two models that were carefully diagnosed for model assumption and outliers.</li><li>We tested our models and constructed prediction intervals; achieved final RMSE of 0.6; presented the result in a report with detailed description and diagrams.</li></ul>		
<b>DataCVG</b>	Intern	<i>May 2021 – Aug 2021</i>
<ul style="list-style-type: none"><li>I performed extract-transform-load (ETL) using Kettle and MySQL; aimed to combine outdated spreadsheets from two data sources onto an updated data source (100+ spreadsheets).</li><li>We constructed the architecture of destination spreadsheets by considering various primary keys along with necessary metadata columns; reengineered incompatible columns through grouping rows, pivoting columns, or modifying column datatypes; loaded the final results through merging transformed spreadsheets by matching primary keys; debugged the final data through a series of rigorous diagnostics for any architectural failures.</li><li>The ETL project helped our client, Fosun Pharma, to upgrade to their newest, higher performing database system.</li></ul>		

### Skills

**High level programming languages:** R, C, Python, Java; **Web development related language:** HTML, CSS, JavaScript; **Assembly language:** RISC-V; **Query language:** MySQL, SQL Server

### Award and Honors

**Anna Sohmen Pao Award:** Awarded for being the top 3 students in class of 2019 (high school)

### Hobbies

**Piano** – Performed on high school graduation ceremony; **Choir** – Participated in ISCMS XII and ISCMS XIII (International Schools Choral Music Society); **Violin** – High school orchestra 2<sup>nd</sup> violinist; **Soccer** – high school varsity team left defensive midfielder.

[1]: [https://zhihao617.shinyapps.io/genshin\\_basic\\_damage\\_calculator](https://zhihao617.shinyapps.io/genshin_basic_damage_calculator)

[2]: <https://www.youngmoe.com>