

# Yu Zhu

Email: drewyu1204@berkeley.edu

## EDUCATION

---

### East China Normal University (ECNU)

Bachelor of Science in Statistics

GPA: 3.67/4.0

Average Grade: 87.9/100

**Honors:** Outstanding Student Scholarship (2020-2021 & 2021-2022)

Shanghai, China

Sep 2020 – Jan 2023

### University of California, Berkeley

Exchange Student

GPA: 3.9/4.0

Berkeley, CA

Jan 2023 – Present

**Relevant Courses:** Concepts of Statistics, Modern Statistical Prediction and Machine Learning, Nonparametric and Robust Statistics

## ACADEMIC PROJECTS

---

### Exploratory Analysis of Soccer Player Dataset

Team Leader, Multivariate Statistical Analysis Course Project

Dec 2022

- Obtained data on 17,588 active soccer players. Ranked players by ratings and recorded their details, including name, nationality, position, height, weight, and various skill scores.
- Performed data visualization for skill score distributions based on player positions (forward, midfield, defense).
- Utilized principal component analysis (PCA) and factor analysis to extract key player abilities.
- Conducted canonical correlation analysis to find the relationship between player's physical qualities and abilities.
- Developed a multivariate logistic regression model to predict player positions based on player's abilities.

### Research on Factors Influencing National Average Lifespan

Individual Regression Analysis Course Project

Nov 2022

- Collected data on national average lifespans and related factors for 193 countries from WHO and UN databases.
- Carried out regression analysis to study the relationships between health, economic, and social factors, and identify key factors influencing national average lifespans.
- Methods include: data preprocessing, establishment of the multiple linear regression model, significance testing of equations, constrained least square methods, confidence intervals, model diagnostics: residual analysis, heteroscedasticity treatment, and multicollinearity tests.

### Optimization of MIXUE Ice Cream & Tea Business Plan from the Perspective of Milk Tea Consumers

Team Leader, National College Market Survey and Analysis Competition

Nov 2021 – Apr 2022

- Collected data through offline and online methods, distributing a total of 2492 questionnaires. Compiled and processed the dataset.
- Performed data visualization to compare MIXUE with other milk tea brands. Analyzed consumer evaluations and preferences for various factors of milk tea brands.
- Built decision tree models to analyze consumer profiles.
- Employed SPSS and R to establish linear regression models to analyze the main influencing factors of consumer loyalty and build logistic regression models to investigate factors affecting consumers' milk tea brand preferences.
- Designed a web crawler to collect text data from Meituan and conducted in-depth interviews to analyze consumer feedback.

### Designing a Game Using Python's Object-Oriented Approach

Individual Python Course Project

Dec 2021

- Created a character class with attributes: name, health, attack power, level, experience, and equipment. Implemented methods for equipping, removing equipment, leveling up, and viewing character stats.

- Developed a map class to store global game information, utilizing a grid-style map to store objects like characters and equipment. Implemented character attacks and movement.

## **Research on China's Foreign Trade Data**

Team Leader, Python Course Project

May – Jun 2021

- Obtained China's foreign trade data from January 2018 to December 2019 for four major industries (mechanical and electronic industry, high-tech industry, agricultural product industry, clothing and apparel industry) from the Ministry of Commerce's data center and trade reports.
- Cleaned and preprocessed data using Python, analyzed data trends through data visualization, and created a visual menu using the tkinter library.

## **INTERNSHIP**

**Capgemini**

Shanghai, China

**On-site Intern: Public Opinion Monitoring for New Energy Vehicles**

Jun – Jul 2023

- **Defining Project Direction and Metrics:** Collaborated closely with the team to identify key evaluation factors, including user satisfaction, complaints, and expectations. Developed methodologies for data collection and quantification, establishing a strong foundation for subsequent project phases.
- **Web Scraping and Data Enrichment:** Utilized Python Selenium for web scraping, successfully extracting commentaries from major new energy vehicle websites. Systematically stored gathered data in MongoDB and MySQL databases, creating a foundational resource for later analysis and modeling.
- **Addressing Natural Language Processing Challenges:** Mastered the Named Entity Recognition (NER) model, enabling the extraction of vital information such as vehicle models and functional components from comment texts; Conducted an in-depth study of sentiment analysis methodologies, particularly the Target-grounded aspect-based sentiment analysis (TG-ABSA) model. Applied sentiment analysis to new energy vehicle comment texts using deep learning neural networks.
- **Exploring Advanced Techniques:** Explored the deployment of ChatGLM-6B and P-Tuning fine-tuning techniques, enhancing keyword extraction and sentiment analysis capabilities.
- **Real-time Integration and BI Display:** Integrated the database with trained models for real-time data processing and analysis. Implemented real-time Business Intelligence (BI) displays aligned with the project's goals.