

Skills

Data Analysis, Visualization and Machine Learning

I am experienced in various data tasks leveraging **R**, **Python** and **SQL**, including data cleaning, feature engineering, statistical modeling, data visualization and interactive reports. Featured projects include (with subjects)

- *Logistic regression, multiple imputation, and resampling.* A **case study** on the most up-to-date and comprehensive Titanic data, exploring survival factors with regression.
- *Text analytics.* A **case study** classifying novels by Brontë Sisters. A similar multinomial classifier is developed to analyze **Animal Crossing Reviews**.
- *SQL for data query and R for Data Visualization:* A **case study** analyzing international debt statistics.
- My own utility R **package**

Front-End Development

I am familiar with **React** (Redux, Next, Gatsby) and **Vue** (Vuex). I also have knowledge in **TypeScript** and **GrapQL**.

- **Data Science Shelf:** A book website for data science built with Vue. Covers authentication, real-time updates and markdown support.
- **Github Profile Viewer:** A React app for exploring Github user statistics.

Work Experience (Internship)

Data Analyst

51Talk, Department of Big Data and Business Intelligence

September 2020 – Now

- At a leading online education company, I contributed to establishing an integrated teacher-evaluation system. With R and Python, I conducted extensive exploratory analysis and research on the influence of hardware setup, speaking speed, facial expressions and time management on conversion rate and customer review. I am involved throughout the project in problem definition, data cleaning, feature engineering, regression modeling and the development of end-to-end web applications.
- I am also involved in an ongoing project to optimize our adaptive placement testing system, based on item-response theory and deep knowledge tracing.

Business Analyst

Analysys, Department of Data Analysis and General User Success

June 2020 – September 2020

- Improved the web and app analytics service in aspects of data visualization tools and user interface
- Supported users across industries by designing tracking events and analyzing user behavioural data
- Translated data into interactive reports and dashboards written in R Markdown and JavaScript, which received immense positive feedback from clients

Data Mining Engineer

DiDi Chuxing, International Division

March 2019 – August 2019

- Contributed to building a fast, scalable and extensible AB testing environment with R and C++
- Conducted overseas geo-marketing research targeting South America and Australia, with density-based clustering, logistic regression, and correspondence analysis

Education

Communication University of China

Beijing - China

Data Journalism, 3.8/4.0

September 2017 – June 2021

- Top 5% with excellent performance in Calculus, Linear Algebra, Probability, Statistics and various courses in programming.

University of California, Los Angeles

Online

Summer Session, **Stats100C: Linear Models**, 4.0/4.0

July 2020 – September 2020

- Top 20% in a proof-based course on the theory of regression analysis. Topics include Gauss Markov theorem, general aspects of maximum likelihood estimation, partial regression, constrained estimation, generalized least squares, and model selection.