

# Fan Fei

## Data Scientist

Chicago, IL 60611

[feifanfiona4\\_t23@indeedemail.com](mailto:feifanfiona4_t23@indeedemail.com)

+1 614 286 8664

MS in Analytics (Data Science) Candidate at University of Chicago looking for Data Scientist/Data Analyst summer internship opportunity.

Expected Graduation: Dec. 2022

Willing to relocate: Anywhere

Sponsorship required to work in the US

## Work Experience

---

### Graduate Researcher

The University of Chicago - Chicago, IL

January 2022 to March 2022

1. Developed and upgraded the fraud comment filtering system by implementing the supervised machine learning algorithm via PySpark on GCP Dataproc and Hadoop RCC to prevent customers from being misled by deceptive high ratings from deficient merchants.
2. Optimized the review rating prediction model and achieved 94% F1-score and 92% AUC using logistic regression, random forest, and linear SVC for restaurant reviews found outside Yelp to enlarge the review database.
3. Performed sentiment analysis in Natural Language Processing to cluster

### Graduate Researcher

The University of Chicago - Chicago, IL

September 2021 to December 2021

1. Designed and developed interactive maps and tables using Tableau to help customers quickly assess hotel and Airbnb information across NY state, including star ratings, price, and dining.
2. Collected, organized, and designed relational databases on MySQL, GCP, and MongoDB based on over 400,000 lines of raw hospitality-industry data.
3. Further guided consumer decision-making with tailored recommendations for cost-effective substitutes.
4. Project also aimed to help growing hotel and Airbnb developers by building regression models using R Studio that study factors impacting popularity and price to determine where to expand.

### Teaching Assistant

The Ohio State University - Columbus, OH

July 2020 to May 2021

Co-hosted R Studio and Python workshops with lecturers by answering questions that students have and helping them to debug their codes.

## **Undergraduate Researcher**

The Ohio State University - Columbus, OH

January 2020 to May 2021

1. Wrote and published thesis paper about how Data Visualization helped with overcoming COVID-19.
2. Performed Data Analysis work such as data cleaning using Python, R, and Tableau to make visualizations for infection rates and mortality rates for COVID and published on public website.
3. Analyzed and compared global data with data from Ohio State along with factors of racial background on infection and mortality rates.

## **Data Analyst**

CTR Market Research - Beijing, China

July 2019 to September 2019

1. Analyzed and visualized raw data set to give clients a better view of consumers' habits for making business decisions.
2. Implemented business interactions during meetings with company branches.
3. Managed to learn a new software (PV) developed by CTR in three days to clean the raw data.

## **Data Analyst Intern**

China International Capital Corporation (Investment Banking) - Beijing, China

March 2019 to July 2019

1. Used data to assess and evaluate investment products' profitability trends.
2. Managed to extract data and plot graphs of product markets using Excel and report to clients.
3. Consolidated data and wrote reports for the Wealth Management team for weekly analysis.
4. Wrote and revised investment product statements and contracts to support further negotiation with clients.

## **Banquet Server**

The Blackwell Inn and Conference Center - Columbus, OH

June 2018 to September 2018

- Prepared and served food under high-pressure environment.
- Set up and cleaned up banquet room with strict organized rules under fast pace.
- Resolved communication problems from customers and managers.

## **Education**

---

### **Master's degree in Data Science**

University of Chicago - Chicago, IL

September 2021 to Present

### **Bachelor's in Data Analytics**

The Ohio State University - Columbus, OH

August 2017 to May 2021

### **High school diploma**

Mother Seton Regional High School - Clark, NJ

August 2015 to May 2017

## Skills

---

- Mandarin
- MySQL (3 years)
- Data Analysis (5 years)
- Python (3 years)
- Data Science (2 years)
- Tableau (3 years)
- R (4 years)
- Open Refine (1 year)
- Google Cloud Platform (1 year)
- Data mining
- Data Visualization
- Big data
- Hadoop
- Apache Hive
- MongoDB

## Languages

---

- Mandarin - Expert
- Korean - Beginner

## Links

---

<https://github.com/feifanfiona>

<https://public.tableau.com/app/profile/fan5584>

<https://www.linkedin.com/in/fan-fei-a44062192/>

## Certifications and Licenses

---

### **2nd Place of Student Research Forum**

February 2021 to Present

## Publications

---

### **SPREAD AND DYNAMICS OF COVID-19: A DATA VISUALIZATION STUDY**

<https://kb.osu.edu/handle/1811/92484>

May 2021

A pandemic of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), named as coronavirus disease 2019 (COVID-19) by the World Health Organization (WHO), emerged in Wuhan, China and it is currently underway resulting in worldwide severe morbidity and mortality. Since the beginning of the pandemic, visualizations have assisted us to understand the unpredictable nature of this virus by providing detailed information based on data and variables on the ground. During the pandemic, visualizations can help us to analyze features such as the infection rate within different groups of people, thus we could interpret the reasons behind the high infection rate. The objective of our study is to overview some of the state-of-art visualization techniques to illustrate the dynamics of the COVID-19 spread for Ohio. Although there are some potential variables, such as age, that are known to be associated with COVID-19, it remains unclear if the affection and death rates for COVID-19 differ significantly for some other factors. Therefore, we are going to analyze COVID-19 data from Ohio and the USA to discover such factors statistically.