

# JIANING (JULIA) CHEN

✉ jchen801@usc.edu · 🔗 LinkedIn · ☎ (765)337-0250

## EXPERIENCE

---

### Hoover Institution, Stanford University

May. 2022–Aug. 2022

**Data Science Intern**, supervised by Dr. Oliver Giesecke: og25@stanford.edu

- Applied web scraping and gathered data from public websites using Selenium WebDriver.
- Built pipeline using optical character recognition (OCR) techniques, PaddleOCR and Tesseract, to extract state and local pension data from annual comprehensive financial reports.
- Created user interface and deployed Dash App for public pension systems in the U.S. 🏠

### USC Keck School of Medicine - Epidemic Modeling for Los Angeles County 🏥

Jan. 2021–Dec. 2021

**Research Assistant**, supervised by Dr. Abigail Horn: hornabig@usc.edu

- Cleaned daily streaming data, such as filling missing data and performed data manipulation.
- Conducted epidemic modeling via a dynamic stochastic model; implemented Approximate Bayesian Computation for parameter estimation with R language; extended model into 342 small counties in Los Angeles.

## PROJECTS

---

### Capstone Project- Behavior Prediction and Safety Control of Clients 🏢

Aug. 2021–Dec. 2021

Kiana Analytics Company

- Applied Lean Six Sigma project management skill to enhance a safety working environment for clients.
- Analyzed large data set and performed data cleaning; generated clients' working schedule patterns using time series model and detected the people gathering area using DBSCAN clustering method.
- Visualized final outcome using Tableau and Matplotlib package and recorded a demonstration video.
- Organized and led meeting as a liaison between stakeholder and team to identify strategic decisions.

### Supervised Learning Projects - Classification Model for Hospital to Prescribe Treatment

Feb. 2021

- Extracted insights by visualizing relationships among predictive features; utilized hypothesis testing and backward selection to define variables; implemented data transformation to amplify significance of statistic.
- Tested multiple classification models such as Random Forest, Logistic Regression, and Gradient Boosting Machine; performed hyper-parameter tuning with GridSearch cross-validation to minimize RMSE.

### Data Management - Track COVID-19 in California Hospitals

Jan. 2021–May. 2021

- Retrieved Hospital data from API save to MySQL; fetched streaming COVID-19 data and store in MongoDB.
- Cleaned web crawling data, transformed and mapped the data on a daily basis, improved ETL processes.
- Designed React and Streamlit web layout and optimized site design to increase user engagement.

## EDUCATION

---

### University of Southern California, Los Angeles, CA

Jan. 2021–Dec. 2022

Master of Science in *Applied Data Science*, GPA: 3.9/4.0.

Coursework: Machine Learning, Database Management, Natural Language Processing, Data Mining

Available to work full-time with CPT or Pre-completion OPT starting on Aug. 22, 2022

### University of California, Santa Barbara (UCSB), CA

Sep. 2019–Dec. 2020

Bachelor of Science in *Statistics and Data Science*, GPA: 3.6/4.0.

Coursework: Bayesian Analysis, Regression Analysis, Probability, Survival Analysis, Optimization, Time Series

## SKILLS

---

- Programming Language: Python, SQL, R, SAS, C, Java, LaTeX
- Database and Cloud: MySQL, MongoDB, AWS EC2, Firebase, Hadoop Spark, Hadoop MapReduce
- Framework: Altair, StatsModels, Matplotlib, Scikit-Learn, Numpy, Pandas, TensorFlow, PyTorch
- BI Tools: MS Excel, Tableau Desktop, Tableau Server