

Haocheng Zhang

7557 Main Street, Houston, Texas | (281)-460-8812 | zh28@rice.edu | GitHub: <https://github.com/jeffreyzhanghc>

EDUCATION

Rice University

Houston, TX

Bachelor of Arts in *Statistics*, Minor in *Data-Science*

Sep 2020 - Jun 2024

- **Cumulative GPA:** 3.7/4.0; **Major GPA:** 3.9/4.0

SKILLS

- **Programming:** Java Script, Solidity, Python, R, SQL, MATLAB, PySpark, Pandas, Scikit-learn, TypeScript
- **Machine Learning:** Linear/Logistic Regression, KNN, SVM, NLP, DNN, QDA, LDA, Naïve Bayes, Survival Analysis
- **Certificates:** IBM Data Analysis and Visualization Certificate, Professional Data Analyst Certificate

INTERSHIP EXPERIENCE

Meson (Web3 Company)

Houston, TX

Data Analyst/Marketing Intern

Oct 2022 - Present

- Created real-time dashboard using on-chain end-to-end observability platform **Sentio.xyz** to monitor the transaction data of Meson.fi to troubleshoot any anomalous event, as well as transaction volume of potential competitors such as synapse bridge
- Obtained self-updating stablecoin transaction data of 4 cross-chain bridges by interacting with smart contracts using **TypeScript**
- Promoted the cooperation with 7+ venture capital firms such as Coinbase VC, and 12+ project teams such as Particle Network

Allied Millennial Partners

Houston, TX

Data Analyst Intern

Feb 2022 - Apr 2022

- Fitted Autoregressive model using data from Citigroup company to predict its future performance, achieving overall accuracy around **93%** after validating it using train-test split method
- Constructed stock selection model with stationary check, statistical testing and data prediction functions, and built ARIMA model in **R** to analyze stock of Citigroup and its competitors to give investment suggestions
- Downloaded Transaction Data of Crude Oil from 2000-2021 to analyze the trend of its stock by simulating the performance of nominal value and daily/monthly return in **Pandas**; results **matching 95%** of curve on Yahoo Finance

Bluebonnet Data

Houston, TX

Data Analyst Intern

Jan 2022 - Apr 2022

- Initiated a **Python** script that helps load multiple CSV files into data frames, aggregate them into one frame using index, fill missing entries with appropriate values, and export the results to CSV files, which **saved 80%** of processing time
- Visualized the vaccination distribution of Philadelphia based on word map and bar-chart by using ggplots in **R** and matplotlib & pandas library in **Python**; final visualization adopted by **every organization canvassing team**
- Imported voting data containing a hundred thousand of entries into **SQL** database to create a script joining raw data tables on desired columns based on voting condition and generated CSV files for report

RESEARCH & PROJECT EXPERIENCE

Texas Children Hospital & Rice D2K Lab

Houston, TX

Machine Learning Scientist

Jan 2023 – Present

- Cleaned patient data from 2012-2023 that has over 1.2 million rows by dropping duplicates, fill in NAs, group by hour/day/month time units, and added 30+ features for model usage; visualized plots showing obvious seasonality as desired by sponsors
- Established ARIMA & Generalized Linear models to predict the patient arrival count to emergency room in next 2-18 hours
- (In Progress) Utilize deep-learning model and Cox-regression to predict ER room waiting time given campus & staffing resource

Perception of Collision & Human-Computer Interaction Lab

Houston, TX

Research Assistant

Jul 2022 - Present

- Used **Python Pandas** to improve the process of participant data comparison, **reducing time costs from 4 days to 1 hour**
- Fixed **MATLAB** code for an equipment that generates the vibration for the time-to-contact experiment using Psychtoolbox, and the equipment is then **used 100% of the time** for every relative experiment
- Helped check python code that used in unity script to build virtual reality scene for collision conception experiment
- Applied QDA and Naïve Bayes classification models to help classify the performance of participants

Stack Overflow Survey Analysis

Houston, TX

Data Analyst

Oct 2021 - Nov 2021

- Implemented **PostgreSQL** script to display the schema of the table, distribution of data, and outliers
- Uploaded a file containing over 70000 responses on programming language usage in **Python Pandas**, cleared the N/A values, grouped the data by programming languages/databases usage frequency, and normalized data
- Visualized data as pie-chart, line-chart, bubble-chart etc. and built dashboard using **IBM Cognos and Excel**