

Junjie (Caspar) Chen

<http://caspar0-0.github.io> | 1950 Elkhorn Court, San Mateo, CA, 94403 | (207)286-4242 | jc5067@columbia.edu

EDUCATION

Columbia University, GPA: 3.43/4.0 New York
M.A. in Statistics, Data Science Track Sep. 2018 - Feb. 2020
Coursework: Data Mining, Databases, Machine Learning, Time Series Analysis, Nonparametric Methods, etc.
University of Wisconsin – Madison, GPA: 3.54/4.0 Madison, WI
B.S. in Applied Mathematics, Emphasis on Actuarial Science, Certificate in Business Sep. 2014 - May. 2018

WORKING EXPERIENCE

Oigetit Fake News Filter Sunnyvale, CA
Data Analyst Intern Mar. 2020 - Present

- Operate Google Analytics for Firebase to track key metrics such as DAU, Screen_View and Notification_Open for mobile app, as well as interpret user behavior from data visualization which resulted in efficient news posting and user growth.
- Collaborate with QA teams to determine adjustment and new features in app that helped to achieve debugging and retention.
- Retrieve log data and conduct ad hoc data analysis from various social media platforms that targeted the correct end users.

China Taiping Insurance Co. Ltd. Shanghai
Product & Marketing Analyst Intern July. 2018 - Aug. 2018

- Scraped data on insurance products of peer companies with insurance apps, cleaned data with R for report in use of product development for actuaries and sales analysis.
- Visualized data with Excel and PowerPoint and presented to product management team with key findings.

Orient Securities Research Center Shanghai
Data Analyst Intern Aug. 2017 - Aug. 2017

- Analyzed financial reports and summarized in Excel with formulas such as Sharpe ratio and betas to determine the portfolio.
- Conducted research on most up-to-date Chinese government new policies on development trend to provide evaluation of profitability for various companies.

Shanghai Well C.P.A Partnership Shanghai
Auditing Department Intern Jun. 2016 - Aug. 2016

- Assisted external audit work of 15 companies' social insurance by verifying salary tables and writing reports with accounting software, and using Excel (VLOOKUP, pivot table) to collect and examine the data.
- Contributed audit work for verifying 277 companies' industry and commerce information disclosure by contacting companies with emails and phone calls under Industrial and Commercial Bureau.

SELECTED PROJECTS

Machine Learning Project Columbia University, Nov. - Dec. 2019

- Implemented the deep learning neural style transfer and Adam optimization algorithm to generate blended images in **Python**.
- Tuned the parameters in image classification model VGG-19, and weight ratios in loss function for different art performance.

Database Application Project Columbia University, Feb. - Apr. 2019

- Developed **ER diagrams** with six entities, front-end web application using **Python** Flask for users to query modern Japanese Animation with custom features and return expected results.
- Applied **PostgreSQL** to extract, transform, and import data into schemas for users to manipulate data.
- Extracted ~9GB data and created database using **Google Cloud Instance** and **Bit Bucket** with shell scripts.

Nonparametric Methods Project Columbia University, Mar. - May. 2019

- Designed **hypothesis testing** exploring relationships between five qualitative features and six quantitative risk factors of Coronary Heart Disease in **R**.
- Analyzed Framingham Heart Study (long-term prospective study of the etiology with 5,209 observations and 16 variables).
- Explored effects of **Survival Analysis** (K-M Curve, Log-Rank test, Cox Regression), Fisher's Exact Test, Correlation Test, **Wilcoxon Rank Sum Test**, Kruskal-Wallis Test, Two-sample t-Test, F-test and Logistic Regression.
- Demonstrated that smokers have a higher probability of getting CHD than non-smokers, and many other findings.

Applied Data Science Project Columbia University, Nov. - Dec. 2018

- Conducted hypothesis testing and studied the most indicative determinants, interaction terms, and functional expressions for multiple linear regression in **R**.
- Manipulated 25,000 observations for males between age of 18 and 70 who are full time workers from a government dataset.
- Explored effects of logistic transformation, bootstrap resampling, stepwise selections, statistical diagnostics (EDA & ANOVA), and model validation (MSE, DFBETA & MSPR).
- Demonstrated that African American males have statistically different wages compared to Caucasian males or all other males.

SUMMARY OF QUALIFICATIONS

- Programming Languages: R, SQL, Python, Tableau, Java, HTML & CSS**
- Extra:** Passed Exam FM/2, Exam P/1; Fulfilled VEE Requirements for Economics, Corporate Finance, & Applied Statistical Methods; Certificates: DataCamp: Cleaning Data in Python, Web Scraping in Python