

Junjie (Caspar) Chen

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

EDUCATION

Columbia University in the City of New York, GPA: 3.43/4.0

New York

M.A. in Statistics, Data Science Track

Sep. 2018 - Dec. 2019

Coursework: Data Mining, Databases, Machine Learning, Time Series Analysis, Nonparametric Methods

University of Wisconsin – Madison, GPA: 3.54/4.0

Madison, WI

B.S. in Applied Math, Emphasis on Actuarial Science, Certificate in Business

Sep. 2014 - May. 2018

WORKING EXPERIENCE

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.
- Visualized data with Excel and PowerPoint and reported to product management team.

Orient Securities Research Center

Shanghai

Data Analyst Intern

Aug. 2017 - Aug. 2017

- Assisted the cross-functional team on analyzing collected data from listed companies and auditing firms.
- Involved in analyzing data by using Excel spreadsheet and suggested on financial formulas to apply into various cases.
- Conducted research on most up-to-date Chinese government new policies on development trend to provide evaluation for various companies.
- Summarized and evaluated from various aspects to provide company a comprehensive research report.

Shanghai Well C.P.A Partnership

Shanghai

Auditing Department Intern

Jun. 2016 - Aug. 2016

- Assisted external audit work of 15 companies' social insurance by making salary tables and writing reports with accounting software and using Excel (VLOOKUP, pivot table) to collect and examine 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

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.

Linear Regression Project

University of Wisconsin-Madison, Mar. - May. 2018

- Created and concluded the final determinant factors of automobile paid losses for collision for a UCI dataset in 1987.
- Performed multiple linear regression model with ANOVA, residual analysis, and VIF as methods in **R**.

SUMMARY OF QUALIFICATIONS

- Programming Languages: R, SQL, Python, Tableau, Java, HTML & CSS, Tensorflow**
- 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