

# Phoebe Cai

yuc4003@gmail.com | 10 River Road, New York | (347)400-9068 | <https://yuc4002.github.io/> | Visa: OPT

## EDUCATION

---

- Cornell University** | New York, NY Dec 2019  
*Master of Science in Biostatistics and Data Science* GPA: 3.88
- Coursework: Biostatistics(R), Epidemiology, Statistical Programming with SAS, Data Management (SQL)
- Hobart and William Smith Colleges** | Geneva, NY June 2017  
*Bachelor of Arts in Mathematics and Economics* GPA: 3.68
- Coursework: Linear Algebra, Econometrics(SAS), Intermediate Programming(Java), Math Statistics(MATLAB)
  - Economics TA for Principles of Econ., Macro Econ & Micro, Statistics (SAS)
  - Mathematics TA for 5 classes for a year, participate and offer suggestions in professors' weekly meeting

## PROFESSIONAL EXPERIENCE

---

- Financial Analyst Intern** 06/2018-08/2018  
*Chengdu Hongqi E-commercial* Chengdu, China
- Collaborated with other teams across the organization to better use and understanding of data
  - Developed KPI and metrics like new customer growth rate, transaction growth rate and retention rate
  - Conducted exploratory data analysis in Python to analyze business trends based on processed transaction data
  - Applied Tableau to build daily, weekly, monthly dashboard and provide insights behind data
  - Developed a predictive model using R, successfully improved the prediction of the fraudulent activates in the remaining months in 2018 by 33.3% accuracy, and decreased the company's monthly loss by 19%
- Data Analyst Intern** 08/2015-12/2015  
*Esperity* Brussels, Belgium
- Redesigned and implemented the database structure in SQL by adding functions to manage the duplicate information in the original users' information database
  - Applied SQL to identify and classify users' behavioral patterns
  - Worked with cross-functional teams, like product managers and engineers
  - Built daily, weekly, monthly dashboard to provide insight about the churn rate, customer order frequency, order volume, total order amount
  - Provide business insights behind data and present to stakeholders

## SELECTED PROJECT

---

- Improve Account Open Rate**
- Helped bank client design a new algorithm to target customers with higher potential to apply for credit card optimize marketing campaign performance
  - Used the scikit-learn in Python to apply Hierarchical clustering to cluster customers into 5 groups
  - Designed A/B testing to compare the result and present the insight to clients
  - Improved the account open rate by 35% within two months
- Prediction on Health Insurance**
- Performed data manipulation with R to process healthcare related insurance data (Place of residence, Wealth, Mother's education, Father's education, Age of the mother & etc.)
  - Built classification models (Logistic Regression and Random Forest) to predict the probability of low-weight infant
  - Optimized model performance using cross-validation
  - The Random Forest suggests age is the most important factor and the accuracy of prediction (90%) is much better than the Logistic Regression (50%)

## SKILLS

---

- Technical Skills: Data Mining, Analytics, Data Analysis, SaaS in cloud, Machine Learning
- Big Data Engineering: SQL, Python, Apache Hadoop, Excel, Data Visualization, SPSS, Cognos
- Soft Skills: Self-learning, Problem-solver, Interpersonal Skill, Communication, Project Management, Risk management, Attention to Detail, Team-oriented, Innovation