

# Heqing Sun

919-279-0193 | [heq.sun@gmail.com](mailto:heq.sun@gmail.com) | [www.linkedin.com/in/heqing-sun](http://www.linkedin.com/in/heqing-sun) | [hq-sun.github.io](http://hq-sun.github.io)

---

## OVERVIEW

Analytically minded professional with a blend of business acumen, entrepreneurial drive, strong analytical knowledge, and excellent communication skills. Key abilities include the following:

- Motivated practitioner with a solid understanding of statistical concepts, predictive modeling, and machine learning and 2+ years' experience using Python, SQL to solve large-scale data analysis problems
- Effective communicator that articulate complex concepts for audiences with diverse backgrounds
- Passionate team player and individual contributor with robust organizational and interpersonal skills

---

## PROFESSIONAL EXPERIENCE

**Data Scientist Intern — Cape Fear Collective, Wilmington, NC** May 2020—Present

- Sourcing, cleaning, and transforming publicly available demographic, population, and economic data to create over 100 county-level socioeconomic indices
- Creating cumulative performance metrics summarizing said indices for over 3000 counties
- Utilizing several machine learning methodologies in R to determine county-level peer groups
- Summarizing insights in a comprehensive dashboard to inform future work on a regional scorecard

**Data Scientist | Practicum Project — Bank of America, Charlotte, NC** August 2019—May 2020

- Partnered with Consumer Anti-Money Laundering team to identify anomalous financial centers
- Evaluated approximately 30GB of data drawn from financial centers and customer interactions to find unknown criteria and latent features indicative of the anomalous financial center activity
- Developed machine learning algorithms in Python to classify financial centers across varying levels of risk and identified important features that drive the anomaly
- Built visualizations in Tableau to present key insights and recommendations to executives and strategy stakeholders

**Data Analyst — Wind Information Co., Ltd, Nanjing, China** June 2018—May 2019

- Sped up FDA data collecting and wrangling process 20-fold by developing an automation tool using Python and SQL
- Incorporated a new Indian generic drug module with 3000+ generic drugs and 100k brands into the existing platform by contributing to module design, data collection, and data wrangling steps
- Administered pharmaceutical-related Private Equity and Venture Capital (PEVC), Mergers and Acquisitions (M&A) modules and conducted both descriptive and inferential statistics
- Boosted product sales by 25% via conducting market research, customer interviews and competitive product analysis to understand the customer needs

---

## SELECTED PROJECTS

**Marketing Campaign Project | (R)** June 2020

- Performed Boruta search to reduce the data dimensionality by 20 times
- Applied several sampling techniques (Under-Sampling, Over-Sampling, ROSE, SMOTE) with logistic regression models to select records to be included in a future campaign and achieved an f1 score 0.12
- Designed A/B testing to understand performance differences attributable to variation in the marketing treatment

**Dental Practice Patterns Anomaly Detection | (Python)** February 2020

- Analyzed 38K+ dentists within the network and performed k-means clustering, bootstrap Mann-Whitney rank test to identify around 1% dentists that might require further investigation since their practice patterns appear to be overly prescriptive towards the more extreme solutions

**Sales Analytics Project | (Python)** January 2020

- Performed Exploratory Data Analysis (EDA) to identify the top advisors with 10K+ transaction data
- Applied XGBoost and Random Forest to explore the relationship between touchpoint and trades
- Used k-means algorithm to determine 3 different types of advisors that help the asset management firm better understand their customer base

---

## SKILLS

- *Software:* Python, SQL, R, SAS, Tableau, GitHub
- *Analytical:* Data Wrangling, Feature Engineering, Dimension Reduction, Data Visualization, Regression, Statistical Modeling, Predictive Analytics, Machine Learning, Clustering, Anomaly Detection, Relational Database Management and Querying, Design of Experiments, Hypothesis Testing
- *Certifications:* SAS Certified Specialist: Base Programming Using SAS 9.4  
SAS Certified Statistical Business Analyst Using SAS 9  
Tableau Desktop Specialist Certification  
Google Analytics Individual Qualification

---

## EDUCATION

**M.S in Analytics** — Institute for Advanced Analytics, NC State University, Raleigh, NC | GPA: 3.67/4.0 May 2020

**M.S in Chemistry** — University of North Carolina at Chapel Hill, NC May 2018

**B.S in Chemistry, outstanding graduate** — Nanjing University, Nanjing, China June 2016