Xi DU
+1 (984) 215-7171; ccjulia@email.unc.edu; www.linkedin.com/in/66JuliaDu; 301 South Rd, Chapel Hill, NC, 27514

EDUCATION

University of North Carolina at Chapel Hill (UNC) *Cumulative GPA: 3.803/4.00*	Chapel Hill, USA
B.S. in ***Statistics and Analytics*** (major GPA: 3.866/4.00); B.A. in ***Mathematics*** (major GPA: 3.733/4.00)	*Aug.2020 - May.2024*

Minor in **Data Science**

- Dean's honor list (Spring2022/ Fall2022/ Spring2023)
- Top 25% in the Hypatia Contest and Euclid Contest, Canadian Open Mathematics Challenge(X2) — *Apr.2018/ Apr.2019*

Carolina Analytics and Data Science Club, *Core Member*	*Jan.2022 - Present*
UNC Mathematics Department, *Undergraduate Learning Assistant for Dr. Kovalsky in Math347 Linear Algebra*	*Aug. - Dec.2022*
UNCFACSS(NGO), *Chief Planner of Event and Project Department*	*Aug.2020 - Dec.2021*
Online Courses: *Introduction to SQL* (Datacamp), *Statistics with Python Specialization* (Coursera)	*Sept.2023 - Dec.2023*

Technical Proficiencies: Python, R, Java, MATLAB, JavaScript, HTML, SQL, Julia, SPSS, Excel

ACADEMIC RESEARCH

Project Member| **Stock Market Time Series Forecasting Project**, *Dr. Zhengwu Zhang*| ***Python, JavaScript*** — *May.2023 - Present*

- Collaboratively scraped and quantified financial news data (Oct. 2021 to Dec. 2023) from 7 main source databases, and applied both traditional deep learning and Large Language Model-based models based on historical stock data from indices like SP500, DJIA, and NDQ, to predict stock market trends across different time intervals, with two main strategies: baseline model (without news and microblogging data) and media data-based model (with such features)

Student Research Assistant| **AMPLab**, *Dr. Alana Campbell*| ***Python, MNE-Python, MATLAB*** — *Aug.2022 - Present*

- Using MNE-Python to preprocess and analyze EEG (electroencephalogram) data related to indicators of infants with autism and performing data visualizations to understand the patients' brain activities using MATLAB

Project Member| **Image Classification Machine Learning Project**| ***Python, TensorFlow, Pytorch, R*** — *Jan. - May.2023*

- Applied transfer learning techniques to classify and predict emotions in dogs based on a dataset of 19,921 dog images, with about 5,500 images for each emotional state (angry, happy, relaxed, and sad), and modified the selected pre-trained model using classifier modification, multi-task learning, fine-tuning, and unsupervised domain adaptation, resulting in substantial improvements in the testing accuracy of the models (54.5% on average VS 63.8%)

Group Member| **2023 ASA DataFest™ at Duke University**| ***R*** — *Mar.2023*

- Assisted with a three-member team in analyzing a dataset comprising the actual conversations between clients and lawyers and determining the significance level of each client request using two criteria: the category of the case and the sentiment of the text
- Developed a prioritization system to reorder requests based on their urgency levels, helping American Bar Association (ABA) lawyers effectively prioritize the most pressing cases

Project Leader| **Honey Production Analysis Project**| ***R*** — *Jun. - Aug.2022*

- Performed extensive data cleaning & merging on large-scale NASS datasets from 2015 to 2020 scraped from Kaggle.com and visualized the data of honey yield per colony and percent loss on a state-by-state map in the U.S to identify key factors influencing the percent loss, with the goal of guiding beekeepers in effectively increasing honey production
- Standardized each predictor to compare their coefficients across models, developed 12 different models to analyze the influencing factors selection, and identified the most appropriate one based on the MAE analysis of each model

PROFESSIONAL EXPERIENCE

Zhejiang Communications Industry Service Co., Ltd.| ***Excel, Business Intelligence***	Hangzhou, China
Financial Data Analyst Intern	*Jul. - Aug.2023*

- Led a cross-functional data project by working collaboratively with the IT and Finance departments to analyze 2023 Q2 financial data from 30 subsidiaries of China Communications Industry Service, including key metrics such as financial document volumes, returned volume percentages, document types & dates, processing time, etc., to evaluate their performance on document quality and processing efficiency
- Collaborated with accountants for in-depth analysis of return rates and types to underlying factors contributing to return trends and formulated targeted recommendations tailored to specific document categories and corporate entities
- Co-authored the 2023 Q2 financial report, incorporating refined data analysis and visualizations into the operational framework of subsidiary companies to provide actionable insights for operations improvement in subsequent quarters

Amazon New York| ***Python, Pandas, NumPy, Matplotlib, Data Modeling***	Remote
Part-time Data Analyst Assistant	*Jul. - Aug.2022*

- Designed experimental processes for A/B testing under the guidance of a Senior Software Engineer from Amazon, analyzed experimental results using Python and proposed specific business proposals based on the findings
- Applied data cleaning, merging, and visualization with Python to transform data into actionable insights, which led to a 0.5% increase in conversion rates after implementing new electrical plugs in Amazon's user interface for products

Zhejiang Transportation Planning and Design Institute| ***Excel***	Hangzhou, China
OD Survey/Route Design Assistant	*Jun.2020 - Jun.2021*

- Conducted origin-destination (OD) survey by collecting raw data from drivers passing through specific OD survey sites on the Keqiao-Zhuji Expressway in Zhejiang Province, China
- Processed and organized the collected data in Excel, utilized the growth rate method to estimate the origin-destination patterns for characteristic years, and assigned the derived OD data to generate new origin-destination plots for further analysis and interpretation