Yangruijue Ma (Anna)

• ym2813@cumc.columbia.edu• +1 9294036608

Education
University of California, Santa Barbara (UCSB)

Bachelor of Science, Financial Math and Statistics   June 2020
Bachelor of Arts, Philosophy   June 2020
Columbia University Mailman School of Public Health

Master of Science, Biostatistics   Expected May 2023

Skills
- *Technical:* Advanced in Python, R, SQL, and SAS.
- Experienced in Microsoft Office Word, PowerPoint, Access, and Excel.
- *Language*: Bilingual Mandarin and English, spoken and written

Work Experience
Yunnan Wangxing Big Data Technology Co., Ltd.   Kunming Yunnan, China
Data Analyst Intern   Oct 2020 - Jan 2021
- Conducted literature searches in academic papers on social network analysis
- Assisted in building a social network analysis model with both Python and R
- Meet with the client, record their requirements, and communicate with the project group.
- Visualized the result of the model with graphs, tables, and interactive plots
- Produced a written report and PowerPoint for project presentation

Projects
Access and My SQL Database Project   May 2022
- Built an Access database with parent, child, and look up table for a hypothetical HIV drug study
- Wrote SQL queries with joins and window functions to create analyzable dataset
- Designed data entry forms, main menu, and reports for summarized group data in Access
- Recreated the tables in MySQL and linked with Access using ODBC connector
- Wrote SQL queries in MySQL to create views, temporary tables, and CTEs to analyze data.

Bayesian & Genialized Linear Model of Hurricane Trajectory Project   May 2022
Team Member
- Cleaned the track data of 702 hurricanes recorded and conducted monthly summary analysis
- Visualized the trajectory of hurricanes in each month on the North American map
- Ran the Markov Chain Monte Carlo Algorithm to generate posterior distribution of model parameters
- Built a Poisson regression model with offset to analysis hurricane characteristics and its damage
- Used LASSO cross validation to select variables for the Poisson Model.

Multiple Linear Regression Project – Crime rate investigation   October 2021
- Used the *httr* R package to interact with APIs and download datasets from webpages.
- Performed descriptive summary statistics using the County Demographic Information dataset
- Fit multiple models, including the full model, models selected by the automatic approaches and models criterion-based approach using R
- Validated the model using training and testing data by cross validation
- Analyzed the performance of the models by comparing their errors and predicative ability
- Produced a written report for the project