Jiaqi Li

100 Haven Avenue, Tower 3 17F, New York, NY 10032
(201) 887-9945 | jl5025@columbia.edu | https://ljq312.github.io/

EDUCATION

Columbia University, Mailman School of Public Health, New York, NY *May 2019*
Master of Science, Biostatistics, GPA: 3.6
Relevant courses: Biostatistical Methods, Data Science, Statistical Learning & Data Mining
University of Washington, Seattle, WA *May 2014*
Bachelor of Arts, Psychology; Bachelor of Arts, Economics, GPA: 3.5

EXPERIENCE

Synyi Artificial Intelligence *Shanghai, China*
RWE Biostatistics Internship (EHR Data) *May 2018 - August 2018*

- Contributed in projects sponsored by pharmaceutical companies, including Merck, AstraZeneca, Eli Lilly, GlaxoSmithKline. Project contents include creating data query specifications, consulting with clinical investigators on sponsors' statistical design needs, editing protocols, building models, and creating detailed analysis reports.
- Built data visuals to display analysis results. The visuals were incorporated into a business advertising presentation.
- Developed a SQL builder prototype, which aimed to speed up query processing by 50% and laid foundation for another internal project.

Columbia University Medical Campus *New York, NY*
Research Assistant (Survey Data) *October 2018 - present*
Structural Determinants & Social Transitions among Adolescents and Young Adults in Rakai, Uganda

- Clustered subjects' self-reported characteristics of sexual partners into different classes using Latent Class with Random Effect Analysis, to solve the conflict of correlation between subject responses and conditional independence assumption of Latent Class Analysis.
- Performed data cleaning with longitudinal survey data to prepare for analysis.
- Compared Latent Class with Random Effect Analysis results with LCA results qualitatively and quantitatively.
- Created codebook to clean historical data and standardize coding rules.

RELEVANT PROJECTS

Flight Delay Patterns -- *R* (website is available here)

- Explored factors associated with flight delay patterns from 2012 through 2016.
- Processed 2.8GB of data and analyzed 29M+ flights records using a virtual server and Linux.
- Created a website to display project findings, including a report, video, and interactive data visuals.

Using Statistical Learning to Analyze Airbnb Listings -- *R*

- Built predictive models of rental prices; a model was selected from various statistical learning algorithms of linear and nonlinear regression based on performance (using methods include ridge, lasso, PLS, splines, GAM).
- Classified high/low review score by selecting model from logistic regression, LDA, QDA, tree methods, and SVM.

Examining the Impact of the New York Advantage Program on the NYC Homeless Population -- *Python* (report is available here)

- Evaluated the effect of the New York Advantage program on the NYC homeless population using a regression model.
- Did explorative analysis and visualization to help understand patterns of change in the NYC homeless population and how patterns related to other relevant variables (population growth, policy changes, seasonal changes, etc.)

Comparing Effects of LAGB and RYGB, and Examine Associations in Weight Loss -- *SAS* (post is available here)

- Built models to identify measured variables associated with weight loss.
- Compared obesity treatments' effects overtime using longitudinal data analysis.
- Created descriptive statistics for relevant variables.

SKILLS

- Statistical languages: R (& Shiny), Python, SAS, SQL
- Bilingual: Chinese (native), English (fluent)