

YUNQIU (LULU) YAO

100 Haven Avenue Apt 19G, New York, NY 10032
yy2827@columbia.edu | (646)-683-6625 | linkedin.com/in/yunqiu-yao

EDUCATION

Columbia University

New York, NY

Master of Science in Biostatistics, GPA 3.95/4.0

May 2019

- Courses: Database, Statistical Learning, Deep Learning, Probability, Inference, Survival Analysis, etc.
- Graduate Teaching Assistant for *P8100 Applied Regression*

Shanghai Jiao Tong University

Shanghai, China

Bachelor of Science, Food Science and Engineering

June 2017

- Academic Progress Scholarship; Academic Excellence Scholarship

SELECTED PROJECTS

Is Venmo Safe to Use?

Sept. 2018 – Present

- Developed Python code to web-scrape the recent public transaction records via Venmo public API
- Constructed a database with PostgreSQL on the Google Cloud Platform, calculated the risk score for each user to remind them of the potential information leakage, and implemented the application with Flask

Study on the Readmission Rate for Diabetes

Mar. 2018 – May 2018

- Analyzed 67,069 electronic medical records regarding the readmission status of patients with diabetes
- Constructed models with kNN, random forest, support vector machine, logistic regression algorithms to evaluate the efficacy of factors such as different treatments on readmission rate and make predictions

Why is a TED Talk Popular?

Nov. 2017 – Dec. 2017

- Collaborated on GitHub and performed regression, sentiment analysis and data visualization in R Markdown with dplyr, ggplot2, shiny and plotly packages to determine the key factors influencing views
- Integrated findings in a scientific report on project website and recorded a screencast for presentation

Empirical Bayes (EB) Method for Haplotype-based GWAS

Jan. 2016 – Oct. 2016

- Utilized R to collaborate on the construction of a linear mixed model (LMM) for GWAS with EB theory
- Spearheaded the initiative to apply the constructed EB-LMM model to haplotype-based GWAS, applied the model to the genome of 1092 subjects, and inferred 3 genes associated with the trait of interest

INTERNSHIPS

Cepheid

Sunnyvale, CA

Biostatistics Intern

May 2018 – Aug. 2018

- Implemented R/SAS to perform/validate analyses for the FDA submission of clinical trials, including Cox PH models, reproducibility studies, sample size calculation, contingency tables, linearity test, etc.
- Developed SAS Macros with SQL for multiple raw clinical datasets to streamline data manipulation
- Summarized and presented the works to the Clinical Affairs group and suggested techniques to improve

Edenred – Accentiv'

Shanghai, China

Data Analyst Intern

Dec. 2016 – Feb. 2017

- Wrote SAS and SQL scripts to automate the data cleaning and produce weekly/monthly sales report
- Utilized R/Python to construct a recommendation engine with clustering, association rule mining, etc
- Analyzed 1.86 million sales records, held discussions to troubleshoot the problems and contributed to future marketing strategies for different groups with different purchase habits

PUBLICATION

- Chen, Z., Yao, Y., Ma, P., Wang, Q., & Pan, Y. (2018). Haplotype-based genome-wide association study identifies loci and candidate genes for milk yield in Holsteins. *PloS one*, 13(2), e0192695.

SKILLS

- Technical: Python, R, SQL, SAS, SPSS, Latex, Perl, MS Office (Word, Excel, PowerPoint)
- Python libraries: numpy, pandas, sklearn, matplotlib, keras, tensorflow, etc.
- R packages: tidyverse (dplyr, ggplot2, tidyr, etc.), shiny, plotly, flexdashboard, etc.