

YUNQIU (LULU) YAO

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

EDUCATION

Columbia University

New York, NY

Master of Science in Biostatistics, GPA 3.8/4.0

May 2019

- Courses: Probability, Data Science, GLM, Statistical Learning, Epidemiology, RCT, Survival Analysis
- Teaching Assistant for *Applied Regression Models*: instructing 50+ students on regression and SAS

Shanghai Jiao Tong University

Shanghai, China

Bachelor of Science, Food Science and Engineering

June 2017

- Academic Progress Scholarship; Academic Excellence Scholarship

INTERNSHIP

Edenred – Accentiv' China

Shanghai, China

Data analyst

Dec. 2016 – Feb. 2017

- Wrote SAS and SQL scripts to automate the data cleaning and produce weekly/monthly sales report
- Implemented cluster analysis in SAS to analyze 1.86 million purchase records from Haagen-Daz
- Presented key findings to supervisor and contributed to recommendations for future marketing strategies for different groups with different purchase preferences and habits
- Collaborated with 3 teammates on the construction of a recommendation engine for L'OCCITANE, including the data cleaning, clustering and association analysis in SAS Enterprise Miner

PROJECTS

Why is a TED Talk Popular?

New York, NY

Columbia University, Data Science course

Nov. 2017 – Dec. 2017

- Collaborated on GitHub to clean raw data in R Markdown and established the website for the project
- Visualized data graphically and conducted sentiment analysis with Shiny and Plotly package
- Performed regression to determine key factors influencing views and built model for future prediction
- Integrated findings in a scientific report on project website and recorded a screencast for presentation

Empirical Bayes (EB) Method for Haplotype-based GWAS

Shanghai, China

Research Assistant, Shanghai Jiao Tong University, Professor Qishan Wang

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, and applied to the genome of 1092 subjects
- Identified 17 significant haplotype blocks and inferred 3 genes associated with the trait of interest
- Drafted a publishable report based on the study findings and presented to PI overseeing the research

Proteome-wide Association Study for Colorectal Cancer

Shanghai, China

Research Assistant, Shanghai Jiao Tong University

June 2016 – Oct. 2016

- Standardized and converted proteomic data from 95 cases and 32 controls into two integrated matrices
- Introduced the Protein Interaction Network (PPI) to pair the mutant proteins with the neighbors
- Applied GLM in R to identify the key mutant proteins significantly affecting neighbors' expression values
- Performed BH correction and found 58 significant proteins out of 1189 candidates for colorectal cancer

Logistic Regression for Metastases in Renal Cell Carcinoma

Shanghai, China

Shanghai Jiao Tong University, Biostatistics course

Nov. 2015 – Dec. 2015

- Conducted a study of 58 patients on metastasis status of renal cell carcinoma based on 5 clinical features
- Performed logistic regression, identified 3 key determinants of metastases, and built the model for further prediction of metastasis based on the clinical features

SKILLS

- Technical: R, SQL, SAS, Python, Perl, MS Office (Word, Excel, PowerPoint)
- Language: English (Proficient), Chinese (Fluent), Korean (Conversational)