XIAOKE (COCO) ZOU

2830 Jackson Ave, Long Island City, NY | 424-365-2099 | xz2809@columbia.edu

A qualified data scientist candidate with extensive understanding of machine learning, data mining, clinical trials analysis, and big data analytics and related workflow.

Skills & Certification

Computer Skills: Advanced Programming in R, Python, PySpark, SQL and SAS. Proficient in Microsoft Office (Word, Excel, PowerPoint, Access), Git, MATLAB and C++
Certification: SAS Certified Base Programmer for SAS 9
Language Skills: Fluent English, Native Mandarin

Education

Columbia University, Mailman School of Public Health — **New York, NY**
Master of Science in Biostatistics: GPA 3.8/4.0 — May 2020

University of California, Los Angeles — **Los Angeles, CA**
Bachelor of Science in Chemical Engineering, Biomolecular Option: GPA 3.5/4.0 — June 2018

Work Experience

Mount Sinai Hospital — **New York, NY**

Bioinformatics Summer Intern, Department of Neuroscience — June 2019 – August 2019

The goal of this project is to quantify cocaine-induced cell-type specific gene expression changes

- Explored, analyzed and performed quality control of more than 1 GB gene expression matrix data and metadata to assess effects of different cocaine challenges on mice brain cells in R
- Performed K-Means and consensus clustering algorithms on gene expression matrix to identify cell types
- Developed differential analysis incorporating different covariates, including stimulation and treatment conditions, to quantify gene expression changes using generalized linear model and hurdle model
- Completed data visualization process by Gene Ontology Enrichment Analysis and cell-type-to-type comparison using Rank-Rank Hypergeometric Overlap

Columbia University, Mailman School of Public Health — **New York, NY**

Teaching Assistant, Data Science and Applied Regression I — September 2019 – December 2019

- Led weekly office hour to assist more than 200 students with performance in R coding and SAS coding
- Graded weekly assignments and provided feedback to help students improve understanding of course materials

Columbia University, Mailman School of Public Health — **Shanghai, China**

Summer Intern, Research & Development Department — Summer 2015

- Cooperated with six colleagues to create relational database using SQL in Access to improve the efficiency of oncology patient data searching

Academic Projects

Columbia University, Mailman School of Public Health — **New York, NY**

Data Science. — November – December 2018

The goal of this project is to study the incident rate of TB among patients with Type II Diabetes

- Chaired code writing for statistical analysis using GLM, Cox Regression and Kaplan-Meier estimator incorporating patient level and hospital level variables to predict the incident rate of TB and survival time of Type II Diabetes patients
- Established a shiny map and plots in website to dynamically demonstrate the predicted survival curve of Type II Diabetes patients based on customized input

Clinical Trails Analysis — February 2020

- Provided consultation to pathologists from New York-Presbyterian Hospital about the reproducibility of NAS score and its correlation with steatohepatitis diagnosis
- Applied repeated measure ANOVA to calculate the ICC to evaluate the accuracy of NAS Score and its components between independent pathologists

Big Data Analysis — November – December 2019

- Built a movie recommendation system to recommend personalized films for each user based on their previous activities using Google Bigquery, PySpark, and GUI with multiple machine learning algorithms