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

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

Education

Columbia University, Mailman School of Public Health — **New York, NY**
Master of Science in Biostatistics: GPA 3.8/4.0 — Expected Graduation: 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 to gene expression matrix to identify cell types
- Used GLM to perform differential analysis incorporating different covariates, including stimulation and treatment conditions, to quantify gene expression changes
- Performed Gene Ontology enrichment analysis and cell-type-to-type comparison using Rank-Rank Hypergeometric Overlap to visualize results

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

Sanofi — **Shanghai, China**
Summer Intern, Research & Development Department — Summer 2015

- Cooperated with six colleges 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 — December 2018

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

- Collaborated with four colleges and chaired code writing for statistical analysis using GLM, Cox Regression and Kaplan-Meier estimator incorporating diabetes age, glucose management, drug usage level, complications level, and daily exercise level 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, including gender, drug usage and exercise level.

Clinical Trails Analysis — February 2020

- Provided consultation to pathologists from New York-Prebysterian 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 — 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 ML algorithms