

Langyi Tian

Phone: +1 347 406 1727

Email: langyi.tian@columbia.edu

GitHub: github.com/lt2710

EDUCATION

M.A. Quantitative Methods in the Social Sciences (Data Science Focus)

12/2019 (expected)

Columbia University (New York, NY)

- Courses: Machine learning, Data visualization, Data structure, Quantitative methodology, Marketing insights
- Research grants: QMSS Travel Grant, ASGC Travel Fund, GSAS Matching Fund
- New York Open Statistical Programming Meetup: attended workshops on visualization and time-series

B.B.A. Global Business Studies and Finance

05/2018

The Chinese University of Hong Kong (Hong Kong)

SKILLS

R (machine learning, clustering, visualization, GIS, statistical inference, web app via Shiny, reporting via Xaringan), **Python** (scikit-learn), **Amazon Web Service** (EC2, S3), **Git**, **SQL**, **Tableau**, **Qlik Sense**, **Apache Spark** (via R)

PROFESSIONAL EXPERIENCE (SELECTED)

Data science contractor

10/2019 to 12/2019 (expected)

A public policy NGO in California

- Built regular reports and a Shiny GUI open to education researchers from public data (ACS, IPEDS, etc.)

Data science intern [\[Code\]](#) [\[Deliverable\]](#)

05/2019 to 10/2019

14 WesTech (Baltimore, MD) 14 WesTech is a service provider in marketing intelligence, operations, and cybersecurity products.

- Defined an ETL process from scratch with 2 call centers to address the data analytical demand from clients
- Presented automated visualization reports with interactive graphs on retention and performance
- Deployed a dashboard on revenue and profit projection with financial analysts for financial planning
- Proposed Agile process and Kanban to manage diverse stakeholders and coordinate cross-functional workflow

Predictive modeling intern [\[Code\]](#) [\[Deliverable\]](#)

02/2019 to 05/2019

First Street Foundation (New York, NY) First Street Foundation is a tech non-profit quantifies the risk of sea-level rise and flood.

- Predicted market values of 3 million properties in South Florida from administrative and demographic data
- Fitted 85 models with regularization (Ridge, LASSO) and trees (random forest and GBM) for city subsamples
- Made functionalities to choose and validate the best parameters to filter input data and improve performance

ACADEMIC RESEARCH (SELECTED)

The impact of parental wealth on offspring living standards [\[Code\]](#) [\[Deliverable\]](#)

10/2018 to 07/2019

- Apply sequence analysis (optimal matching) and K-means clustering to categorize job trajectories
- Modeled education, income, and wealth with OLS, quantile estimates, cumulative link, and logit regressions
- Presented at the American Sociological Association Annual Meeting (New York University, Aug 2019)

Consumer segments: exploration with unsupervised learning [\[Code\]](#) [\[Deliverable\]](#)

05/2017 to 09/2018

- Used Multiple Correspondence Analysis (PCA for categorical variables) to visualize consumer dimensions
- Applied bootstrapped hierarchical clustering, K-means, and K-medoids to generate 4 consumer clusters
- Visualize the geographical distribution of consumer group clusters with GIS packages
- Presented at the International Sociological Association Summer Meeting (Princeton University, Aug 2019)