

# Langyi Tian

Phone: +1 (347) 406 1727

Email: [langyi.tian@columbia.edu](mailto:langyi.tian@columbia.edu)

GitHub: [github.com/lt2710](https://github.com/lt2710)

## SUMMARY

Aspiring data science professional with market research experience. Having leveraged exploratory, inferential, and predictive algorithms as well as delivered insights via automated reports and dashboards professionally. Experienced in communicating insights with non-technical stakeholders and general audiences. Three quantitative papers under review.

## EDUCATION

**M.A. Quantitative Methods in the Social Sciences (Data Science Focus)** 12/2019 (expected)

*Columbia University, Institute for Social and Economic Research and Policy (New York)*

- Courses: Machine learning, Data visualization, Data structure, Quantitative methodology, Marketing insights
- Grants: QMSS Travel Grant, ASGC Travel Fund, GSAS Matching Fund
- New York Open Statistical Programming Meetup: meet for socializing, training and latest updates on R

**B.B.A. Global Business Studies and Finance (double major program)** 05/2018

*The Chinese University of Hong Kong (Hong Kong)*

- Grants: SHSS Travel Fund, Undergraduate Research Award

## SKILLS

**R** (machine learning, clustering, visualization, GIS, statistical inference, dashboarding via Shiny, reporting via Xaringan), **Python** (scikit-learn), **Amazon Web Service** (EC2, S3), **GitHub**, **SQL**, **Apache Spark** (via R), **Tableau**, **VBA**, **Stata**

## PROFESSIONAL EXPERIENCE (SELECTED)

**Data science/people analytics intern (R, AWS)** 05/2019 to 10/2019 (expected)

*14 WesTech (Baltimore, MD)* 14 WesTech serves clients with marketing intelligence, operations, and cybersecurity products.

- Led to define an ETL process from scratch to integrate structural and unstructural data within 2 call centers
- Presented automatized visualization reports with interactive graphs on factors of retention and performance
- Applied random forest algorithm to predict agent performance metrics based on early career information
- Developed a dashboard with financial analysts to leverage the insights on retention and profitability analysis
- Proposed Agile process and Kanban to manage diverse stakeholders and coordinate cross-functional workflow

**Predictive modeling intern (R, AWS)** 03/2019 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
- Performed data wrangling, feature engineering, and missing value imputation
- 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
- Built automatized report parsing and visualizing model performance delivered to the head of data science

**Market research consultant** 11/2018 to 03/2019

*Porteñas (New York, NY)* Porteñas is a bar startup offering premium South American drink and food.

- Explored Chinese market entry solutions for yerba mate (an Argentina drink) portfolio on targeting and branding
- Led 4 focus groups, 3 virtual ethnographies and a sequential monadic survey to test product

**Finance intern (Excel)** 02/2017 to 05/2017

*PwC/PricewaterhouseCoopers (Shenzhen, CN)*

- Applied VBA to automate Excel input tasks for regular data entries including dates and institution names

**Marketing research intern** 09/2016 to 11/2016

*D'Addario & Co. (Shenzhen, CN)* D'Addario is one of the world's largest manufacturers of instrument accessories.

- Sourced and interviewed 7 customers to portrait customer persona for consideration of low-end market entry

## RELEVANT RESEARCH (SELECTED)

**The impact of parental wealth on offspring living standards (R, AWS)** 05/2019-today

*Columbia University*

- Aggregated data by restructuring more than 20 relational data sets from a large-scale panel survey
- Apply sequence analysis (optimal matching) and K-means clustering to categorize job trajectories
- Modeled education, income, and asset with OLS, quantile estimates, cumulative link, and logit regressions
- Presented the paper at the American Sociological Association Annual Meeting (New York, Aug 2019)

**Consumer segments: an exploratory approach with unsupervised learning (R)** 05/2017 to 09/2018

*The Chinese University of Hong Kong*

- Used Multiple Correspondence Analysis (MCA) to visualize dimensions of consumer segments
- Applied bootstrapped hierarchical clustering, K-means, and K-medoids to generate 4 subgroups
- Visualize the geographical distribution of consumer group clusters with GIS packages
- Presented the paper at the International Sociological Association RC28 Summer Meeting (Princeton, Aug 2019)