Langyi Tian

Phone: +1 (347) 406 1727 Email: langyi.tian@columbia.edu GitHub: github.com/lt2710

SUMMARY

Aspiring data science professional with prior business experience. Having leveraged exploratory, inferential, and predictive algorithms as well as delivered insights via automated reports and dashboard apps in a professional setting. Real-world exposure to call center, real estate, and large-scale consumer survey data. 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: Applied machine learning, Machine learning with Python, Modern data structure, Data visualization
- Grants: QMSS Travel Grant, ASGC Travel Fund, GSAS Matching Fund
- New York Open Statistical Programming Meetup: meet with R community for socializing, updates and training

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, maps, inference, dashboarding via Shiny, reporting via Xaringan), **Python** (scikit-learn), **Amazon Web Service** (EC2, S3), **GitHub**, **SQL** (via R), **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.*

- Defined an ETL process from scratch to integrate two call centers' data from HRIS, Five9, and QA
- Presented automized visualization report on factors impacting retention and performance to client executives
- Applied random forest algorithm to predict agent performance over tenure based on early career performance
- Developed a web dashboard via Shiny for HR and financial analysts to use in strategic planning meetings
- Proposed Agile development and Kanban to work with diverse stakeholders and cross-functional colleagues

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 value imputation to improve model performance
- 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 automized 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)

- Processed ERP journal entries, Excel data and made financial statements to assist annual internal audit project
- 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.*

- Led intern team to source and interview artists to facilitate string products performance improvement
- Screened 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

- Apply sequence analysis (optimal matching) and K-means clustering to categorize job trajectories in AWS
- Modeled living standard measures with OLS, quantile estimates, cumulative link models, and logit regression
- Presented the paper at the American Sociological Association Annual Meeting (New York University, Aug 2019)

Lifestyle segments: an exploratory approach with data mining (R) **05/2017 to 09/2018**
The Chinese University of Hong Kong

- Applied MCA, bootstrapped hierarchical clustering, K-means, and K-medoid to generate 4 lifestyle clusters
- Visualize geographical distribution of clusters in lifestyles with GIS visualization packages in R
- Presented the paper at the International Sociological Association RC28 Summer Meeting (Princeton, Aug 2019)