Always latest here: https://yuchentang.github.io/

Yuchen Tang

yt2754@columbia.edu (646)220-7903 LinkedIn

Education

Columbia University | New York, U.S. **Jan. 2021–May 2022**
Master of Statistics GPA: 3.97/4.0
Core Courses: Statistical Inference, Linear Regression Models, Time Series Modelling, Sample Surveys

University of International Relations | Beijing, China **Sept. 2015–Jun. 2019**
Bachelor of Engineering in Management Information Systems GPA: 3.73/4.0
Core Courses: C Programming, Data Structure and Algorithms, Principles of Database, Data Mining

Peking University | Beijing, China **Sept. 2016–Jun. 2019**
Bachelor of Economics GPA: 3.47/4.0
Core Courses: Intermediate Microeconomics & Macroeconomics, Econometrics, Game Theory

Skills

Hard Skills MS Office Toolkit, R == Statistics > SQL > Python, Jupyter Notebook, (R)Markdown, Latex, Tidyverse, RShiny, Data Visualization (ggplot2, seaborn), Hypothesis Testing, ML Models: Linear Regression, Logistic Regression, Tree-based Models, Neural Networks, Clustering

Soft Skills Self-learning, Multi-tasking, Solution-oriented

Internship

Teaching Assistant | Dept. of Statistics, Columbia University | New York **Sept. 2021–May 2022**
- Reviewed homework and Q&A for undergraduate students on statistical theories and R coding

Data Management Intern | Qichacha Tech. | Suzhou, China **Jul. 2020–Oct. 2020**
- Took part in the design of a new section of bidding information, including determining the organization and features
- Collected 300+ data sources manually; Wrote SQL queries to check the incorrect and missing data, and communicated with the web crawler team to do corrections

Assistant Analyst | TokenInsight Tech. | Beijing, China **Apr. 2019–Jul. 2019**
- To identify the wash trading of cryptocurrency exchanges, proposed the methodology that the Power Law can be used to check the distributions of orders of exchanges, then identify the unusual patterns and estimate real trading volumes
- Realized the idea with R and Python using the raw order book data, validated its effectiveness by data manipulation (bins, outliers), visualization (histograms, scatter plots), and curve fitting (linear regression)
- Analyzed 10+ exchanges, found the evidence and volumes of their washing trading; Finished the corresponding parts in the reports; Became the first company to release reports on cryptocurrency exchanges washing trading in China

On-campus Projects

How To Increase Happiness **2022**
- Preprocessed data using Python, like one-hot/label encoding, standardization, and aggregation of features; Modeled logistic regression and gradient boosting trees with the scikit-learn package, tuned the hyperparameters by grid search and cross-validation, evaluated them by F-1 score and ROC/AUC
- Extracted the common important features from them by coefficients and feature importance, found that comparison with peers, family's social/economic status, and health are important, and gave corresponding practical suggestions

Loan Origination Prediction/P2P Bad Debt Forecasting **2021/2019**
- **(2021)** Did EDA and visualization by the ggplot2 package in R, balanced the dataset using SMOTE-NC algorithm, and used the FAMD technique to generate principal components for mixed data
- Modeled with logistic regression with L-1 regularization for feature selection; Did a simple performance comparison with random forest, XGBoost and ANN; Drove accuracy from 65% to 72%
- Did research about the credit domain and concluded the lack of key variables like credit history
- **(2019)** Similar preprocessing and modeling using R; Found that variables selected by the stepwise logistic regression performs better in neural networks

Does the High Housing Price Lower the Fertility Rate **2021**
- Collected and joined different data sources, like World Bank, to generate a panel data; Preprocessing like imputing missing values and standardization in R; Did visualization to check the integrity and understand the data intuitively
- Used fixed effect regression to deal with time-invariant confounders; Did literature review and determined time-varying confounders, like household wealth
- Inspected parameter uncertainties by grouping the data. Found that high housing prices don't lead to low fertility rates, and even have a positive effect in low-income regions and for rich families. Corresponding explanations given