# FRANK XU

https://medium.com/@frankxu0124

(919)-308-3187 | frankxu0124@gmail.com

## PROFILE

- **A DUKE Data Science graduate student & certified writer for *TowardsDataScience.com***
- **Experienced in Data Cleaning, Machine Learning and Databases;**
- **Focused on Model Interpretation during internship at Wells Fargo;**

## TECHNICAL SKILLS

- Data Analysis and Visualization: **Python, R, Tableau, Excel;**
- Machine Learning and Deep Learning: **Scikit-Learn, Keras, TensorFlow, H2O;**
- Data Management: **SQL/NoSQL (MySQL, SQLite, Hadoop, Spark);**
- Languages: Fluent English; Native Mandarin, Shanghainese;

## EXPERIENCE

**Intern Data Scientist**                                                                     Aug. 2019 – Present
*SNAP* group, TradePending                                                                          Carrboro, NC
- Using a GCP VM to deal with a 72-million rows vehicle transaction and trying to improve the existing model using geographical information and clustering techniques.

**Data Science Intern**                                                                    June. 2019 – Aug. 2019
Wells Fargo Bank, Data Management and Insights                    San Francisco, CA/Charlotte, NC
- Analyzed customer data and built/tuned RF/GBM models to recommend products.
- Focused on different ways of measuring feature importance of RF models, including H2O, Shap, Boruta, Ranger, Lime.

**Data Analyst Assistant**                                                                     Jan. 2018 – Feb. 2018
EPA, government of Dezhou City                                                                 Shandong, China
- Analyzed hourly air pollution and monitored excessive emission; modeled with R and presented more than 20 charts every day for the government's project of improving air quality.

## EDUCATION

**Master of Interdisciplinary Data Science** (expected 2020)                   Aug. 2018 – Present
Duke University, Information Initiative at Duke                                                  Durham, N.C.
- Enrolled in the inaugural class with an acceptance rate less than 5%
- Core courses: Bayes Inference, Machine Learning, Text Analysis, Stochastic Process, etc.

**Bachelor of Science in Environmental Science**                                              2014 – 2018
Fudan University (Major GPA: 3.54/4.0, top 10%)                                          Shanghai, China
- Awards: Outstanding Student Scholarship for three consecutive years (2015-2017)

**Minor Study in Accountancy**                                                                     2015 – 2017
Shanghai University of Finance and Economic                                               Shanghai, China
- Core courses: Principles of Economics

## PROJECTS

**Kaggle Competition: Identifying Solar PV in Aerial Imagery**               Jan. 2019 – Feb. 2019
- Used a Convolutional Neural Network built based on MobileNet to achieve a 97.8% accuracy on identifying solar panels in aerial images and ranked 10% tier.

**Analyzing Data for Real Talk**                                                             Sept. 2018 – Dec. 2018
- Real Talk is a mobile app available for iOS geared for teenager users. The app allows teens to submit anonymous stories about topics such as puberty, bullying and relationships.
- Ran text & regression analysis with R on a dataset of around 1400 stories, unveiled the connection between teenager psychometric status with demographic information.

**Analyzing Annual Pollution Data of Shanghai**                                                2017 - 2018
- Used R to scrape back trajectory data from NOAA, built linear regression models for the air pollution data of Chongming, Shanghai from 2017 – 2018.