(217)-898-5526

San Jose, CA, 95129 [yangshawnlong@gmail.com](mailto:yangshawnlong@gmail.com)

**Technical Skills**

https://github.com/ShawnXiaolongYang https://[www.linkedin.com/in/xyang105/](http://www.linkedin.com/in/xyang105/) https://profiles.udacity.com/p/4464378536

*Languages:* Python, R, Matlab, SQL, Html, JavaScript, Java Servlet, C#, VB, C, C++ *Libraries:* Tensorflow, NumPy, Pandas, Scikit-Learn, NTLK, Graphlab, gglot2, Flask *Database:* Neo4j, MongoDB, SQL server, Mysql

**Xiaolong Yang**

*Software:* Tableau, Octave, Anaconda, RStudio, Jupyter, Dreamweaver, Visual Studio, Myeclipse

# Work Experience

## Data Mining Intern

June 2017 - now, MobileIron.co

Focus: Data Mining, Postgresql, Mysql, Tableau

* Used P-sql and Mysql to collect data in different product and different sources from company.
* Cleaned the data collect from different sources and found the synonyms name in different products with Open refine
* Analyzed result using R and ploted the graph using Tableau
* Applied machine learning algorithm to make recommendation

## Data Analyst

January 2017 – April 2017, Udacity Focus: Data Visualization, A/B test

*A/B Test* - https://github.com/shawnxiaolongyang/A-B-test

* Designed an A/B test to find whether add new features would affect the courses enrollments and payments
* Analyzed the result of the former test, and design a followingtest.

*Data Visualization* - https://shawnxiaolongyang.github.io/

* Analyzed Divvy Bike over 2 million lines data, found the pattern of customers’ behavior.
* Plotted the riding paths to show the distribution of customers’ behavior at time and geography.
* Applied dimple.js and d3.js to plot the charts and hierarchy bundle graphs.

## Research Assistant

June-December 2016, FORWARD Data Lib, University of Illinois

Focus: Graph Analysis, Deep Learning, Recommendation System, Website Develop

*Recommendation System Design***-** https://github.com/shawnxiaolongyang/Researchers-Recommendation

* Developed a recommendation system to recommend valuable papers and potential researchers to users.
* Parsed and cleaned data collected from dblp in different format, transferred 5GB data in graph database Neo4j.
* Combined graph analysis algorithms Page-rank and deep learning algorithm Node2Vec for recommendation.

*Web-app Design***-** https://github.com/shawnxiaolongyang/Researchers-Recommendation/tree/master/WebRoot

* Developed a professional social network app for researchers
* Designed the UI with Bootstrap and Back-End with Java Servelt Page and Flask
* Combined the Recommendation System and Researcher Application together and published on Linux Server

## Data Analyst

June 2015 – May 2016, Udacity

Focus: Machine Learning, Text Mining, Data Wrangling

*Fraud Email Classification* **-** https://github.com/shawnxiaolongyang/Enron-Fraud-Email-Classification

* Identified who was fraud in Enron Scandal by analyzing the 500,000 employees’ emails.
* Applied algorithms such as Gaussian NB, TFIDF, Adaboost, Random forest and KNN to classify employees.
* Verified results using Recall-Precision, used Gini importance to select features.

*Exploration Data Analysis* - https://github.com/shawnxiaolongyang/Propsper-Loan-Data-Analysis

* Analyzed Prosper Loan over 110,000 lines data with 81 variables using exploratory data analysis.
* Explored univariate, bivariate and multivariate analysis using correlation, ANOVA, AIC via R.
* Applied dplyr to subset, validate and summarize data and ggplot2 to plot Heat map and GIS map.

*Data Wrangling* **-** https://github.com/shawnxiaolongyang/Open-Street-Map-Data-Wrangling

* Assessed 1 GB Tianjin City’s Map JSON data for validity, accuracy, completeness, consistency, and uniformity.
* Assembled MongoDB for data storage and applied Regular Expression and MongoDB query in data cleaning.
* Deep-cleaned address data written in different formats, using synonymous names or with wrong postal codes.

## Web developer

Focus: Web development, Data Visualization

November 2014 – March 2015, Tianjin Maritime Safety Administration, China

### *Questionnaire Design Website*

* Developed a questionnaire survey website for users to test customer satisfaction
* Helped user easily design questions in a tree style frame and stored in Sqlserver
* Automated plotted the survey result with Chart module in C#

September - October 2014, C.G Xiao’s Hospital, China

### *Login Website*

* Developed a website for Doctors to login an intelligent medicine cabinet system
* Used VB for Front End design and Mysql to save the data

# Education

#### University of Illinois at Urbana-Champaign

*Computer Science-Data Science, M.S.* GPA: 3.7 / 4.0

*Double Major: Technology Management, M.S.*

Courses in Data Mining, Cloud Computing, Applied Machine Learning, Models of Applied Statistics, Statistical Learning, Data Visualization, Data Cleaning, Text Mining and Information Retrieval, Project Management.

#### Udacity

*Data Analyst,* Machine Learning Engineer *Nanodegree*

Courses in Data Wrangle, EDA, Feature Engineer, Reinforcement Learning, Deep Learning, Data visualization, A/B test.