

# YUJI SHEN, M.S.

Software Engineer & Data Analyst

yuji.im.shen@gmail.com

(412) 228-1560

<http://yuji.im>

## TECHNICAL SKILLS

Python | R | Java | SQL | Git | D3.js | Apache Spark | Hadoop | Lucene | HTML | CSS

NumPy | Pandas | Scikit-Learn | Matplotlib | caret | ggplot2 | AWS

Certificates of Achievement from **JHU, MIT, Duke, Stanford, Berkeley** in Data Science & Big Data

## WORK EXPERIENCE

**Data Analyst Intern** - *Changsha Excellent Internet Info Service, China* 2013

- Led group of interns in participation of construction of Business section to attract investment
- Scraped information in web crawler tools, filtered through Java to collect sustainable companies' info
- Improved intern group data collecting productivity using data technologies

## ACADEMIC & PERSONAL PROJECTS

**U.S. Sexual Opinion Research** - <http://yuji.im/sex-edu> 2015

- Composed research report on sexual opinion based on GSS and Fair's extramarital affairs data
- Applied hypothesis testing in **R** to show the relationship between education and opinion on affairs
- Created plots for trends and comparisons in **ggplot2**, generated HTML report in **R Markdown**

**Display Ads Challenge** 2015

- Predicted users' click-through rates on display ads based on large dataset
- Configured IPython server and **Apache Spark** cluster on **AWS** across two EC2 instances
- Performed feature engineering to handle categorical data, reduced sparse dimensions, predicted probability using Logistic Regression in Spark **Python** API

**Pittsburgh Analysis & Visualization** - <http://yuji.im/vizburgh> 2014

- Analyzed socio-economic data of Pittsburgh, built interactive visualization of data
- Utilized visualization techniques and **K-Means clustering** to analyze the data in R
- Visualized data by maps and multiple plots in **D3.js**, reduced workload by localStorage in HTML5

**Twitter Hashtag Search Engine** - <http://yuji.im/hashtag> 2014

- Constructed, deployed search engine on AWS to recommend Hashtag based on user's tweet
- Extract fields and built inverted index from UDI Twitter dataset in **Java**
- Implemented searching algorithm in **Apache Lucene** with Vector Space Model and Boolean Model
- Optimized relevance by combining scores from different fields and using fuzzy query

**E-Commerce Database System** - <http://yuji.im/ecom-db> 2013

- Developed an E-Commerce OLTP system and monitored database performance
- Devised 15 tables in 2NF with 1M rows dummy data and built B-tree indexes in **MySQL**
- Implemented product searching pages and top sales report functions with **PHP**

## EDUCATION

**Data Science Mentee** - *SlideRule Data Science Workshop, CA* 2015

- Mentored by data scientists from Deloitte and Glassdoor
- Built portfolio site and completed personal projects throughout the entire pipeline of data science

**M.S. Information Science** - *University of Pittsburgh, PA* 2015

- Graduated *summa cum laude*
- Collaborated with colleagues to accomplish academic projects using state-of-art technologies

**B.B.A. E-Commerce** - *Hunan University, China* 2013

- Graduated *cum laude*