

# YUAN ZHANG

✉ yuanz2299@gmail.com 📁 yzhang0.github.io 🌐 yzhang0 📍 Looking for Jan 2019 internship

## SKILLS

---

**Languages** Python (pandas, scikit-learn, flask), Java, SQL, C++, C, D3.js, R (dplyr, xgboost)  
**Tools/OS** Hadoop (Spark, MapReduce, Oozie, Hive, Sqoop), git, bash, Linux, LaTeX  
**Interests** Applied Machine Learning, Data Visualization, Time Series, Deep Learning

## EXPERIENCE

---

**Salesforce** May 2018 - Aug 2018  
*Software Engineering Intern* San Francisco

- Develop **ETL pipelines** and **UDF** to efficiently store, query and analyze data in **Hadoop (Oozie, Hive)**.
- Optimize platform components using **Python** and **Java** with replication and scalability features to transport large scale data from key-value store to RDBMS.

**Scotiabank** Sep 2017 - Dec 2017  
*Data Science Intern* Toronto

- Designed and developed a model using **supervised learning** in **Hive** and **R** to predict digital banking likelihood of new customers with an accuracy of **73%**.
- Cleaned and analyzed millions of customer browsing data points using the **Hadoop Framework (HDFS, MapReduce, Spark)** to investigate the efficiency of third party data for marketing campaigns.
- Delivered presentations with data visualizations on how to leverage machine learning on the new digital platform for leadership of different technical familiarity.

**IBM** Jan 2017 - Apr 2017  
*Data Analytics Developer* Ottawa

- Wrote **Python** scripts to detect anomaly and user trends using hourly warehouse data from CouchDB.
- Built an **automated** system to create scheduled data visualization using D3.js and Jinja2.
- Made recommendations on data visualizing libraries based on experiments with Kibana, Plot.ly, and NVD3.
- Built and tested a real-time data processing program using **reactive programming** to generate log files.

**Agriculture and Agri-Food Canada** May 2016 - Aug 2016  
*Web Statistics Specialist* Ottawa

- Assisted users in interpreting **web statistics** reports and investigated possible discrepancies.
- Trained a team of two interns and five full time and wrote a 15-page reference manual on web analytics.

## PROJECTS

---

**Bookclub** Jul 2018

- Created a website where users can sign in to review 5000 books from the GoodRead API.
- Interacted with database using **SQLAlchemy** and created web application using **Flask** and **Bootstrap**.

**Predict and Visualize Software Internships Trends at UWaterloo** Jan 2018

- Built a reactive directed graph of 86 most recurrent companies and their number of interns using **D3.js**.
- Collected company ratings, revenue, etc. from Glassdoor using **Selenium** and **Requests** in Python.
- Trained an **ensemble model** to predict the most fitting company given a student's experience.

**Best Selling Mangas Bubble Chart** Mar 2017

- Built a **bubble chart** with data from Wiki and external images from Mangapark using **D3.js** and Python.

## ACTIVITIES AND VOLUNTEER EXPERIENCE

---

**Future Blue Lead, IBM** Jan 2017 - Apr 2017

- Co-Organized bi-monthly career development workshops and social events for 30 interns.

**Dept. of Statistics Representative, Women in Math Committee** Oct 2015 - Aug 2017

- Represented female students in the Statistics program during committee meetings.

**Student Life Award 2015, MarianArtist Club** Sept 2014 - Apr 2015

- Co-Directed a Christmas bazaar and fundraised 1000\$ for Nepal earthquake victims.

## EDUCATION

---

**University of Waterloo, Canada** 2015 - 2019 (Expected)  
Honours B. Math., Joint Major in Computer Science and Statistics  
Coursework: Object-Oriented Programming, Data Structures and Algorithms, Forecasting, Applied Linear Models, General Linear Models, Intro to Combinatorics.