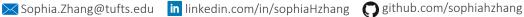
# Sophia Zhang

Data science and visualization expert who solves real world challenges with analytics.









# **SUMMARY**

Quantitative critical thinker with 2 years of statistical data science and 8 years of engineering experience. Proficient in Python, R, SQL, Tableau, Git & Matlab. Experience in data visualization, communication, cloud platforms (AWS), and machine learning.

# DATA SCIENCE PROJECTS



#### Avocado Price Prediction

Applied linear and logistic regression, principal component analysis (PCA), and K-nearest neighbors (KNN) algorithms in Python Scikit**learn** to predict US market price trends.



## **Airbnb Listing Price Prediction**

Predicted Cambridge Airbnb prices using multivariate regression in R



## **Heart Attack Survival Analysis**

**Bioinformatics** model in R using Kaplan-Meier estimator (probability of survival). Applied machine learning algorithms: logistic regression, KNN, and random forest



## Natural Language Patterns

Bayesian predictive model in R JAGS for human reaction time and classification of conversational pauses



#### **Factors for Heart Disease**

Exploratory data analysis in Python analyzing patient demographics, cholesterol, blood pressure, and medical ECG readings.

# **EDUCATION**



# Master of Science in Data Analytics **Tufts University** | Medford, MA, USA

- **GPA 4.0**
- Expected graduation: September 2021



# MicroMaster in Statistics & Data Science Massachusetts Institute of Technology

Expected completion: July 2021



Bachelor of Science in Chemical Engineering, 2012 **University of Alberta** | Edmonton, Canada

## **EXPERIENCE**



## **Process Engineer**

Bantrel | Edmonton, CAN | Feb 2017 - Sep 2020

- Forecasted equipment performance and probability of failure using regression analysis
- Modeled physical processes using technical drawings, equipment, reports, and databases
- Eliminated \$10 million in infrastructure costs by analyzing time series pressure data in a pipeline system, identifying risk patterns and developing solutions to mitigate overpressure



## **Project Engineer**

CEPro Energy & Environment | Edmonton, CAN | Sep 2016 – Feb 2017

- Presented clear visual analysis and saved company unprofitable entry into natural gas market



## **Process Engineer-In-Training**

Stantec Stantec | Edmonton, CAN | May 2012 - May 2015

- Predicted consequences of pipeline risk scenarios based on time series and non-linear data
- Reduced millions in infrastructure cost by proposing simple changes to control and operation

## **Process Engineering Coop**

Stantec | Edmonton, CAN | May 2011 - Dec 2011

- Presented comparative visualizations for efficiency of industrial separation technologies



## **Engineering Coop**

KemeX | Calgary, CAN | Jan 2010 - Aug 2010

- Developed a desalination seawater properties model using regression and correlation analysis



## Pipeline Integrity Engineering Coop

TC Energy | Calgary, CAN | Jan 2009 - May 2009

- Reduced \$3 million in expenditures by building a database that identifies redundant repairs

## **LEADERSHIP**

APEGA | Edmonton Branch Executive Chair 2020

(Sponsorship Chair 2019 | Mentorship Chair 2018)

Manage volunteer operations and outreach to 14,919 professional engineers, geoscientists, and 1,067 corporate members.