SOPHIA ZHANG
DATA SCIENTIST

✉ sozhang89@gmail.com   **in** LinkedIn: [sophiaHzhang](#)   **GitHub:** [sophiaHzhang](#)   🌐 [sophiaHzhang.github.io](#)

EXPERIENCE

SEP 2021 – NOV 2021
Data Science Fellow | Correlation One Data Science Program
- Led group of 5 data scientists to apply machine learning to predict household electric vehicle adoption for marketing and policy initiatives
- Mined travel data, built and tested classification models (regression, decision trees) using Python to achieve 96% predictive accuracy
- Deployed a Tableau dashboard to visualize EV penetration rates, energy costs, and charging station density by geographic location

MAY 2021 – AUG 2021
Data Scientist (Intern) | GlaxoSmithKline
- Developed a scalable data generator and data pipeline for processing laboratory data which saves $50k/year in software fees
- Integrated multiple Python scientific packages and Oracle database access for extraction of large-scale pharmacology and chemistry data
- Automated high volume mass spectra analysis via a TensorFlow neural network and reduced analysis time for chemists by 50%

FEB 2017 – AUG 2020
Applied Scientist / Process Engineer | Bantrel
- Optimized production for multinational energy companies ($20BN+ revenue) by applying data driven recommendations across cross-functional teams of engineers, scientists, and leadership
- Restored 100% target capacity on a utility network by implementing a distributed operations model and avoided $1 million loss in production
- Extracted data from historical logs and technical documents to effectively forecast equipment performance and predict repairs

SEP 2016 – FEB 2017
Business Analyst | CEPro Energy & Environment
- Evaluated strategy for small/medium oil & energy outfits to reach sustainability targets and access $30 million in regulatory credits
- Identified database of 8,000 clients for marketing of environmental services by analyzing company operation metrics and market trends

MAY 2012 – MAY 2015
Applied Scientist / Process Engineer | Stantec
- Eliminated $10 million in infrastructure costs by analyzing time series and non-linear data to model industrial processes in a pipeline system and develop solutions to mitigate risk

EDUCATION

MSc. Data Analytics
Tufts University | GPA 4.0/4.0
Aug 2021

MMs. Statistics & Data Science
Massachusetts Institute of Technology MITx
Aug 2021

BSc. Chemical Engineering
University of Alberta
Jun 2012

SKILLS

Programming
- Python, R, SQL

Data Science
- Regression, Classification, NLP, Neural Networks, A/B Testing

Data Visualization
- matplotlib, seaborn, plotly, Tableau

Tools
- MongoDB, Spark, TensorFlow, Git, AWS