

# STEPHEN RO, Ph.D.

stephensro@gmail.com | [GitHub](#) | [Personal Website](#) | [LinkedIn](#)

Analytically minded and independent starter with 15 years of research experience collaborating with science researchers, computer scientists, and engineers around data, computing, and astrophysics.

## Projects

### The Data Incubator Capstone Project

- Designed and created a wine app recommending best-priced wines using user reviews and price-sentiment analysis: [OVINO](#). Selected as the highest quality DS projects amongst 50+ students.
- Scraped 31 million wine reviews, isolated 150k with prices, curated 6k price-sentiment training set.
- Trained price-sentiment models using sentence transformers (Sentence-BERT), TF-IDF, and star ratings. Deployed XGBoost model with 0.90 PR AUC to my wine app.
- Technologies used: BERT, Scikit-learn, Pandas, SQL, GitHub Actions, Flask, Render, spaCy, Google API.

### LCBO Scraper Bot

- Wrote GitHub Actions workflows that scrapes live inventories of 13.6k alcohol products across 800 LCBO locations. Automatically cleans data, uploads to my SQL database, and updates my wine app.

### The Data Incubator Certification Projects

September 2022 - March 2023

- Stack Overflow veteran-user classifier (PySpark + SQL + ML); NYC cuisine-borough violations (SQL); CNNs and transfer learning an inception network (TensorFlow); Polarizing term identifier in Yelp reviews (NLP & Naive Bayes); time-series temperature, and Yelp rating predictions (random forest, linear/SGD regressors, FFTs);

## Experience

### UC Berkeley, Astronomy Department

July 2020 - August 2020

#### *Course Instructor: Introduction to Scientific Python for Astronomers*

Berkeley, CA

- Designed and taught a python course for astronomy projects. Topics: large data, time-series, numerical simulations, good visuals, scientific writing, and their applications to astrophysics

### UC Berkeley, Theoretical Astrophysics Center

September 2017 - August 2020

#### *Postdoctoral Research Fellow*

Berkeley, CA

- Reduced acoustic-shock prediction errors by 50x across 3-orders of magnitude in param space using time-dependent analyses and mathematical models.
- Accelerated bottlenecked evolution calculations by 10,000x in speed using mathematical models.
- Discovered and analyzed a serious failed test case in gold standard astrophysics code used worldwide. Consulted developers on reproducing the failure and developing solutions.

### University of Toronto, Department of Astronomy & Astrophysics

September 2011 - August 2017

#### *Doctoral Candidate*

Toronto, Canada

- Wrote and designed a telescope image synthesizer that uses real-world statistics from instruments and physics. Used by 4 graduate students for world-class instrument proposals.
- Wrote a relaxation solver to model stellar winds (Python, FORTRAN, C++). Analyzed thousands of stellar wind solutions and extracted parameters to disprove leading ideas on stellar surfaces.

## Education

Data Scientist Certification | The Data Incubator Fellowship Program | March 2023

PhD in Astrophysics focus on Computational Fluid Dynamics | University of Toronto | November 2017

BSc in Mathematical Physics | Queen's University | May 2011

## Technical Skills

**Data Science:** Supervised Learning (Linear | Logistic | KNN | Trees | SVM | Naive Bayes) | Unsupervised Learning (K-means | HDBSCAN | UMAP | PCA) | Big Data (PySpark) | Natural Language Processing | Processing & Visualization (Pandas | Matplotlib | Flask) | Analysis (NumPy | SciPy)

**Programming Languages:** Python | SQL | Bash | YAML | C++ | FORTRAN | HTML / CSS

**Industry Knowledge:** Mathematical Modeling | Statistical Analysis | Predictive Modeling | Quantitative Analysis | Data Management | Data Collection | Dynamical Systems | Lecturing, Teaching & Mentoring | Communication at all levels | Fluid Dynamics | Course & Curriculum Design