|  |
| --- |
| Summary |
| Data scientist with a talent for identifying critical questions and efficiently solving large-scale problems. Analytical experience in every stage of data collection and wrangling large datasets, identifying crucial features, modeling, and presenting results to expert and non-expert audiences. Excels in collaborative team environments. |
| Skills |  | Expertise |
| Machine Learning | MLR | GLR | PCA | Regularization Random Forest | Boosting | scikit-learn | Team leadershipWritten communicationPresentationIndependent workExternal collaboration |
| Visualization | ggplot | matplotlib | seaborn | shiny |
| Data Wrangling | scrapy | selenium | dplyr |
| Languages | Python | R | SQL | LaTeX | MATLAB |
| Education |
| **Certificate,** Data Science | NYC Data Science Academy | 2018 |
| **PhD,** Chemistry | Brown University | 2016 |
| **BS,** Chemistry &French | University of Connecticut | 2010 |

|  |
| --- |
| Experience |
| Data Science Fellow | NYC Data Science Academy | 2018 |
| Immersive 12-week program covering topics in R, Python, SQL, Spark, and Hadoop* [**Data visualization**](http://jamesbudarz.com/2018/07/30/american-pollution/)**:** Designed an interactive Shiny web app mapping air pollution in U.S. cities, annual and seasonal trends, hazard levels, and analysis of population metrics
* [**Web scraping**](https://github.com/jimbudarz/sofariders-profile-analysis): Uncovered fundamental problems for the growth of a popular social network based on aggregated demographics and usage habits
* [**Machine learning**](https://github.com/jimbudarz/MST-housing-price-prediction): Predicted housing prices using thorough feature engineering and application of multiple linear regression and gradient descent boosting. Created insights into potential advantages for both buyers and sellers
 |
| Postdoctoral Fellow | EPFL Lausanne & PSI, Switzerland | 2016 - 2018 |
| Led international teams of between 4 and 12 doctoral students and postdocs in over a dozen collaborations in Switzerland, Italy, Germany, and the United States.* Wrote data acquisition pipeline to process raw instrumental data to suitable data for modeling, leading to 6 papers currently in preparation by graduate students
* Saved hundreds of thousands of dollars in lab access by automating hardware control
* Communicated results in seminars and international conferences
 |
| Research AssistantTeaching Assistant | Brown University &Stanford Linear Accelerator Center | 2010 - 2016 |
| Produced the world’s first [Molecular Movie](https://www.youtube.com/watch?v=HhmBKd-_sRM) from 100TB of molecular images (Press coverage in [Engadget](https://www.engadget.com/2015/06/26/crazy-fast-x-ray-laser-catches-chemical-reactions-in-the-act/), [Nature](http://www.nature.com/nature/journal/v522/n7557/full/522395a.html), and [Physics](https://physics.aps.org/articles/v8/59))* Developed protocol for data cleaning to remove artifacts (dark levels, background levels, nonlinear response, and electronic cross-talk) from an experimental pixel array detector
* Developed pipeline for filtering and sorting desirable images based on metadata
* Wrote generalized additive model to fit physical phenomena
* Designed, constructed, and fully implemented a novel scientific instrument
* Published cover article of top physics journal [*Physical Review Letters*](https://journals.aps.org/prl/abstract/10.1103/PhysRevLett.114.255501) describing the experiment and results and a methodological paper detailing the data-processing routines
* Presented findings and methods in international seminars and conference talks
 |