

João Moreira

✉ joao.ag.moreira@gmail.com | 🏠 jagmoreira.com | 📍 jagmoreira | 🌐 joao-moreira

Summary

Data scientist with a strong statistics background and experience in big data, machine learning, and data visualization. Passionate about explaining data science to non-technical audiences. Strong critical thinker who always welcomes a challenge.

Education

- | | |
|--|-----------|
| Ph.D. Chemical Engineering, Northwestern University
Thesis: <i>The Effect of Gender Diversity in Creative Teams</i> | 2012–2017 |
| M.Sc. Physics, University of Lisbon, Portugal
Thesis: <i>Evolutionary Dynamics of Cooperation in Multiplayer Games</i> | 2009–2011 |
| B.Sc. Physics, University of Lisbon, Portugal | 2006–2009 |

Skills

- Machine Learning** regression, classification, clustering, neural networks, data cleansing, feature engineering
- Statistical Modeling** hypothesis testing, time-series analysis, cross-validation, regularization, bootstrapping, MapReduce
- Programming** Python (numpy, scipy, pandas, scikit-learn, matplotlib, requests), R, BASH, LaTeX
- Tools** Jupyter notebook, Sublime Text, Git, Mercurial, Excel
- Databases** MySQL, MongoDB
- Web Design** HTML, CSS, Javascript, D3.js

Experience

Data Scientist and Instructor

Northwestern University

Jan. 2012–Sep. 2017

- Used maximum likelihood estimation to parameterize the distribution of the impact of scientific publications (~1TB dataset).
- Constructed a Django web application with interactive D3.js visualizations to showcase my data-driven ranking of scientific research: <http://forecite.us>.
- Demonstrated the existence of historical gender discrimination in the movie industry by performing time-series analysis on IMDb metadata of +15k movies.
- Created and delivered original course materials on web-scraping and object-oriented programming to over 100 students in 2 sessions of a 1-week long Python programming bootcamp.
- Organized and led discussion groups, and graded homeworks for all students in 4 undergraduate courses.
- Mentored 2 younger graduate students and 2 high-school interns on the principles of scientific research and basics of programming.

Data Science Intern

GadflyZone

Apr. 2016–Sep. 2016

- Constructed MySQL database of patents on Amazon Database Relational Services built from publicly available patent information.
- Used natural language processing and citation analysis to design a multidimensional scoring engine to predict a company's innovativeness.

Software Engineer

Narrative Science

Jun. 2014–Sep. 2014

- Developed system in agile environment to automatically analyze performance of the company's artificial intelligence platform.