



# Thiago Marzagão

Data Scientist

+55-61-99645-0448

thiago.marzagao@gmail.com

<http://thiagomarzagao.com>

<https://github.com/thiagomarzagao>

## Profile

I'm a data scientist with a strong statistics background and seven years of experience with impactful machine learning projects. I'm full stack: I can collect the data (scrape it or query it from SQL and NoSQL databases), clean it up, make sense of it, use it to train a predictive model, assess the model's performance, present the results to non-technical audiences, apply the model.

## Technical Skills

machine learning	● ● ● ● ●
statistics	● ● ● ● ●
Python	● ● ● ● ●
R	● ● ● ● ●
SQL	● ● ● ● ●
web development	● ● ● ● ●
AWS	● ● ● ● ●

## Recent Experience

### Data Scientist, Comptroller-General of Brazil

2016 - Present

- I trained a predictive model that helps the Comptroller-General of Brazil choose which municipalities it will audit each year. The model impacts which instances of corruption will be discovered and hence which mayors and government employees will go to jail, how much taxpayers' money will be recovered, and which local policies will be discontinued. The model was introduced in 2017. The total auditable budget was US\$ 3.6 billion in 2018 alone.
- I have worked in several other projects at the OPS, like a model that predicts which government suppliers will fail to deliver and a bot that automates the background check of new government employees.

### Data Scientist, Brazil's antitrust agency

2014 - 2016

- I developed and maintained CEREBRO, an app that mines government procurement data looking for evidence of bid rigging. CEREBRO uses a mix of machine learning, time series econometrics, and graph theory. CEREBRO screens all non-classified federal government bidding processes in Brazil. It was introduced in 2015. The total auditable budget was US\$ 5.9 billion in 2018 alone.

### Adjunct Professor, several universities

2016 - 2018

- I've taught stats and machine learning at a number of private and public universities in Brasília: UnB (machine learning; graduate level), IESB (machine learning; graduate level), IDP (statistics; undergraduate level).

## Education & Research

### Ph.D. in Political Science

2010 - 2014

#### Ohio State University

I took lots of stats classes and in my dissertation research I used natural language processing and machine learning to create the first machine-coded democracy index.

### Online Coursework

*Udacity: Nanodegree Deep Reinforcement Learning (2018). Coursera: Deep Learning Specialization (2018), Machine Learning (2013), Introduction to Data Science (2013). edX: Introduction to Apache Spark (2016).*

### Peer-Reviewed Publications

See <http://thiagomarzagao.com/publications/>