Jorge Cimentada - 697 382 009 - Madrid - cimentadaj@gmail.com

@linkedin: Jorge Cimentada | @website: cimentadaj.github.io | @github: github.com/cimentadaj | @Twitter: @cimentadaj

PROFILE

Data scientist with background in statistics, predictive analytics and machine learning. Experience working with causal inference,
big data and R programming. Passionate about teaching R, data science and explaining my research to non-technical audiences.
For examples of my work, make sure to visit my website: cimentadaj.github.io

EDUCATION

Pompeu Fabra University   Jan 2016 - Present
PhD Candidate Quantitative Sociology - Expected January 2019

My research involved collapsing data on over **3 million students** to study educational policies and its impact on inequality.
The key takeaway is that inequality depends heavily on the level of vocational education of the country. **Awarded best PhD paper in 2017**

Pompeu Fabra University   Sep 2014 - Sep 2015
MSc Sociology and Demography
Best master's thesis: First of class of 25. Featured in master's website.

PUCMM University   Sep 2009 - Sep 2013
B.A. in Business

WORK EXPERIENCE

Data Scientist: Kernel Analytics   July 2018 -
[R, Python, Impala, Hadoop, Airflow]
Working full time as a Data Scientist at Kernel Analytics, among the best consultancy firms specialized in advanced analytics.
My work involves querying 500GB datasets with Impala, Hadoop and Hive and applying state of the art algorithms to
understand complex patterns with Python and R. Currently, I predict customer churn and client segmentation.

Data Scientist: Research and Expertise Centre for Survey Methodology (RECSM)   Jan 2018 - July 2018
[R, MySQL, Web scraping, API access]
This project involved designing and implementing two web applications (based on Shiny apps). I applied **text mining** techniques
to extract and translate text using the Google Translate API. I created a library that **reimplemented four algorithms** for
adjusting **model predictions** for measurement error.

Data Scientist: Pompeu Fabra University   Feb 2016 - Jan 2018
[R, Rmarkdown, LaTeX]
I exercised the complete research pipeline of accessing **noisy data**, cleaning, analyzing and reporting. In one case, I used several
classification algorithms to estimate the propensity of mobility between social classes. I compared **Linear Discriminant
Analysis**, **Support Vector Machines** and **Naive Bayes** algorithms and landed on LDA because the predictive classes were stable
and had the lowest out-of-sample error.

International Business Analyst: Trovit   Jul 2015 - Feb 2016
My work was focused on analyzing click-through rates and price-per-clicks to create a better strategy for different country
managers. My analysis helped launch the internationalization of 4 new products (+2 million potential each), all decisions using
time-series analysis.

Statistical consultant: Population Services International   Jul - Aug - 2014
[Causal Inference, Reporting]
I designed, implemented and analyzed the **randomized controlled trial** related to HIV prevention. The project increased HIV
awareness by 7% in the population of armed forces resulting in a **renovation of funding** from the Department of Defense.

Junior data analyst: Newlink Research   2013-2014

TECHNICAL SKILLS

Machine Learning: classification, regression, clustering, feature engineering
Statistical Methods: regression models, hypothesis testing and confidence intervals, principal component analysis and dimensionality reduction, text mining, multilevel models, propensity score matching.
Software and Programming Languages: R (+ 3 years, dplyr, ggplot2, Rmarkdown, caret, tidyverse), Python (numpy, pandas), SQL, Impala, Hadoop, Airflow, Regular Expressions, Git, Bash, Docker, Linux, Microsoft Excel, LaTeX

MY SOFTWARE

Github ranking: R (http://git-awards.com/users/cimentadaj)
3rd out of 126 in Madrid
13th out of 482 in Spain

R libraries:
opendataes: Interact with the API of datos.gob.es and extract data from all of Spain [github.com/cimentadaj/opendataes]
sqpr: Access over 40,000 question from the SQP API and adjust your estimations for measurement error
essurvey: Download data from the European Social Survey [CRAN, ropensci.github.io/essurvey]
perccalc: Estimate continuous distributions from categorical variables [CRAN, github.com/cimentadaj/perccalc]
cimentadaj: Personal package [github.com/cimentadaj/cimentadaj]
corrr: Exploring correlation in R in a 'tidy' framework. **Contributor** [github.com/drsimonj/corrr]

LANGUAGES

English: fluent
Spanish: native

SELECTED PRESENTATIONS, TEACHINGS AND PROJECTS

 "Introduction to R programming", RECSM Data Science Summer School 2016 [Jul 2018]
"Introduction to the tidyverse and functional programming", UPF PhD training, March 2017
"Introduction to the tidyverse" - PyData Conference, May-2017 [invited talk]
"The essurvey package" - ESS conference at the University of Southampton, January 2018 [invited talk]

SELECTED INDEPENDENT PROJECTS
See at cimentadaj.github.io/projects/

* Scraping and analyzing the scripts of How I Met Your Mother
* Analyzing bicycle waiting time in Barcelona
* How I programmed a Twitter bot to clean and visualize data
* Web shiny application to visualize educational inequality for over 70 countries in 15 years
* All solutions to the R for Data Science book

Or check out my blog for shorter posts on machine learning and data analysis at cimentadaj.github.io/blog/. Some example posts:

* Brief Analysis of Independent/Unionist Vote in Catalonia
* Hold and Cross-validation series of posts