JACQUELINE ARAYA

DATA SCIENTIST

London, UK 🏠 | 077-6131-4054 📞 | ja3076@columbia.edu ✉ | jacquelinearaya.github.io 🌐 | arayajacqueline 💼

A researcher-turned data scientist eager to transform my experience in applied research into actionable methods to tackle business problems using data as the vehicle to get insights, predictions and recommendations that speak to every business stakeholder

---

EDUCATION

M.Sc. in Data Science | Columbia University
NEW YORK, USA | SEPT 2018 – MAY 2020
Capstone Project: "Asset tracking: identifying patterns in spatio-temporal trajectory datasets" in collaboration with GE Research Centre

Professional Engineer - Industrial Engineering | University of Chile
SANTIAGO, CHILE | JUL 2013 – MAR 2015
Thesis: "Analysis of scientific virtual communities of practice"

BSc. in Engineering Science | University of Chile
SANTIAGO, CHILE | MAR 2008 – JUN 2013

SKILLS

Python	Numpy, Pandas, TensorFlow, Keras, Scikit-Learn, Matplotlib, NLTK, PySpark, Web Scraping
R	Data Visualization: ggplot, Data Manipulation: dplyr, tidyr
Documenting	Git, Jupyter, RMarkdown, Latex
Linux	Command line, bash
Others	SQL, BigQuery, Stata, Matlab, Java

PROFESSIONAL EXPERIENCE

Research Associate | Columbia Business School
NEW YORK, USA | AUG 2016 – OCT 2018
Worked with the Management Division faculty over two years in the technical side of applied research projects related to organizational behaviour, entrepreneurship, firm strategy, labour market dynamics, and others.

- Designed and executed automated Python/bash scripts for a field experiment to contact 15,000 job seekers in one of the largest job-seeking online platforms and to track subsequent engagement.
- Data collection of 6.5 million records from one of the major e-sports game, Dota2, using Python to connect to the service API and Linux-bash scripts to parallelize collection.
- Data web scraping of multiple static and non-static websites using Python (Selenium and BeautifulSoup).
- Translated and adjusted Java code into Python code to calculate a variety of similarity measures between artists for a music dataset of 10,000 unique artists and 115,000 songs.
- Analysed survey data of 606 Columbia MBA students using a multinomial logistic regression model to study the effects of different choices when facing job prospects upon graduation.
- Mentored research assistants in the use of Linux-based high-performance servers and helped interns and Ph.D. students to translate research needs into coding implementation.
- Worked closely with the Information Technology Group Director in the promotion of a data-driven culture within the research community of the school by introducing new Linux-based tools.

Project Analyst | Nic Chile Research Labs (Chile's DNS research centre)
SANTIAGO, CHILE | JUN 2012 – MAR 2015

- In charge of project applications for public funding, effectively securing up to $40,000 for one of our major projects.
- Proposed business models for software developed by our engineering team, considering the needs of different business stakeholders.
- Forged a public-private partnership with one of Chile's largest Internet service providers to financially support an app produced by our software engineering team that measures the quality of service of mobile Internet at a national level.

Digital Marketing Intern| Bank of Credit and Investment (BCI)

SANTIAGO, CHILE | SUMMER 2012

- Performed data analysis of customers' overdraft credit behaviour datasets to design more effective marketing campaigns to target companies and entrepreneurs' segments differently, resulting in a 3% lift for the summer email campaign.
- Built a data visualization dashboard of forecasted versus actual sales for the area manager integrating the area's 5 credit products, with weekly maintenance.

DATA SCIENCE ACADEMIC EXPERIENCE

Researcher (Engineering Thesis) | Centre for Business Intelligence (CEINE)

UNIVERSITY OF CHILE, SANTIAGO, CHILE | NOV 2013 – FEB 2015

Quantitative Analysis of scientific virtual communities of practice using topic models and social network analysis.

- Applied natural language processing methods (NLP) to extract and process text data from 4,845 academic articles.
- Estimated probabilistic topic models (LDA) to build a network based on topic similarity that improves network properties and sub-communities detection over co-authorship networks.

Intern | University of Notre Dame (iCeNSA - Interdisciplinary Centre for Network Science and applications)

SOUTH BEND, IN, USA | SUMMER 2014

Worked exploring ways on how to build a network that would feed on multiple economic news and conditions to predict future crises in Latin American countries.

VOLUNTEER EXPERIENCE

Logistics Volunteer | Latino Impact Summit at the United Nations HQ

NEW YORK, USA | DEC 2018

Summit for the progress and growth of America, proposing solutions for the 2030 Sustainable Development Agenda

- Monitored registration process for 400 attendees.
- Assisted nearly 50 speakers to keep track of time during talks.
- Served as a direct contact for attendees during workshops with communication/logistic requests.

Organizing Committee, Co-chair | LatinCrypt (2nd International Conference on Cryptology and Information Security in Latin America)

SANTIAGO, CHILE | OCT 2012

Managed a budget of $21,000 to organize the 3-day conference for 70 attendees (event planning, registration, logistics).

ACADEMIC HONORS

Scholarship for internships abroad for engineers| CONICYT (Commission for scientific and technological research of Chile)

SUMMER 2014

Dean's List Industrial Engineering| University of Chile

SEPT 2012 – SEPT 2013

RELEVANT COURSEWORK

M.Sc. in Data Science	Algorithms for Data Science, Statistical Inference, Machine Learning, Deep Learning, Data visualization, Computational systems and Databases, Product Recommendation, Causal Inference
Professional Engineer - Industrial Engineering	Statistics for Business and Economics, Applied Econometrics for Business and Economics, Operations Management, Marketing Strategy, Marketing Analytics, Consumer Behaviour, Financial Analysis, Financial Engineering, Data Mining, Data Warehousing, Industrial Organization, Information and Communication Technologies for Management
BSc. in Engineering Science	Single and Multivariate Calculus, Linear Algebra, Ordinary Differential Equations, Computer Science, Electromagnetism, Probability and Statistics, Modelling and Optimization, Operations Research, Microeconomics, Macroeconomics