

CARLOS A. HARO

Data scientist with experience in reproducible pipelines for data products. Specialized in visualizations, but with knowledge of cloud and on-premise data engineering practices.



EXPERIENCE

Jan. 2019
|
present

Sr. Data Scientist

Mexico's Central Tax Administration Office
(Servicio de Administración Tributaria)

📍 Mexico City

- Development of a **supervised model** for identification of tax evasion schemes using **cloud architecture** | *Spark (pySpark, koalas, SQL), python (scikit, pandas)*
- Development and orchestration of a **bayesian model** for predicting aggregate employment using tax receipts on **cloud architecture** | *R (tidyverse, STAN), Python (scikit-learn, pandas), Spark (SQL)*
- Development and deployment of a **supervised model** for classifying tax debt | *R (tidyverse, ggplot2), GNU Make*
- Designed and taught three courses for the institution's **staff training**: intro to R, basics of EDA, building DS pipelines with Make | *Rmarkdown (xaringan for slides)*
- Designed a **pipeline for automatic** generation of frequent data visualization reports | *R (tidyverse, visnetwork, ggraph)*

2018

Jr. Data Scientist | Economic Analyst

EnergeA (Energy Sector Consulting Firm)

📍 Mexico City

- Developed a **statistical model** for identifying anti-competitive practices between the mid-stream natural gas providers | *base R*



MISCELLANEOUS

2020

Cloud certifications

Microsoft certified in Azure fundamentals (certification badge)

2020
|
present

Podcast host

Co-founder and host in "Cuestion de datos", a data related podcast. (Spotify link)

2020

Undergraduate teaching

Taught 3 undergraduate courses (*pro bono*)

The **objective** was to teach political science, management, and economics students about reproducible workflows associated with modern computational techniques for data analysis and modelling. Taught in **spanish** for **ITAM students**; software included R, python, SQL, and git through RStudio, VScode, jupyter, and the terminal.

Course videos, syllabus and materials: (1st course), (2nd course), (3rd course)

2020
|
2021

Graduate teaching

Taught one course to economics master's degree students

Content was similar to the undergraduate one, however, special emphasis was made into developing a toolbox for reproducible academic research.

Course videos, syllabus and materials: (Dropbox link)



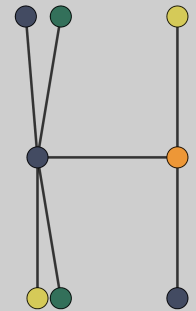
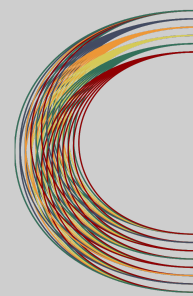
EDUCATION

2014
|
2018

(ITAM) Instituto Tecnológico Autónomo de México

B.S., Economics

📍 Mexico City, Mexico



CONTACT INFO

✉ haro_ca@outlook.com

🐦 @haro_ca_

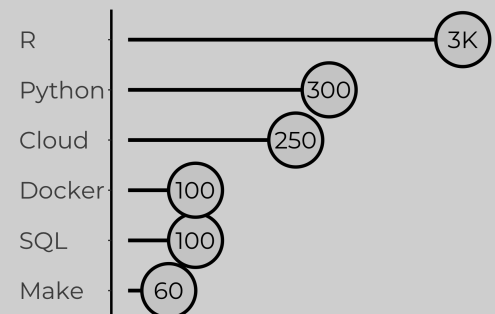
📞 haro-ca

📞 (+52) 55-49-85-53-34

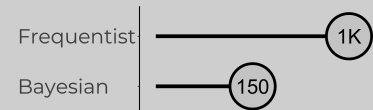
SKILLS

Outliers, by Malcolm Gladwell, states that mastery over a skill can only be achieved by 10,000 hours of productive work. I find it a good proxy for maturity of a skillset, so I summarize my current state in the charts below:

Programming (aprox. hours of work)



Statistics (aprox. hours of work)



Main experience with frequentist methods, but with high interest to develop in bayesian's.

Fun facts

Vim enthusiast; in the terminal, in VScode, in RStudio, in Chrome (with Vimium), even in the whole OS! (with Karabiner)

Chess fanatic (chess.com profile)

Salsa dancer (though not while COVID)

*This resume was made entirely with R. Raw format with **pagedown**, logo with **ggraph**.*

*The source code is available at my **personal repo** and, obviously, it can be used freely.*

Last updated on 2021-01-25.