Curriculum Vitae

Luís Assunção

Email / GitHub / LinkedIn / Website

Employment

Hotmart

Staff Data Scientist | April 2020 - present

- Developed an in-house AB hierarchical testing framework with optional stopping
- Consulted for and developed randomized controlled trials
- Estimated causal effects in non-randomized experiments
- Estimated pricing elasticity for digital products using multilevel models
- Classified evergreen vs launching sales strategies using hidden state models
- Improved quality of course assigments using Item Response Theory models

Oper

Data Scientist | Oct 2018 - March 2020

- Consulted for companies such as AB InBev and GTB in statistical projects
- Modeled spatial pricing elasticity for beverages using Gaussian Processes
- Estimated revenue attribution in multi-touchpoint marketing campaigns

IRIS

Intern | 2015 - 2017

- Collected, wrangled and described survey data
- Researched policies to advance human rights in the digital matters

Education

B.S in Statistics

Federal University of Minas Gerais (UFMG) | Belo Horizonte, MG - Brazil | 2017 - 2021

- Researched and authored a reproducible monograph (in portuguese with an abstract in english) on exponential random graphs applied to epidemiology
- Co-authored Frequency and burden of neurological manifestations upon hospital presentation in COVID-19 patients: Findings from a large Brazilian cohort

Examples

Blog

Posts on data analysis using tools such as Python, `polars`, `pymc`, `pulp`, `seaborn`:

- Picking a fantasy football team: In this post, I delve into the data for the 2022 season of a brazilian fantasy football league; formulate a mixed integer linear program to pick the optimal team; and present initial concepts for forecasting player scores using mixed effects linear models.
- Decomposable non-monotonic models: In this post, I compare empirical and parametric approaches to model non- monotonic relationships using a Digit Span verbal working memory cognitive test dataset.

Repositories

- site: My website and blog post codes using Quarto Markdown
- mldc2020: Recommendation system and 7th place solution to the Mercado Libre Data Challenge 2020
- rstanbtm: Biterm Topic Model implementation in Stan
- qlm: Generate predictive SQL queries from linear models in R
- tophat: Scheduled shell script to fetch and save fantasy football data

Others

- Pod e Dev podcast episode, where I talk (in portuguese) about the challenges in pricing digital products and causal assumptions we made to overcome these challenges in our model at Hotmart. We also discuss good and bad use cases for large language models, as well as how models with 2 parameters can be as useful as models with 200 million parameters.