

David Damián Arbeu
(+52) 771-183-8002 | david.mathdatasc@gmail.com | [LinkedIn](#) | [GitHub](#)

SUMMARY

As a dedicated and experienced data science professional I am committed to the belief that together we can create better AI and data science solutions that could help enhance people's lives and help organizations achieve greater efficiency and success. Experienced in delivering end-to-end data science solutions — from business problem understanding to data engineering, modeling, and scalable deployment — driving strategic decisions and optimizing KPIs across the retail and fintech industries.

EDUCATION

- **Applied Deep Reinforcement Learning Diploma:** Diploma program in *Applied Deep Reinforcement Learning* at [Colegio Bourbaki](#). Certification available [here](#).
- **Instituto Tecnológico Autónomo de México (ITAM). México, CDMX**
M. Sc. in Data Science - GPA 9.6/10. Aug. 2022 - Aug 2024
Thesis: Computational uncertainty in scalable probabilistic machine learning.
- **Universidad Autónoma del Estado de Hidalgo (UAEH). México, CDMX**
BASc. in Applied Mathematics - GPA 9.59/10. Aug. 2016 - Dec 2021
Thesis: Causal effects of climate change on Mexico's wellbeing indicators.

EXPERIENCE

RappiCard México. México City, MX. (September 2024 – actual)

Sr. Data Scientist → Lead Data Scientist

- **Promoted as Lead Data Scientist:** Took technical ownership of critical projects with greater autonomy in decision-making and initiative prioritization, while mentoring two junior Data Scientists and providing peer guidance to strengthen team delivery.
- **Customer Segmentation for Profitability Policies:** Led data-driven profitability strategies by developing unsupervised ML segmentation using behavioral and transactional and credit risk. Deployed models on AWS SageMaker for scalable inference and built a GenAI agent to generate business-focused segment reports.
- **Credit Line Increases data driven:** Drove a \$4M MXN revenue increase in 6 months for each data driven credit line increase (CLI) campaign. Led randomized control trials and applied causal inference methods like ITS and DiD to rigorously measure impact.
- **LTV model:** Developed and deployed an ML model on AWS to predict credit card customer Lifetime Value, improving forecasting accuracy by 10% and supporting Growth and Strategic Finance teams decisions.
- **Merchant Categorization Project:** Deployed scalable pipeline to improve merchant categorization by combining DENUE data, browserless Google searches, and LLMs for classification and reclassification. Validated through an A/B test in a growth discount campaign, the system reduced manual effort, improved MCC accuracy, and enhanced targeting precision, creating a framework for future classification challenges.
- **Credit Line primacy causal study:** Spearheaded a causal analysis quantifying the impact of credit line primacy across the portfolio. The findings, derived from a Double Machine Learning model, directly informed a revamped credit line increase framework to strategically defend and capture primary lender status.
- **Portfolio Risk Models:** Powered data-driven credit line increase campaigns, scalable client risk segmentation, and strategic decision-making across the organization developing SOTA deep learning models for tabular data to assess risk, achieving AUCs of up to 0.90.
- **Led regular executive briefings to communicate results and actionable insights to stakeholders and the CEO.**

PhiQus Data Science. México City, MX. (January 2024 – May 2024)

Data Scientist Consultant

- **Contributed to the founding of PhiQus, formed by the Data Science team at the company I previously worked with.**
- **Soriana's Data-Driven Delivery System:** Designed and deployed Soriana's first data-driven route optimization algorithm using Azure Functions, reducing delivery time and distance for online orders. Conducted fieldwork to evaluate impact and enabled scalable, data-informed logistics.
- **Time Series Forecasting for Retail Sales Optimization:** Secured Home Depot México as a client through a forecasting PoC that combined ARIMA and machine learning, demonstrating measurable business value.

- **Promotion effectiveness analysis for Soriana:** Applied quasi-experimental methods like difference-in-differences to transaction and inventory data to quantify the causal impact of promotional periods, identifying a ~12% average sales uplift and reducing ineffective campaigns by 25%.
- **Data Science Documentation Standards:** Initiated the first culture of documentation within the company by designing and leading the implementation of a comprehensive data science project framework—bridging business context and technical code. This initiative significantly improved code development effectiveness, fostered cross-team collaboration, ensured reproducibility..
- **Conducting regular follow-up sessions to present results and main insights to stakeholders.**

ENKI, México City, MX. (October 2023 – December 2023)

Jr Data Scientist Consultant

- **Sales forecasting:** Improved the accuracy of biweekly online sales forecasts by combining ARIMA models with machine learning techniques, achieving a 20% average reduction in RMSE and enhancing Found Rate and the efficiency of recommendation algorithms in production.
- **Data Analysis:** Conducted data analysis on online delivery orders at Soriana, leading to the development of a sales forecasting model for products sold online.
- **Conducting regular follow-up sessions to present results and main insights to stakeholders**

PERSONAL AI LAB

LLM fine tuning and calibration via RLHF: Fine-tuned a LLaMA model with LoRA and calibrated its behavior using a reinforcement Learning from Human feedback (RLHF) agent for deceptive text detection.

Recommendation System via Offline Reinforcement Learning: Built an end-to-end movie recommendation engine using an Actor-Critic agent trained on historical user data. Leveraged PyTorch and offline RL techniques to optimize long-term user engagement; evaluated performance with precision@k on a 100k+ rating dataset.

AWARDS

- **Full Scholarship to study Data Science at ITAM:** Covering 100% of tuition and fees for Master of Data Science program based on academic excellence.
- **Mérito Académico:** Top graduate at UAEH, awarded for academic excellence with the highest GPA in my cohort.
- **Mérito Garza:** Recognized for excellence in the General Exam for Undergraduate Graduates (EGEL).

COURSES AND CERTIFICATIONS

- **Generative AI Nanodegree – Udacity:** On going.
- **Machine Learning Engineering at AWS Nanodegree – Udacity:** On going
- **Agentic AI Nanodegree – Udacity:** On going.
- **Machine Learning in Production – DeepLearning:** Certification available [here](#).
- **Spark and Datalakes – Udacity:** Certification available [here](#).
 - Designed and implemented a scalable lakehouse architecture on AWS using AWS Glue, Amazon S3, PySpark, and Python to process and curate semi-structured data for data science workflows as part of an end-to-end data engineering solution. Project repository available at this [link](#).

SKILLS

- **Cloud and big data:** AWS (SageMaker, S3, Lambda, Step Functions, Athena, EC2, Glue), Azure (Functions and Azure DevOps), Snowflake.
- **Data Science:** Python, R Software, SQL.
- **AI frameworks:** ScikitLearn, econML, XGBoost, Hugging Face, PyTorch.
- **Business Intelligence:** Shiny and Streamlit.
- **Bayesian Inference:** Stan.