The Journal of Open Source Software

# PyMC-Marketing: Bayesian Marketing Mix Models and Customer Analytics in Python

**William Dean** [1], **Juan Orduz** [1], **Colt Allen**[1], **Carlos Trujillo**[1], **Ricardo Vieira**[1], **Benjamin T. Vincent**[1], and **Thomas Wiecki**[1]

**1** PyMC Labs

**DOI:** 10.xxxxxx/draft

**Software**

- Review
- Repository
- Archive

**Editor:** Open Journals

**Reviewers:**

- @openjournals

**Submitted:** 01 January 1970
**Published:** unpublished



## Summary

PyMC-Marketing is a comprehensive Python library implementing Bayesian marketing analytics, built on PyMC (Salvatier et al., 2016). Commercial marketing analytics tools typically provide limited transparency into their models, while open-source alternatives like Meta's Robyn and Google's Meridian focus primarily on media mix modeling (Google Inc., 2023; Meta Platforms Inc., 2022). PyMC-Marketing provides a unified framework spanning multiple marketing domains, including: Media Mix Modeling, Customer Lifetime Value analysis, Bass Diffusion Models, and Customer Choice Models. All outputs include full posterior distributions rather than point estimates, enabling explicit risk assessment in business decisions.

## Statement of Need

Marketing organizations struggle to attribute sales outcomes to specific marketing activities across multiple touchpoints and delayed conversion effects. Existing solutions suffer from: (1) black-box proprietary models with limited customization; (2) oversimplified approaches failing to capture marketing dynamics; and (3) lack of uncertainty quantification for high-stakes decisions.

PyMC-Marketing addresses these gaps by bridging marketing science research and practical applications. It operationalizes advanced Bayesian methods—hierarchical modeling, experimental calibration, and uncertainty quantification—within a user-friendly, scikit-learn compatible API. Key innovations include time-varying coefficients using modern Gaussian process approximations optimized for marketing contexts, and a novel experimental calibration framework that integrates lift test results directly into model likelihood. While frequentist approaches like Robyn provide bootstrap-based intervals, all PyMC-Marketing outputs include full Bayesian posterior distributions, enabling decision-makers to assess risk explicitly.

## Installation and Dependencies

PyMC-Marketing is available via conda-forge and pip. Core dependencies include PyMC ( 5.0), NumPy, Pandas, ArviZ (Kumar et al., 2019), and scikit-learn. Optional dependencies enable GPU acceleration (JAX), advanced samplers (NumPyro, Nutpie), and production deployment (MLflow (Zaharia et al., 2018), Docker).

## Key Features

PyMC-Marketing provides four distinct modules addressing comprehensive marketing analytics:

The Journal of Open Source Software

1. **Media Mix Modeling (MMM)**: Multiple adstock functions, saturation curves, time-varying parameters via HSGP (Solin & Särkkä, 2020), experimental calibration for causal inference, budget optimization with business constraints, time-slice cross-validation, and marginal effects analysis (Arel-Bundock et al., 2024).

2. **Customer Lifetime Value (CLV)**: BTYD models (Fader & Hardie, 2020) including BG/NBD, Pareto/NBD, and Gamma-Gamma frameworks with hierarchical extensions and individual-level uncertainty.

3. **Bass Diffusion Models**: Product adoption forecasting (Bass, 1969) with flexible parameterization for innovation and imitation effects across multiple products.

4. **Customer Choice Models**: Discrete choice analysis (Train, 2009) based on random utility theory, including multinomial logit and multivariate interrupted time series models.

**Production Ready**: All modules feature MLflow (Zaharia et al., 2018) integration, Docker containerization, multiple MCMC backends (NumPyro (Bingham et al., 2019), Nutpie), variational inference (ADVI), MAP estimation, data connectors (e.g., Fivetran), and comprehensive diagnostics via ArviZ.

# Key Advantages

PyMC-Marketing provides uncertainty quantification through full posterior distributions, experimental calibration anchoring observational models to causal ground truth, and flexible budget optimization with business constraints. The scikit-learn compatible API ensures seamless integration into existing data science workflows. The library has been successfully deployed by companies including HelloFresh and Bolt for production marketing analytics. Comprehensive tutorials, example notebooks, and video resources are available in the online documentation at https://www.pymc-marketing.io/en/stable/, with community support from over 70 contributors and translations in Spanish.

# Community Guidelines

- **Issues**: Report bugs and feature requests on GitHub
- **Discussions**: Join community discussions on GitHub Discussions
- **Documentation**: Comprehensive guides and tutorials available at pymc-marketing.io
- **Contributing**: See CONTRIBUTING.md for development guidelines
- **Support**: Professional consulting available through PyMC Labs

# Funding

PyMC-Marketing is a community-driven project built primarily through volunteer contributions from over 70 developers. Some contributions by PyMC Labs affiliates have received partial funding.

# Acknowledgments

We acknowledge the PyMC development team for the foundational probabilistic programming framework and the broader PyData ecosystem contributors. Special recognition goes to the marketing science research community for developing the theoretical frameworks that PyMC-Marketing operationalizes (Bass, 1969; Fader et al., 2005; Jin et al., 2017; Train, 2009).

# References


Arel-Bundock, V., Greifer, N., & Heiss, A. (2024). How to interpret statistical models using marginaleffects for R and Python. *Journal of Statistical Software*, *111*(9), 1–32. https://doi.org/10.18637/jss.v111.i09

Bass, F. M. (1969). A new product growth for model consumer durables. *Management Science*, *15*(5), 215–227. https://doi.org/10.1287/mnsc.15.5.215

Bingham, E., Chen, J. P., Jankowiak, M., Obermeyer, F., Pradhan, N., Karaletsos, T., Singh, R., Szerlip, P. A., Horsfall, P., & Goodman, N. D. (2019). Pyro: Deep universal probabilistic programming. *Journal of Machine Learning Research*, *20*(28), 1–6. http://jmlr.org/papers/v20/18-403.html

Fader, P. S., & Hardie, B. G. (2020). Customer lifetime value: What's next? *Journal of Interactive Marketing*, *49*, 26–37. https://doi.org/10.1016/j.intmar.2019.08.001

Fader, P. S., Hardie, B. G., & Lee, K. L. (2005). Counting your customers: Who are they and what will they do next? *Management Science*, *51*(9), 1371–1387. https://doi.org/10.1287/mnsc.51.9.1371

Google Inc. (2023). *Meridian: Bayesian media mix modeling at scale*. https://github.com/google/meridian. https://github.com/google/meridian

Jin, Y., Wang, Y., Sun, Y., Chan, D., & Koehler, J. (2017). *Bayesian methods for media mix modeling with carryover and shape effects*. Google Inc. https://research.google/pubs/pub46001/

Kumar, R., Carroll, C., Hartikainen, A., & Martin, O. A. (2019). ArviZ a unified library for exploratory analysis of bayesian models in python. *Journal of Open Source Software*, *4*(33), 1143. https://doi.org/10.21105/joss.01143

Meta Platforms Inc. (2022). *Robyn: Marketing mix modeling*. https://github.com/facebookexperimental/Robyn. https://github.com/facebookexperimental/Robyn

Salvatier, J., Wiecki, T. V., & Fonnesbeck, C. (2016). Probabilistic programming in python using PyMC3. *PeerJ Computer Science*, *2*, e55. https://doi.org/10.7717/peerj-cs.55

Solin, A., & Särkkä, S. (2020). Hilbert space methods for reduced-rank gaussian process regression. *Statistics and Computing*, *30*(2), 419–446. https://doi.org/10.1007/s11222-019-09886-w

Train, K. E. (2009). *Discrete choice methods with simulation*. Cambridge University Press. https://doi.org/10.1017/CBO9780511805271

Zaharia, M. A., Chen, A., Davidson, A., Ghodsi, A., Hong, S. A., Konwinski, A., Murching, S., Nykodym, T., Ogilvie, P., Parkhe, M., Xie, F., & Zumar, C. (2018). Accelerating the Machine Learning Lifecycle with MLflow. *IEEE Data Eng. Bull.*, *41*, 39–45. https://api.semanticscholar.org/CorpusID:83459546