

# 1 PyMC-Marketing: Bayesian Marketing Mix Models 2 and Customer Analytics in Python

3 William Dean  <sup>1</sup>, Juan Orduz  <sup>1</sup>, Colt Allen<sup>1</sup>, Carlos Trujillo<sup>1</sup>, Ricardo  
4 Vieira<sup>1</sup>, Benjamin T. Vincent<sup>1</sup>, and Thomas Wiecki<sup>1</sup>

5 1 PyMC Labs

DOI: [10.xxxxxx/draft](https://doi.org/10.xxxxxx/draft)

## Software

- 6 ▪ [Review](#) ↗
- 7 ▪ [Repository](#) ↗
- 8 ▪ [Archive](#) ↗

---

9 **Editor:** [Open Journals](#) ↗

10 **Reviewers:**

- 11 ▪ [@openjournals](#)

12 **Submitted:** 01 January 1970

13 **Published:** unpublished

## License

14 Authors of papers retain copyright  
15 and release the work under a  
16 Creative Commons Attribution 4.0  
17 International License ([CC BY 4.0](#))  
18

## 19 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.

## 29 Installation and Dependencies

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

## 34 Key Features

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

- 36    **1. Media Mix Modeling (MMM):** Multiple adstock functions, saturation curves, time-varying  
37    parameters via HSGP ([Solin & Särkkä, 2020](#)), experimental calibration for causal inference,  
38    budget optimization with business constraints, time-slice cross-validation, and marginal effects  
39    analysis ([Arel-Bundock et al., 2024](#)).
- 40    **2. Customer Lifetime Value (CLV):** BTYD models ([Fader & Hardie, 2020](#)) including BG/NBD,  
41    Pareto/NBD, and Gamma-Gamma frameworks with hierarchical extensions and individual-level  
42    uncertainty.
- 43    **3. Bass Diffusion Models:** Product adoption forecasting ([Bass, 1969](#)) with flexible parameteri-  
44    zation for innovation and imitation effects across multiple products.
- 45    **4. Customer Choice Models:** Discrete choice analysis ([Train, 2009](#)) based on random utility  
46    theory, including multinomial logit and multivariate interrupted time series models.
- 47    **Production Ready:** All modules feature MLflow ([Zaharia et al., 2018](#)) integration, Docker  
48    containerization, multiple MCMC backends (NumPyro ([Bingham et al., 2019](#)), Nutpie), varia-  
49    tional inference (ADVI), MAP estimation, data connectors (e.g., Fivetran), and comprehensive  
50    diagnostics via ArviZ.

## 51    Key Advantages

52    PyMC-Marketing provides uncertainty quantification through full posterior distributions, exper-  
53    imental calibration anchoring observational models to causal ground truth, and flexible budget  
54    optimization with business constraints. The scikit-learn compatible API ensures seamless  
55    integration into existing data science workflows. The library has been successfully deployed by  
56    companies including HelloFresh and Bolt for production marketing analytics. Comprehensive  
57    tutorials, example notebooks, and video resources are available in the online documenta-  
58    tion at <https://www.pymc-marketing.io/en/stable/>, with community support from over 80  
59    contributors and translations in Spanish.

## 60    Community Guidelines

- 61
  - **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](#)

## 66    Funding

67    PyMC-Marketing is a community-driven project built primarily through volunteer contributions  
68    from over 60 developers. Some contributions by PyMC Labs affiliates have received partial  
69    funding.

## 70    Acknowledgments

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

## 75 References

- 76 Arel-Bundock, V., Greifer, N., & Heiss, A. (2024). How to interpret statistical models  
77 using marginaleffects for R and Python. *Journal of Statistical Software*, 111(9), 1–32.  
78 <https://doi.org/10.18637/jss.v111.i09>
- 79 Bass, F. M. (1969). A new product growth for model consumer durables. *Management Science*,  
80 15(5), 215–227. <https://doi.org/10.1287/mnsc.15.5.215>
- 81 Bingham, E., Chen, J. P., Jankowiak, M., Obermeyer, F., Pradhan, N., Karaletsos, T.,  
82 Singh, R., Szerlip, P. A., Horsfall, P., & Goodman, N. D. (2019). Pyro: Deep universal  
83 probabilistic programming. *Journal of Machine Learning Research*, 20(28), 1–6. <http://jmlr.org/papers/v20/18-403.html>
- 85 Fader, P. S., & Hardie, B. G. (2020). Customer lifetime value: What's next? *Journal of  
86 Interactive Marketing*, 49, 26–37. <https://doi.org/10.1016/j.intmar.2019.08.001>
- 87 Fader, P. S., Hardie, B. G., & Lee, K. L. (2005). Counting your customers: Who are  
88 they and what will they do next? *Management Science*, 51(9), 1371–1387. <https://doi.org/10.1287/mnsc.51.9.1371>
- 90 Google Inc. (2023). *Meridian: Bayesian media mix modeling at scale*. <https://github.com/google/meridian>. <https://github.com/google/meridian>
- 92 Jin, Y., Wang, Y., Sun, Y., Chan, D., & Koehler, J. (2017). *Bayesian methods for media  
93 mix modeling with carryover and shape effects*. Google Inc. <https://research.google/pubs/pub46001/>
- 95 Kumar, R., Carroll, C., Hartikainen, A., & Martin, O. A. (2019). ArviZ a unified library for  
96 exploratory analysis of bayesian models in python. *Journal of Open Source Software*, 4(33),  
97 1143. <https://doi.org/10.21105/joss.01143>
- 98 Meta Platforms Inc. (2022). *Robyn: Marketing mix modeling*. <https://github.com/facebookexperimental/Robyn>. <https://github.com/facebookexperimental/Robyn>
- 100 Salvatier, J., Wiecki, T. V., & Fonnesbeck, C. (2016). Probabilistic programming in python  
101 using PyMC3. *PeerJ Computer Science*, 2, e55. <https://doi.org/10.7717/peerj-cs.55>
- 102 Solin, A., & Särkkä, S. (2020). Hilbert space methods for reduced-rank gaussian  
103 process regression. *Statistics and Computing*, 30(2), 419–446. <https://doi.org/10.1007/s11222-019-09886-w>
- 105 Train, K. E. (2009). *Discrete choice methods with simulation*. Cambridge University Press.  
106 <https://doi.org/10.1017/CBO9780511805271>
- 107 Zaharia, M. A., Chen, A., Davidson, A., Ghodsi, A., Hong, S. A., Konwinski, A., Murching,  
108 S., Nykodym, T., Ogilvie, P., Parkhe, M., Xie, F., & Zumar, C. (2018). Accelerating  
109 the Machine Learning Lifecycle with MLflow. *IEEE Data Eng. Bull.*, 41, 39–45. <https://api.semanticscholar.org/CorpusID:83459546>