

What is GitHub: GitHub is a proprietary developer platform that allows developers to create, store, manage, and share their code. It uses Git to provide distributed version control and GitHub itself provides access control, bug tracking, software feature requests, task management, continuous integration, and wikis for every project.

-Who We Are: Business Insights Team at Github. Our Mission is to **Deliver data, insights and strategies that sharpen GitHub's ability to see around corners.**

-What We Do:

- ◆ **Synthesize & Contextualize:** Use data to create “one version of the truth” of our business for Hubbers

Identify Growth Opportunities: Provide meaningful insights from our data and market analysis, Insights that inform strategies and investments

 **Present Market Insights:** Analyze market intelligence, and **Identify emerging industry trends** and challenges

Mudda's role: Senior Data Analyst of Time-Series Forecasting for Copilot 1 Year out. We have Copilot MAU historical Actuals at the end of each month for the last 2 years Aug 2023' - Aug 2025' which is the main dataset.

What is Copilot: GitHub Copilot is a code completion and programming AI-assistant developed by GitHub and OpenAI that assists users of Visual Studio Code, Visual Studio, Neovim, and JetBrains integrated development environments by autocompleting code.

What is MAU engagement metric? MAU are distinct users who have visited [Github.com](https://github.com), used our API, or made a git request on at least one day in the trailing 28 days. **Copilot MAU**, in this case, is anyone who uses Copilot, that could mean accepting a suggestion, issuing a Copilot Chat prompt, or other Copilot-related activity, at least one day in the trailing 28 days..

How we work:

Remote-first: We collaborate online and value asynchronous conversation

Date-Informed, Not Data-blind: We use analytics to guide choices but balance gut instinct, creativity, and stakeholder feedback.

Leadership:

Mudda -Data Analyst. US Eastern

Tala - Mudda's Boss, director of business insights at Github which falls under Operations in Github

The Business Insights Team

 Leadership

-Tala Karadsheh

-Fintan Ryan-Director of Market Insights. From the U.K.

 Data Analysts

Purpose: Predicting Marketing Impact on Copilot MAU

Understanding the impact of marketing initiatives on Monthly Active Users (MAU) is critical to optimizing growth strategies for Copilot. As Copilot continues to evolve as a multi-modal AI companion across platforms, marketing plays a pivotal role in driving user engagement, retention, and expansion. This methodology outlines a data-driven approach to quantify and predict how specific marketing activities—ranging from thought leadership, organic awareness posts, and product/features launches —affect MAU trends over time. For example, Mudda works with the director of competitive marketing intelligence and copilot Product and Marketing leads. Mudda can go to these people and ask them (Product/Marketing/Director of competitive market intelligence teams) whether they think they expect a new copilot feature release to be strategized as a feature release to be more of an Engagement Play, Acquisition play, or Monetization Play.

By integrating behavioral analytics, attribution modeling, and time-series forecasting, this framework aims to isolate the causal effects of marketing efforts from broader usage patterns and ARR impact. The goal is to empower stakeholders with actionable insights that inform how future campaigns decisions can enhance Copilot's reach and engagement by users.

METHODOLOGY AND APPROACH

Predictive Modeling of Marketing Impact

focus on building a predictive framework to estimate the future impact of marketing activities on key performance metrics: Copilot MAU, ARR, and GitHub sign-up rates. This leverages statistical modeling and machine learning techniques to quantify relationships between marketing inputs and user or revenue outcomes.

The plan is to use the DARTS framework in Python. The five models provide a robust time-series forecasting pipeline that leverages machine learning, probabilistic models, and ensemble techniques to predict the impact of marketing activities on Copilot MAU, ARR, and GitHub sign-up rates. This stage involves:

- Data Preparation: Determining what information to integrate into the models as covariates, in addition to the outputs from Stage 1
- Model Training: Add agreed upon covariates into the selected predictive models
- Backtesting & Validation: Test models to evaluate performance over rolling windows and identify the MAPE scores. Visualize prediction intervals to assess uncertainty.
- Scenario Forecasting: Simulate future marketing moments by injecting hypothetical covariates. Forecast expected changes in MAU, ARR, and sign-ups under different campaign scenarios.

The quality of inputs/covariates can affect the quality of our forecasts.

->Covariates for Time-series forecasting using Models via Darts (theta, Prophet, XGBoost, LightGBM, and Exponential Smoothing):

- 1)Historical Copilot-All MAU (Copilot MAU at the end of each month last 2 years Aug 23' - Aug 25') and MoM Growth rates of Copilot MAU
- 2)Copilot-All Monthly Signups last 2 years
- 3)Copilot_free Monthly Signups
- 4)Copilot-All Current seats per plan/sku by Geo
- 5)Future Weighted Internal Covariates (scenario-based weighting/**Strategic Alignment Scoring**. e.g. Forecast 1 if we expect Copilot MAU as an Engagement Play: MAU 50%, Signups 35%, ARR 15%. **Monetization Play** for Forecast 2: Weight ARR 50%, MAU 35%, Signups 15%. Forecast 3: **Acquisition Plays**: Weight Signups 50%, MAU 35%, ARR 15%).
- We will not include global expansion framework data , since we're not forecasting by geography
- No external covariates (data outside of our internal data) will be included

Exact names of covariates in csv files in local machine:

CopilotAll-Monthly-Signups-Sep2025.csv
Copilot-Adoption-MonthlySignups-Sep2025.csv
CopilotFree-MonthlySignups-Sep2025.csv
Copilot-All-CurrentSeats-ByGeo-Sep2025.csv
GithubOverall-MAU-Engagement-Last3Years-Sep2025.csv

->Note:

-If we think any of the above covariates add more “noise” than value, we can remove any of the above mentioned covariates, more covariates doesn’t always equal better, being strategic about the covariates fed into the ML models is usually the ideal route.

-No external covariates (data outside of our internal Github data warehouse) will be included

Stage 3: Real-Time Monitoring and Campaign Optimization

With predictive models in place, deploying a real-time feedback loop that continuously monitors marketing activity and its impact on key metrics—Copilot MAU, ARR, and GitHub sign-up rates. This stage transforms the forecasting engine into a tactical asset, enabling dynamic adjustments to campaigns based on observed performance and updated predictions.

During this stage, Mudda plans to integrate live data streams into a continuously updating forecasting pipeline powered by DARTS. The idea is to refresh predictions regularly based on actuals so that the models continue to be informed by shifts and therefore become more informed with their predictions. Insights from the predictions will be shared to calibrate investments, future planning, and improve marketing ROI.