

Bolt-Process Document

**Table of Contents**

[1. Background about Bolt 2](#_Toc80365116)

[2. Business Questions 2](#_Toc80365133)

[3. Scope 2](#_Toc80365134)

[4. Model Approach 3](#_Toc80365135)

[5. Modeling Process 3](#_Toc80365136)

[6. Output/Results: 7](#_Toc80365137)

# Background about Bolt

Bolt is a ride sharing company. The main advertising channel is digital, with spending predominantly on Facebook, Google, Snap, Apple Search, Tik-Tok. Referral codes are also used heavily to incentivize first and repeat rides. More recently OOH has been used to drive awareness.

# Business Questions

* What is the contribution of base vs incremental?
* What are the key drivers of Fist time activations?
* What are the incremental activations driven by Media advertising?
* What is the ROI / CAC for all marketing drivers?
* How does media and promotion work together to impact first time user activation?
* How do we measure and optimize the impact of the advertising campaigns or marketing budget?

# Scope

**KPI**

1. First time activations.

2. First year LTV

Time Period: 730 days – 11th Jun. 2019 – 9th Jun. 2021

Granularity: Daily level data

**Model Measures**

Measures considered for the analysis are categorized into following groups:

* Base
* Macro Environment
  + Temperature
  + Precipitation
* Calendar
  + Weekend Flag
  + Holiday
* Price
  + Avg Distance Price
  + Avg Supply Demand Multiplier
* Others
  + Bolt ETA
  + Mobility Data
* Media
  + Facebook Impressions
  + Google Impressions
  + Apple search Impressions
  + OOH Impressions
  + Twitter Impressions
  + Snap Impressions
  + Tik-Tok Impressions
  + Influencer’s & Blog visitors’ data
* Non-Media
  + Signup’s cost
  + Event Cost
  + Referral cost
  + Lifecycle data (Total messages that sent to the unique users)



















# Model Approach

Started with the ‘First time activations.’ (KPI) using our platform Demand Drivers Edge (DDE). DDE runs statistical models by blending media, promotions, macro-economic indicators data and other bolt specific inputs required to quantify the relationship on KPI.

**KPI=Intercept+β1\* Base+β2\* Media+β3\* Promo+....**

# Modeling Process

1. **Base** –
2. Started with base model by including Weather, ETA, Holidays, Weekend effect, Covid mobility, lockdown status



1. **Media** –
2. **Next included media (one variable at a time).** To test media, we considered spend share of media variables to identify the significant ones to begin with.
3. Suitable Transformations (Ad stock /Gamma) have been selected and parameters are identified by running multiple iterations and comparing model fits for each.
4. List of media variables used in the model.

|  |  |  |  |
| --- | --- | --- | --- |
| **Media** | Transformation used | Granularity | Spend share (with in Media) |
| **Facebook, Google, Apple** | Ad stock/Gamma on impressions | Used by Objective | Facebook (34%), Google (36%), Apple (4%) |
| **Twitter, Snap, Tik-Tok, OOH** | Ad stock/Gamma on impressions | Aggregate | Twitter (0.2%), Snap (3%), Tik-Tok(0.6%), OOH (5%) |
| **OOH Q2-2021, Influencer** | Ad stock/Gamma on reach | Aggregate | OOH Q2-201(13%), Influencer (3%) |



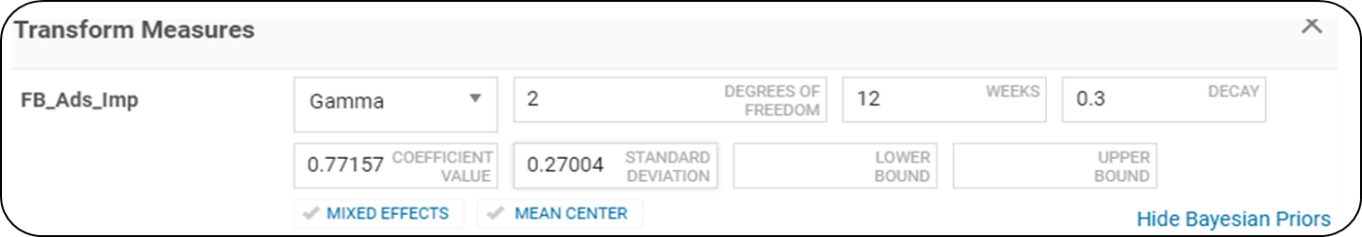
1. Model validation is done using statistical and business validity.
   1. Statistical validity is done by checking – R-square, MAPE / Hold out MAPE, Coefficient Sign, VIF, P value & T stat.
   2. Business validity is done by checking contributions vs. spend share.
2. For each media, multiple iterations are run and tracked how the above-mentioned metrics are varying.

****

1. Based on multiple iterations range of coefficients / contributions are tracked for media variables.



1. An average of these coefficients is used as Prior, and the range noticed is used to determine standard deviation for the respective variables.



1. Once priors are incorporated model stability has been validated through statistical diagnostics for each iteration to ensure overall model fit is intact.
2. Holdout MAPE is used as the criteria for ensuring in-sample model fit validation.

|  |  |  |
| --- | --- | --- |
| **R2** | **MAPE** | **HoldOut MAPE** |
| 75.7% | 35.7% | 36.3% |

1. **Promotion** –

|  |  |  |
| --- | --- | --- |
| **Type** | **Granularity** | **Spend share  (within Promotions)** |
| Referral bonus | by offer | 42% |
| Sign-up bonus | by year | 54% |
| Event campaign cost | By campaign | 4% |

1. **Referral bonus**: Used heavily as a marketing tool during the launch and reduced in later periods. As we see below, variation in referral costs is strongly mimicking the variation in activations (high correlation). Such behavior would result in model attributing high proportion of activations to referrals. Similar pattern is observed for sign-up bonus as well.

Chart

Description automatically generated

To ensure we get unbiased read, we used

* Total referral cost by offer value eg: 8gb off, 10gbp off etc. and used directly in the model
* Used the redeemed no. of users by offer value to validate the results.

1. **Sign up costs** – Sign up bonuses are given to users throughout the modelling time. They are advertised in media and through lifecycle and the user has a choice to use the promo code on the first ride.

Since sign up costs are heavily linked with first time activations, their correlation is high with the KPI. Hence, using this variable directly in the model might lead to its impact being inaccurately captured.

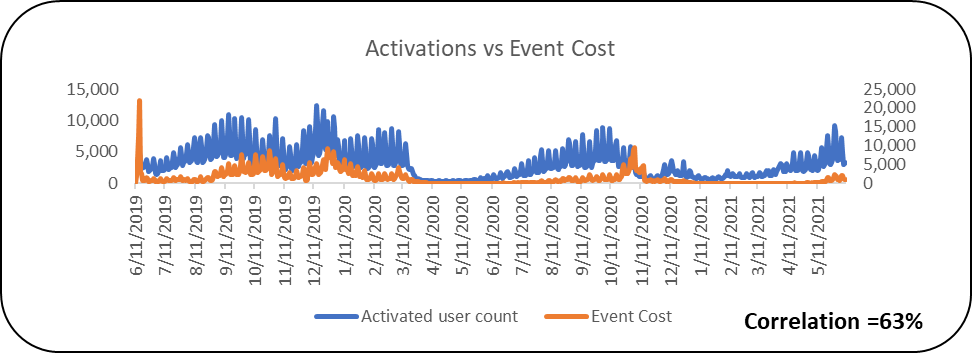
Chart

Description automatically generated

To ensure we get unbiased read, we analyzed

* Sign up bonuses by year (e.g.: 2019,2020 etc.) and used directly in the model.
* We used Redeemed no. of users by year to validate the results.

1. **Event campaign cost** – Bonuses linked with any specific marketing events executed by Bolt



To ensure we get unbiased read, we used

* Event bonus data – by event campaign (only the major campaigns that have the highest spend share are considered & everything else clubbed under others) and used these campaigns directly in the model.
* Redeemed no. of users by campaign used to validate the results.

1. **Lifecycle** –
2. We have tested in the model by breaking the variables by message type.
   1. We applied Lag transformation to the variables where Lag value is iterative.
   2. Lag transformation will consider the lag it takes for a user to activate after receiving the message.

# Output/Results:

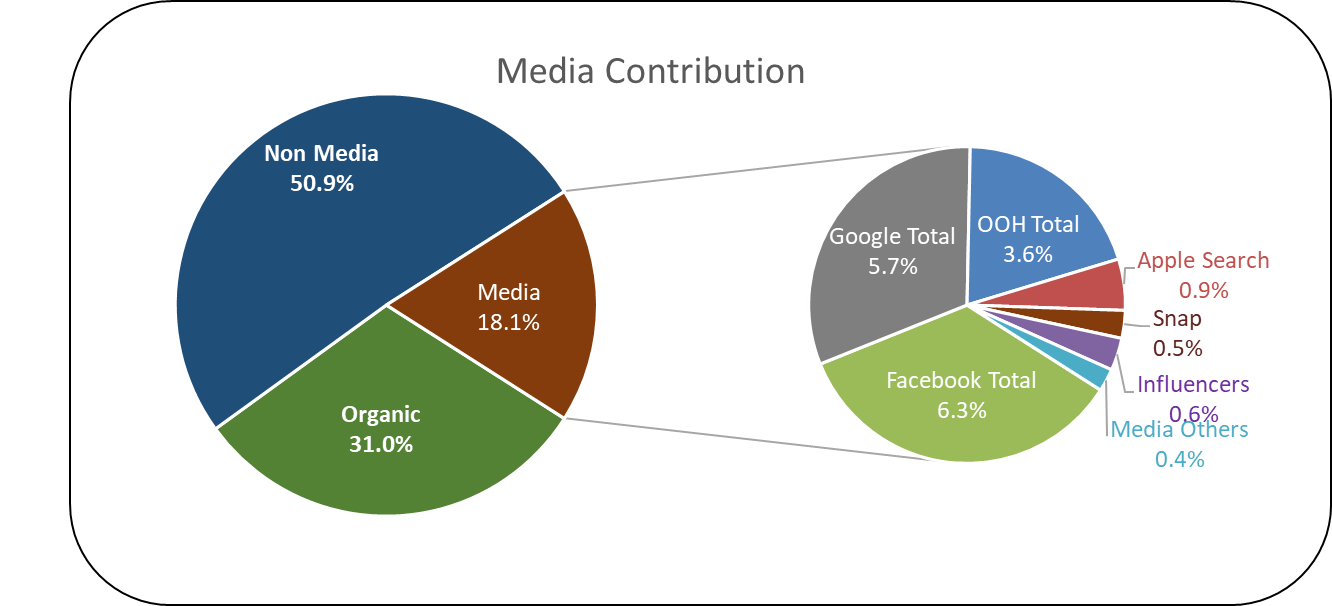
1. **Model Fits – First time Activations KPI** -- The model was developed on data from 11th Jun’19 – 9th June’21 and the resulting model fit is robust based on statistical diagnostics.

**Chart, line chart, histogram

Description automatically generated**

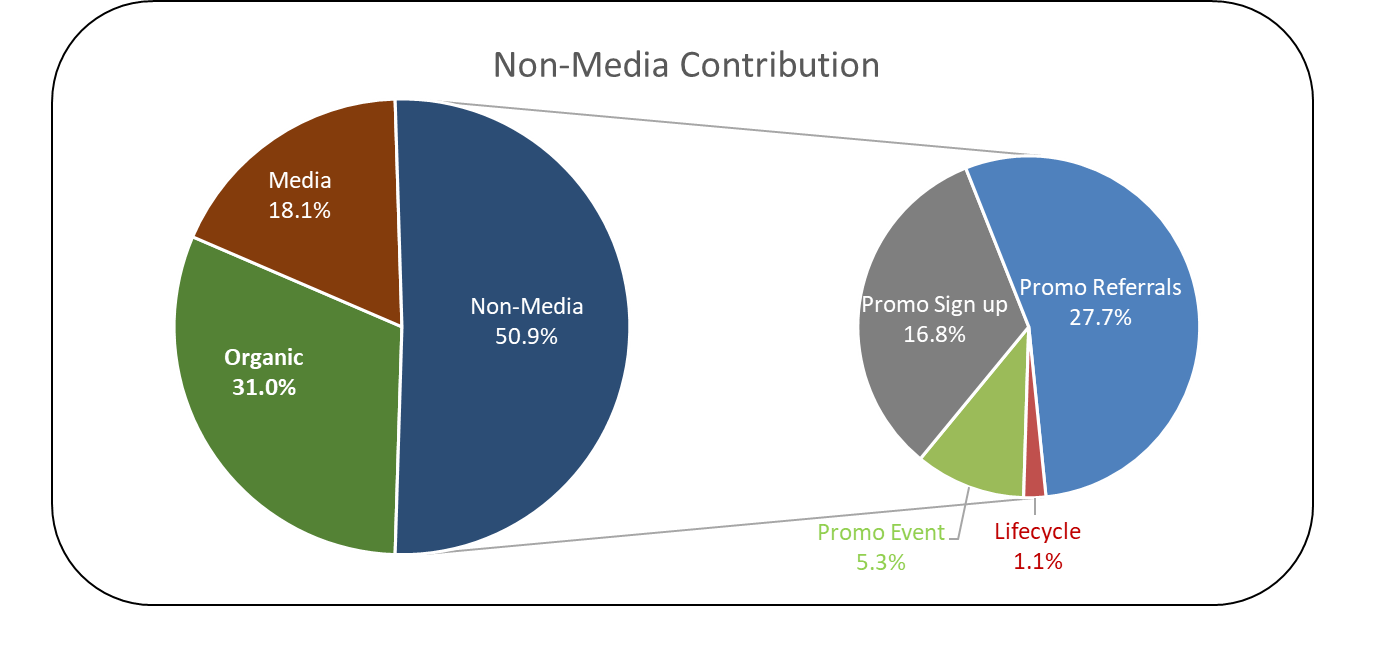
1. **Contribution** – Given the current levels of execution what is the individual contribution of each media/promotion.
2. **Media:**

* 18.1% of the activations from 11th Jun 2019 to 9th Jun 2021 was incremental, driven by media with the majority from Facebook (6.3%), Google (5.7%) & OOH (3.6%).
* 31% of the Total activations is Organic.i.e. On an Average around 31% of Organic activations is achieved in absence of media during "11th Jun-19 to 9th Jun-21“which essentially were driven by Core Users.



1. **Non-Media:**

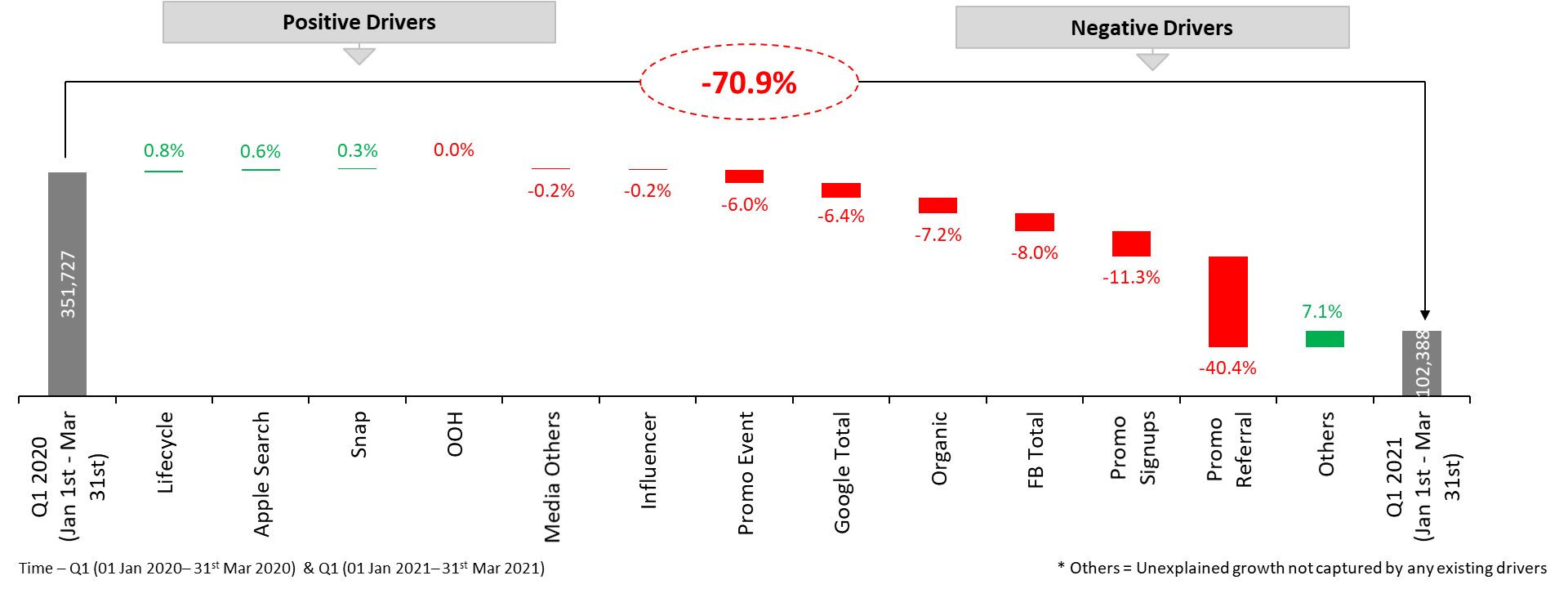
* ~51% of the activations was incremental and driven by Promotions with the majority from Referrals (~28%) & sign up (16.8%).



1. **Due-To Change** -- Due-to change (%) indicates the change between any two periods in KPI and model breaks down the change and attributes to key drivers in the model.

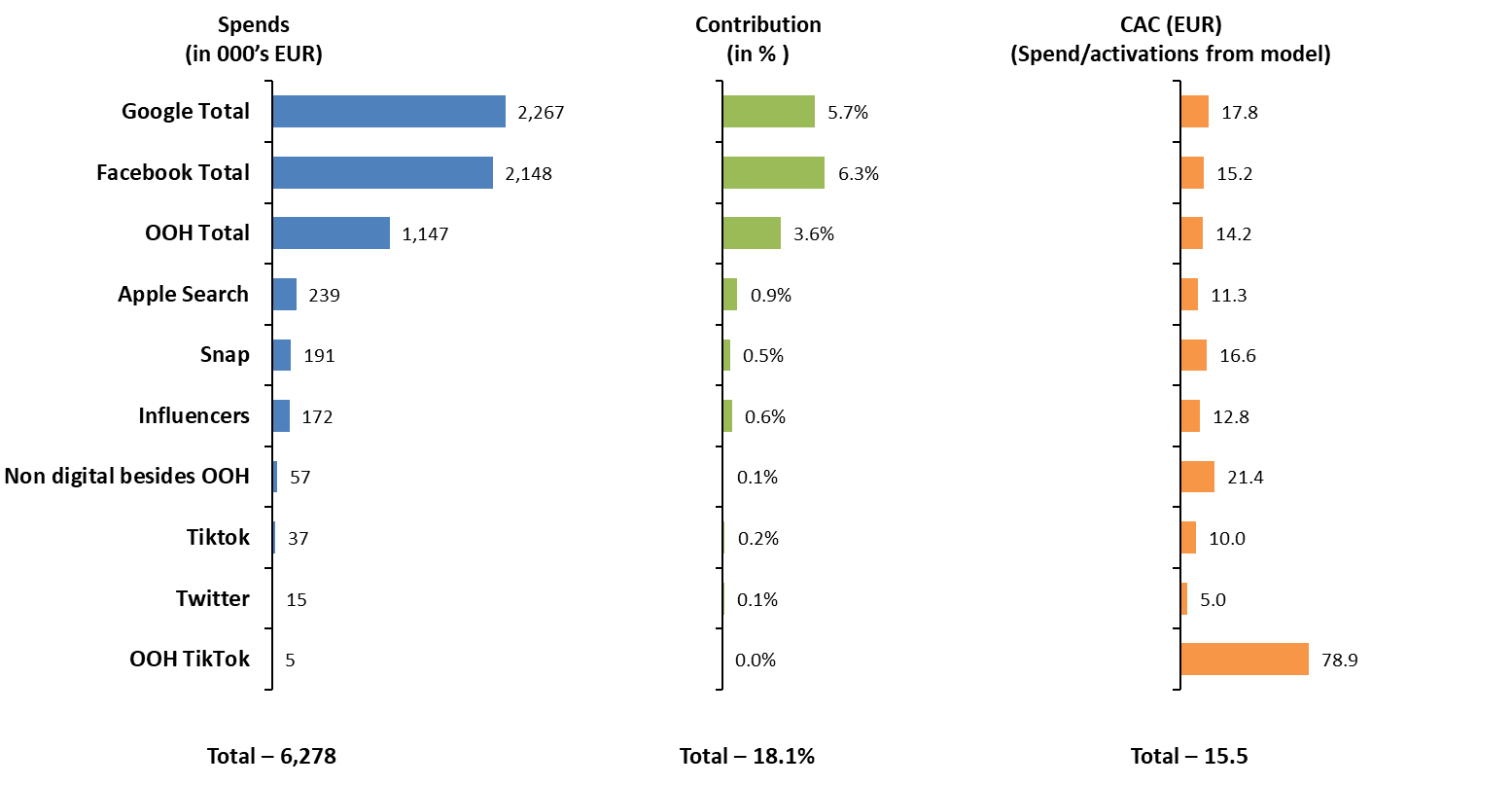
* In many cases, there is a gap between actual change and model predicted change and this is grouped under the bucket “Others” (Model Error).
* **Drivers of change:** Activations decreased (~ -71%) in Q1 (Jan’20 – Mar’20) vs Q1 (Jan’21 – Mar’21) mainly driven by significant drop in Referrals. Signups & Facebook support in Q1 2021

**Q1’20 vs. Q1’21**

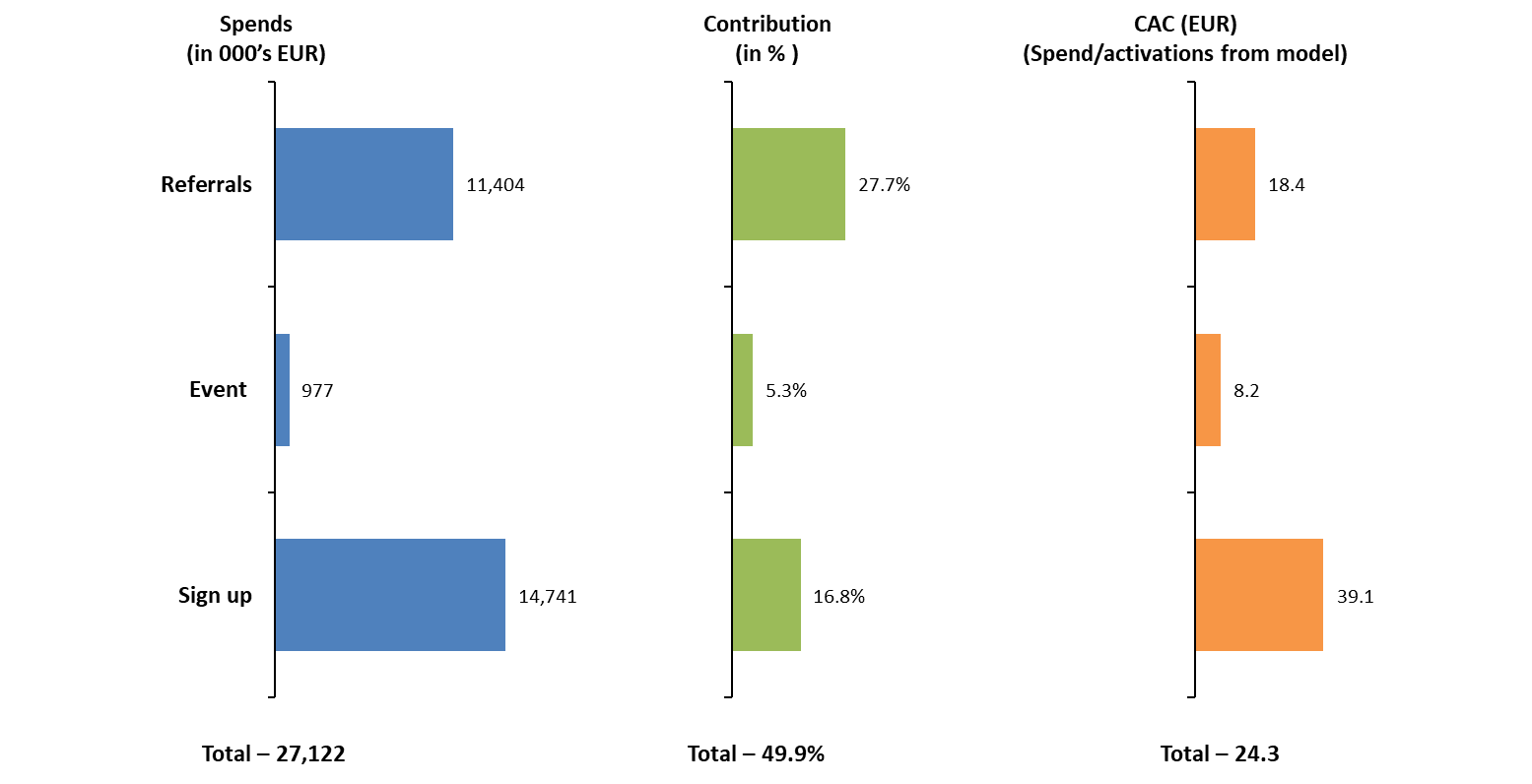


1. **Cost of Acquisition** – Cost of acquisition is calculated as Spends (for respective media) divided by Number of first-time activations.

**a) Media**: Among the largest media spends channels, Apple has a lower CAC (11.3EUR) relative to OOH (14.2EUR) & Facebook (15.2EUR) in the 11th Jun’19- 9th Jun’21 time period.

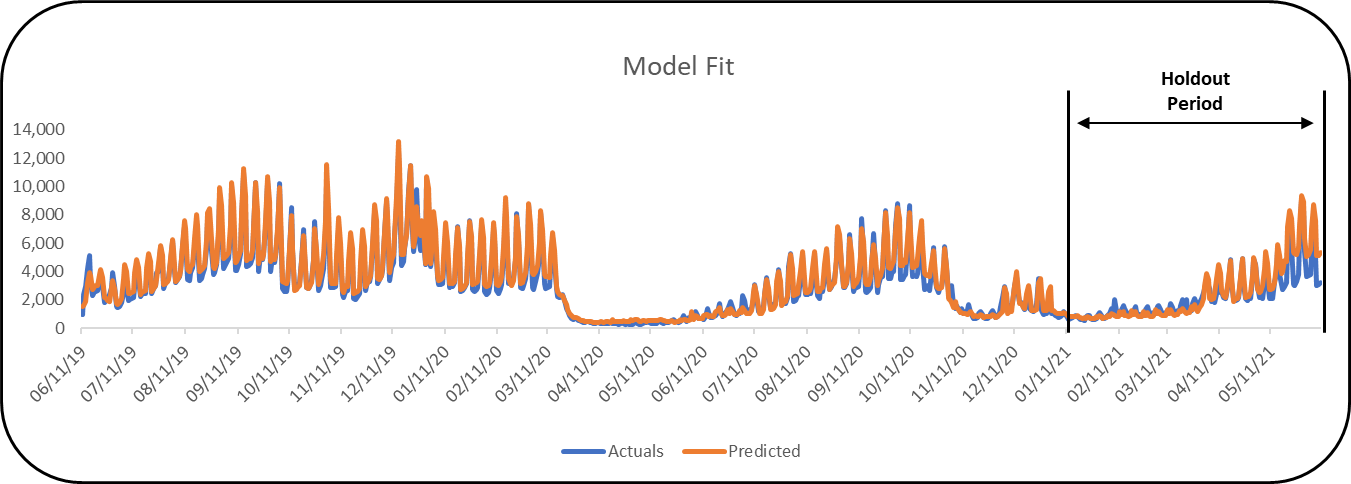
****

**b) Non-Media:** Total Promotion CAC was 24.3EUR, while the Event costs has lower CAC (8.2 EUR) compared to Referrals and Sign ups.

****

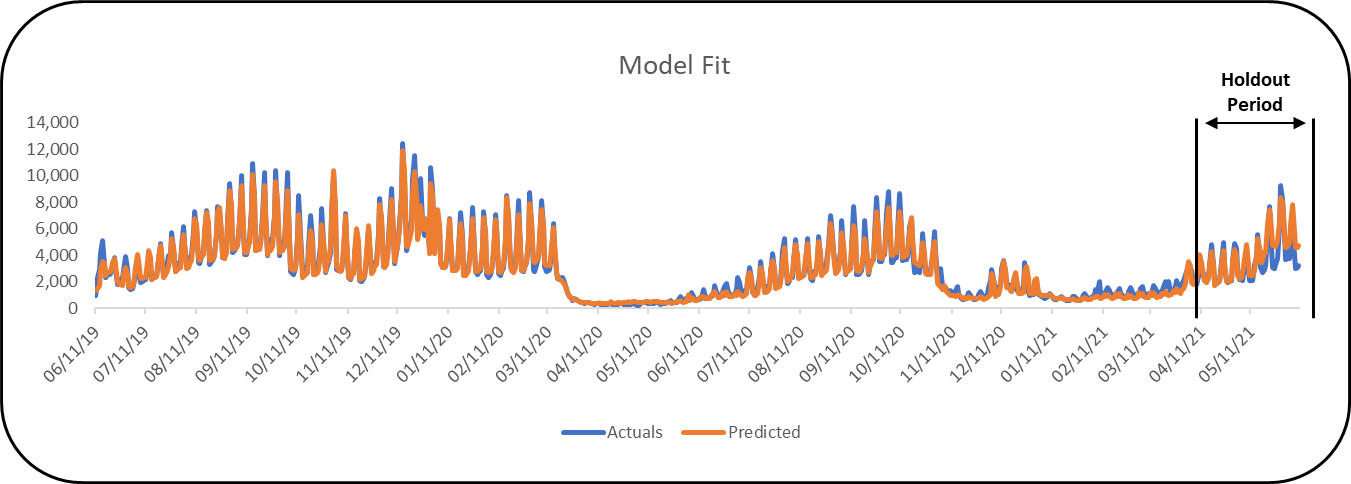
1. **Forecasting Model Fits – First time Activations KPI**

* Model validation is done by predicting the activations for 6 months Jan’21 – Jun’21 (holdout period) of data and the prediction was fairly accurate.



|  |  |
| --- | --- |
| **Daily Model Statistics** | |
| **R Square** | **96.0%** |
| **MAPE** | **17.5%** |
| **Hold Out MAPE (6 Months)** | **21.7%** |

* Further ran the model validation using additional two months of data (1st Apr’21 to 9th Jun’21)



|  |  |
| --- | --- |
| **Daily Model Statistics** | |
| **R Square** | **96.1%** |
| **MAPE** | **13.7%** |
| **Hold Out MAPE (2 Months)** | **17.1%** |