

Bolt-Process Document

**Table of Contents**

[1. Background about Bolt 2](#_Toc77333946)

[2. Business Questions 2](#_Toc77333963)

[3. Scope 2](#_Toc77333964)

[4. Model Approach 3](#_Toc77333965)

[5. Modeling Process 3](#_Toc77333966)

[6. Promotion 5](#_Toc77333967)

[7. Output/Solution: 7](#_Toc77333968)

# 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. Started with base model by including Weather, ETA, Holidays, Weekend effect, Covid mobility, lockdown status



1. **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.
2. Suitable Transformations (Ad stock /Gamma) have been selected and parameters are identified by running multiple iterations and comparing model fits for each.
3. 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%) |



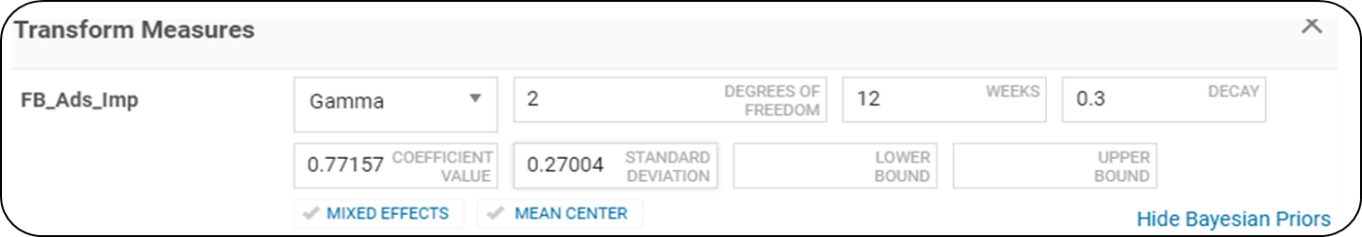
1. Model validation is done using statistical and business validity.
   1. Statistical validity is done by checking – R-square, MAPE / Hold out MAPE, Coefficient Sign, VIF, P value & T stat.
   2. Business validity is done by checking contributions vs. spend share.
2. For each media, multiple iterations 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% |

# Promotion

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

To ensure we get unbiased read, we recommend

* Breaking the total cost into offer value level variables eg: 8gb off , 10gbp off etc and using directly in the model
* Using the Redeemed no. of users by offer value at a daily level to validate the results.

Chart

Description automatically generated

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.

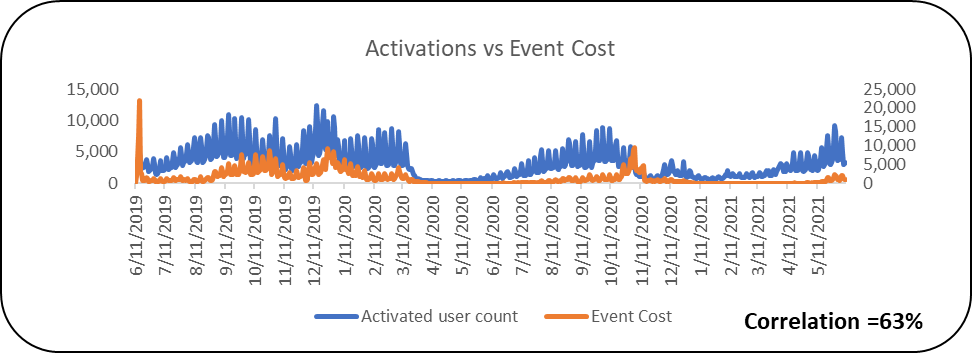
To ensure we get unbiased read, we recommend.

* Using Sign up bonuses by offer value (e.g.: 8gb off , 10gbp ) similar to referral cost (or any other granularity available)
* Redeemed no. of users by offer value at a daily level (No of users who availed offer per day) would also help in either using directly in the model or validating the results from the model

Chart

Description automatically generated

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



To ensure we get unbiased read, we recommend

* Using Event level bonus data – For each event campaign what was the cost
* Redeemed no. of users by offer value at a daily level (No of users who availed offer per day) would also help in either using directly in the model or validating the results from the model.

**Non-media --**

Lifecycle –

1. 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/Solution:

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

Chart, pie chart

Description automatically generated

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).
* Model error to be kept as low as possible.

Chart

Description automatically generated with medium confidence

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

**How confident are we on the contributions from media/promo?**

* We will use **Statistical parameters** like coefficients, standard errors, t-value etc. to validate the model.
* **Holdout MAPE** is used to measure or verify the accuracy of a prediction.
* **Out of sample** is another way to evaluate forecasting performance.

Use one or two months of additional data to predict the KPI and then comparing the how near the prediction is to the actuals.