

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%) |



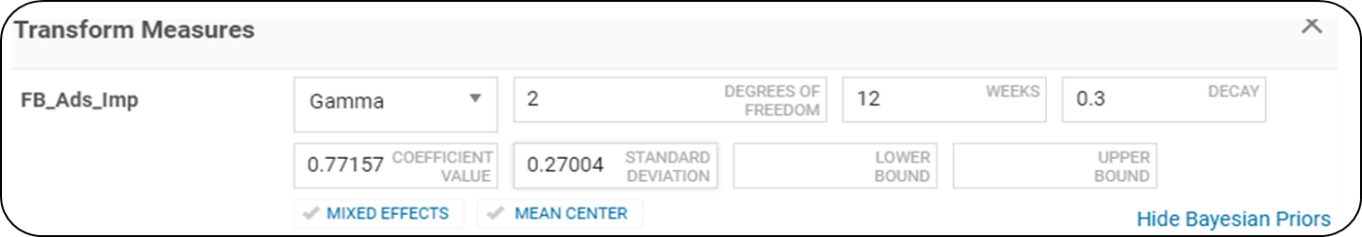
1. Model validation is done using statistical and business validity.
   1. Statistical validity is done by checking – R-square, MAPE / Hold out MAPE, Coefficient Sign, VIF, P value & T stat.
   2. Business validity is done by checking contributions vs. spend share.
2. For each media, multiple iterations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

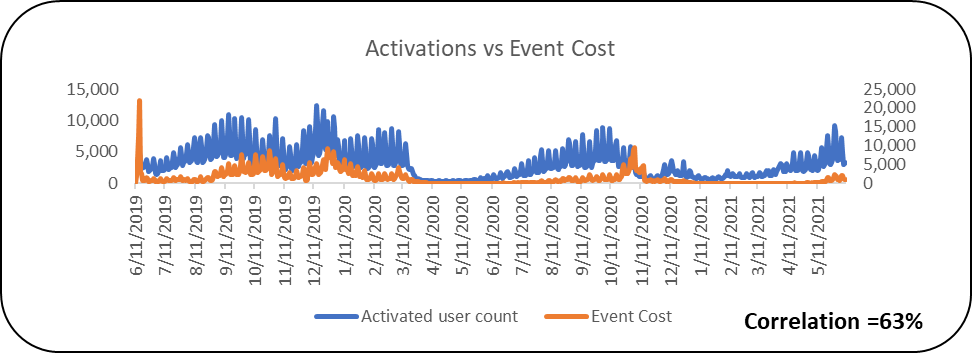
To ensure we get unbiased read, we recommend.

* 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.

Appendix:

**Modelling Approach in detail:**

Modelling breaks down business metrics into differentiate between contributions from marketing (Media) and promotional activities (incremental drivers) vs. other (base) drivers. These factors affecting KPI can be defined as:

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

1. **Base Drivers:** The outcome of base variables is achieved without any advertisements.

Base measures which are used in the model are as follows:

|  |  |  |  |
| --- | --- | --- | --- |
| **Type** | **Measures** | **Transformation** | **Assumptions** |
| External Factors | Precipitation, Temperature | Direct | These uncontrollable factors greatly influence the KPI performance.  E.g.: There is a high chance that people would opt for cabs during rainy season |
| Key Holidays |  | Direct |  |
| Categories ETA | XL ETA, Bolt ETA | Direct | Tried to add only those variables that have a significant impact in the model. |
| Weekend Dummy |  | Direct | Since our activation model has an on(weekend) and off trend throughout the modelling period, year wise weekend dummies are introduced. |
| Covid Mobility | Workplaces | Direct | Direct impact of COVID-19 on KPIs seen due to reduction in the number of visitors to workplaces during the lockdown period and hence used workplaces in our model. |

**Model Fit- Activations (Base)**

Chart

Description automatically generated

1. **Incremental Drivers:**

* Activations generated by marketing activities like Digital spends, Promotions, lifecycle etc. These total incremental activations are split into activations from each input to calculate contribution to total activations.
* As mentioned, we tested each **media variables** based on its spend share and ran multiple iterations along with the base measures, tested its statistical diagnostics (model stats) to arrive at the final model.
* Following are the media measures used in the model, Fit and the Summary:

**Model Fit- Activations (Base+ Media)**Chart, line chart

Description automatically generated

**Media Summary**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Metrics** | **Impression (000')** | **Spend** | **Spend %** | **Contribution Value** | **Contribution %** | **CAC** |
| Snap | 94,318 | 190,545 | 3.0% | 11,488 | 0.5% | 16.6 |
| Google Ads (no search, no video) | 468,917 | 1,962,900 | 31.3% | 107,657 | 4.8% | 18.2 |
| Google Ads (no search, Video) | 14,335 | 58,906 | 0.9% | 820 | 0.0% | 71.9 |
| Google Ads (Search, Brand) | 1,829 | 168,799 | 2.7% | 9,978 | 0.4% | 16.9 |
| Google Ads (Search, Non-Brand) | 1,900 | 76,017 | 1.2% | 8,670 | 0.4% | 8.8 |
| Apple Search Brand Ads | 276 | 35,912 | 0.6% | 7,388 | 0.3% | 4.9 |
| Apple Search Non-Brand Ads | 1,213 | 203,433 | 3.2% | 13,833 | 0.6% | 14.7 |
| Tik-Tok | 48,310 | 36,748 | 0.6% | 3,679 | 0.2% | 10.0 |
| OOH Tik-Tok | 2,916 | 5,179 | 0.1% | 66 | 0.0% | 78.9 |
| OOH\_Imp\_without\_Q2\_2021 | 268,526 | 356,493 | 5.7% | 35,112 | 1.6% | 10.2 |
| OOH Sum of Ad Plays | 122,590 | 790,877 | 12.6% | 45,850 | 2.0% | 17.2 |
| Influencers spend | 0 | 172,435 | 2.7% | 13,467 | 0.6% | 12.8 |
| Twitter- Total | 14,365 | 14,913 | 0.2% | 2,987 | 0.1% | 5.0 |
| Non-digital besides OOH spend | 0 | 56,875 | 0.9% | 2,655 | 0.1% | 21.4 |
| FB Ads | 343,031 | 1,813,104 | 28.9% | 107,404 | 4.8% | 16.9 |
| FB Brand Acquisition Ads | 7,613 | 23,750 | 0.4% | 1,380 | 0.1% | 17.2 |
| FB Post Brand Awareness | 22,222 | 48,745 | 0.8% | 7,700 | 0.3% | 6.3 |
| FB Post engagement | 34,777 | 201,011 | 3.2% | 19,948 | 0.9% | 10.1 |
| FB Post Reach | 29,582 | 61,173 | 1.0% | 4,838 | 0.2% | 12.6 |
| Total | 1,476,721 | 6,277,816 | 100.0% | 404,919 | 18.1% | 19.5 |

* Below table shows the treatment of **non-Media variables**:

|  |  |  |
| --- | --- | --- |
| **Promo variables** | **Granularity** | **Assumptions/Reasons to use it in the model** |
| Referrals Cost | By offer value | * Reduces high correlation and multicollinearity with the KPI |
| Event Cost | Analyzed at the Campaign level | * Breaking down the variables by campaign to understand which of them are driving more activations. |
| Sign up cost | Analyzed on an annual basis | * Break down costs by year to accurately measure activation impact |
| Lifecycle | By type of messages sent | * Used Lag transformation * Lag transformation will consider the lag that it takes for the user to activate after receiving the message. |

**Model Fit-Activations (Base+ Media+ Promotions +Lifecycle data)**

Chart, line chart

Description automatically generated

**Non-Media Summary**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Metrics** | **Spend** | **Spend %** | **Contribution Value** | **Contribution %** | **CAC** |
| Referral\_Cost\_12gbp | 2,916,023 | 10.8% | 133,918 | 6.0% | 21.8 |
| Referral\_Cost\_8gbp | 2,422,378 | 8.9% | 167,847 | 7.5% | 14.4 |
| Referral\_Cost\_10gbp | 2,120,648 | 7.8% | 102,523 | 4.6% | 20.7 |
| Referral\_Cost\_14gbp | 2,007,667 | 7.4% | 99,774 | 4.5% | 20.1 |
| Referral\_Cost\_100pct | 141,183 | 0.5% | 4,291 | 0.2% | 32.9 |
| Referral\_Cost\_Others | 1,796,345 | 6.6% | 113,089 | 5.0% | 15.9 |
| Event\_BAPFTR10GBP\_Cost | 71,590 | 0.3% | 13,924 | 0.6% | 5.1 |
| Event\_FreetripWaitingVol1\_Cost | 104,286 | 0.4% | 25,179 | 1.1% | 4.1 |
| Event\_FreetripWaitingVol23\_Cost | 166,470 | 0.6% |  | 0.0% |  |
| Event\_OOH\_Campaign\_Cost | 136,398 | 0.5% | 30,721 | 1.4% | 4.4 |
| Event\_RMNFTR10GBP\_Cost | 161,510 | 0.6% | 28,930 | 1.3% | 5.6 |
| Event\_RMNFTR8GBP\_Cost | 115,496 | 0.4% | 20,931 | 0.9% | 5.5 |
| Event\_Cost\_Others | 107,560 | 0.4% |  | 0.0% |  |
| UK, Demand Boosting FTR | 113,980 | 0.4% |  | 0.0% |  |
| Signups\_Costs\_2019 | 6,633,221 | 24.5% | 132,665 | 5.9% | 50.0 |
| Signups\_Costs\_2020 | 4,696,047 | 17.3% | 175,624 | 7.8% | 26.7 |
| Signups\_Costs\_2021 | 3,411,483 | 12.6% | 69,024 | 3.1% | 49.4 |
| Total | 27,122,284 | 100% | 1,118,440 | 49.9% | 19.8 |

**Lifecycle-Summary**

|  |  |  |
| --- | --- | --- |
| **Lifecycle by type** | **Contribution Value** | **Contribution %** |
| Adhoc | 4,008 | 0.2% |
| User Campaign | 19,758 | 0.9% |
| Total | 23,766 | 1.1% |

1. **Overall Model Fits – Activation 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. **Activation Contribution:**
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.

Chart, pie chart

Description automatically generated

1. **Non-Media:**

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

Chart, pie chart

Description automatically generated

1. **Drivers of Change – Activations:**

* Activations decreased (-9.4%) in H1 (1st Jan – 9th Jun’20) vs H1 (1st Jan – 9th Jun’21) was driven largely by decrease in Referrals & Facebook spend in H1 2021
* Increased investment in OOH and sign up offset the large negative impact coming from Referrals & Facebook

Graphical user interface, application

Description automatically generated

#Pending: Model validation (done by predicting the activations for June 21(holdout period)