**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 and incremental sales?

 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. Predicted first year value for new activations.
2. First year LTV

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

Granularity: Daily level data

Model Measures:

Typically, measures to be modeled can be grouped into the following categories:

* 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

**Model treatment – (specific to Bolt’s marketing variables)**

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

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

**Media –**

1. To test media, we started with some of the most important media variables. Either use business logic or look at the spend share of media variables to identify the biggest ones
2. Transformation parameters are determined by running multiple iterations and comparing model fits for each.

Below is the list of media variables used in the model.

|  |  |  |
| --- | --- | --- |
| Metric (Impressions) | Transformation used | Granularity |
| Facebook | Ad stock/Gamma | Used by Objective |
| Google | Ad stock/Gamma | Used by Objective |
| OOH | Ad stock/Gamma |  |
| Twitter | Ad stock/Gamma |  |
| Snap | Ad stock/Gamma |  |
| Tik-Tok | Ad stock/Gamma |  |
| Apple | Ad stock/Gamma | Used by Objective |
| Influencers’ reach | Ad stock/Gamma |  |

1. **Use the base model and start by inserting one media at a time.**

**Chart

Description automatically generated**

**Chart, histogram

Description automatically generated**

1. The validation checks we have done after adding a variable.
   1. R square – Does the R square increase after inserting this variable.
   2. MAPE / Hold out MAPE – Does the MAPE reduce by inserting this variable.
   3. Coefficient Sign – The sign is indicative of the relationship with the KPI. Media execution should have a positive effect on your KPI so the coefficient should be positive.
   4. VIF – VIF must be low to ensure there isn’t Multicollinearity with any other independent variable.
   5. P value & T stat– these are used to check the confidence of the coefficient value
2. Keep a track of the above metrics after each iteration.
3. For each media, multiple iterations can be run and check how are the metrics are varying.

Graphical user interface, application

Description automatically generated

1. This will give a range on how media changes based on other variables.
2. An average of these coefficients can be used a Prior and the range noticed can be the std deviation to be used for the respective variables. For E.g.:

Graphical user interface, application

Description automatically generated

**Promotion –**

**Challenges:**

1. Correlation of promo variables with the KPI is relatively high and Using cost directly in the model would be good for Model fit and MAPE but may not be accounting for the consumer behavior surrounding this variable.

Chart

Description automatically generated

Chart

Description automatically generated

1. Coupons is something which is always present and moves along with the total activations. When we model this promo variables to predict activations, it is carrying the total variation away and we would get an artificially suppressed impact of media this is one of the predictive modelling challenges.

**Possible Solutions:**

1. One approach we tried is to bucketizing the promo variables (Sign up/Event cost) into ranges because it is happening almost every day. And any predictive model will not be able to tease it out so effectively as it has continued value. So, creating ranged variables based on per user cost eg: 0-5gbp, 5-10 gbp, 10-15 gbp etc
   1. This would result in reducing the high correlation that direct cost has with the KPI.
   2. Applying a transformation would also help in capturing consumer behavior associated with this variable.
   3. Transformation can be Moving average of x number of days.
   4. Number of days can be iterative based on how well it spreads the data whist retaining the variation.
2. Using Flag variables
   1. Creating a flag variable is a good way to test promotion since it tests for on vs off and eliminates the high correlation of cost with dependent.
   2. Cannot use flag directly with this variable since its continuous
   3. Can use flag after creating the Ranges for per user cost.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Activated User | Signup cost | Per User Signup Cost | Variables created | | |
| 5-10gbp | 10-15gbp | >15gbp |
| 2,918 | 5,836 | 2 | 0 | 0 | 0 |
| 4,192 | 41,920 | 10 | 1 | 0 | 0 |
| 5,102 | 40,816 | 8 | 1 | 0 | 0 |
| 3,205 | 57,690 | 18 | 0 | 0 | 1 |
| 2,298 | 20,682 | 9 | 1 | 0 | 0 |
| 2,569 | 12,845 | 5 | 0 | 0 | 0 |
| 2,652 | 42,432 | 16 | 0 | 0 | 1 |
| 3,236 | 38,832 | 12 | 0 | 1 | 0 |
| 3,836 | 11,508 | 3 | 0 | 0 | 0 |
| 2,942 | 44,130 | 15 | 0 | 1 | 0 |

* 1. Downsides of using flag variable:
     1. If the data for any range is continuous for a long period of time, model would not be able to attribute the right impact from it.
     2. Since data is only on vs off, the day-to-day variation in consumers using this promotion can get diminished.

1. We can break out the coupons by year to measure the activation impact properly.
2. Using Exponential decay – Since Rate of activations is declining over a period of time, it’s impact can captured through this transformation.

**Non-media --**

Lifecycle –

1. We have tested in the model by breaking the variables by message type.
   1. Can apply 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 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 one of the ways to evaluate forecasting performance.

Here we take one or two months of new data in order to see how accurately our model predicts the results and determine whether the statistics of errors are similar to the original model.