**Step 2.1 - Model walkthrough (Model setup, Transformation & Model iteration)**

We will be walking you through the modeling process using Gume data as an example. Since we have already covered the ingestion and classification of data in the training videos, we will be jumping straight to the modeling process.

**Parameter**

In the Modeling module a user is first required to define the model parameters.

1. **Dependent variable -** It is the KPI being modeled. In this example its Revenue sales
2. **Model duration -** The platform in default selects the total time-period ingested in the platform but we also have the option of selecting a more recent time-period
3. **Model type -** It has 3 options
   1. **Unpooled -** When the data is available only at a single dimensional level. We are only presented with this option. Which in our case in this example.
   2. **Pooled & Mixed effects - When** the data is multi-dimensional, at a regional or category level, we are also presented with the options of pooled and mixed effects. **Mixed effects** in particular is applied when a/group of media tactics are expected to have varying elasticities across the dimensions and if the results are required at individual dimension level. Also, a prerequisite in selecting the mixed effects is to also select the **mean center** option.
4. **Model form –** It has 2 options
   1. **Additive -** For this example we have selected the additive form
   2. **Multiplicative –** It is recommended when the KPI has high seasonality and/or when the media tactics are expected to have synergy effect between them. It is worth mentioning that the multiplicative form is a bit more complex than the additive form, however the platform takes care of this and provides us results like an additive model

**Correlation**

1. It allows us to check for correlation between the KPI & the independent variables and check for multicollinearity among the independent variables themselves
2. The platform also offers the options of selecting the time-period for which the correlation check is required and a feature to set a threshold for correlation %, where the platform highlights a tactic which crosses the set threshold

**Create transform measures**

1. This tab contains a list of transformations usually used in MMM (Marketing Mix Modeling)
2. It is a simple click and transform kind of an approach, which does not involve any manual calculation
3. We simply pick the transformation of interest and select the tactic to the transformed
4. We can look at this in detail, with an example, as we build a model shortly

**Switch to Auto mode**

1. Switching to this mode enables the platform to create multiple Ad-stock transformations of a/multiple variables and returns the transformation with the highest correlation with the KPI
2. It saves time on manually creating transformations; however, the platform only applies Ad-stock transformations now

**Variable selection**

1. Scrolling down further we have 2 panes, the pane on the left contains the list of variables ingested in the platform and the pane on the right is where the variables to be tested in the model are moved to. **(Halloween, Update\_2ndOct)**
2. To select the variables, we shall simply check the variables of interest and move to the selected measures area (right pane) or search for the variables in the search bar (on top of the left pane) and move it to the selected measures area
3. The list of variables in the selected measures pane will be tested in the model

**Priors**

1. Before running the model, we have the option of setting priors I.e., informing the model of empirical evidence of relation between the independent variables and the KPI.
2. We can provide co-efficient and standard deviation to a tactic(s). This will give us the out of interest, contribution, which we could use to calculate performance measures like effectiveness and ROI (Return on Investment).
3. We shall look at this in detail, with an example, while we build the model

**Model output Page**

The model output page displays the results of the given iteration. Let us look at features available in it.

**Select model –** It is a list of all the saved and unsaved iterations of the model. We shall scroll through it to view any iteration of interest for future reference. To save an iteration we can simply provide a name in the “**Enter name to save model**” section and click on save.

Below we have the various tabs of results in the model output page. We shall discuss each of them in detail one after the other.

**Model fit**

1. **Model fit chart –** It consists of the predicted sales (red line), modeled sales (orange line) and blue columns representing residual (error i.e., difference between predicted and modeled sales). The chat also displays the model stats (RSQ & MAPE- Mean Absolute Percent Error), where the **acceptable range for a daily model is RSQ > 80% & MAPE < 20%**
2. **Model fit data –** It consists of the results a variable level, where the list of variables selected to be tested in the model, their respective coefficients along with its statical validation measures are displayed
   1. **Standard error –** it is a measure of uncertainty. The higher the value, the more uncertain is the coefficient of a given variable
   2. **T Stat -** A T-stat of +2/-2, depending on the nature of the variable, indicates that the coefficients are significant
   3. **P value –** A P value of <=0.05 indicates that the coefficients are significant
   4. **VIF (Variable Inflation Factor) - it** is a measure to indicate multi-collinearity among the independent variables. A value above 25 highlights the presence of multi-collinearity.

**Model iteration**

1. Now that we have viewed the results of the selected base variables, we shall try a**dding a media variable** and see how the model reacts.
2. **Update Model –** In order to edit or add new variables to a current iteration a user must click on the **update model** button that is available on the top right of the model output window. This takes us back to the model setup window.
3. **Adding media variable –** A raw media variable cannot be included in the model because the carryover impact will be accounted for in the model. To apply that lets transform the media variable **(Search\_Apple\_Imp)** 
   1. **Transformation –** As mentioned earlier, select the transformation of interest **(select gamma)** and the variable to be transformed and click on continue.
   2. Here, we input the parameters for transformation. In the case of gamma transformation. There are 3 parameters,
      1. **Decay –** Indicates the % of impact that is being carried over to the next day **(input 0.4)**
      2. **DOF (Degree of Freedom) - Provide** the build aspect to the transformation, for example, a DOF of 1 indicates that the impact is immediate, whereas when the DOF is 2 it indicates that the real/majority of the impact occurs from the 2nd day **(Input 2)**
      3. **Week -** (days in this case, as it is a daily model) indicates the period for which the impact is allowed to carryover to. For example, if week/day is 5 the carryover impact stops at the 5th day **(Input 5)**
   3. The platform also allows us to rename the transformed variable. It is suggested to add a meaningful suffix to the input name. For example, g425. Indicating gamma, decay, DOF & week/day respectively.
   4. There is an additional option of creating multiple transformations at the same time. By clicking on the “+” sign and adding as many new transformations as possible for the selected variable.
   5. To generate a new transformed variable, a user must click on save and only then click on generate measure.

Required variables to input:

1. Initially input base variables (Click on “Updates” and move all the variables to the right pane (selected measures Pane)

R-SQ 91.4 %; MAPE 46.0%

\*\*\*Mention about current RSQ & MAPE and proceed to adding one media variable\*\*\*

\*\* We added all the variables which belongs to updates because we see that most of the spikes in the model are driven by the updates. once we start adding the media variables KPI will be distributed across all inputted variables

1. Select **FB\_Ap\_June** 🡪 apply gamma transformation decay (0.2), DOF (1), Weeks (7)
2. \*\*\*Mention how the RSQ & MAPE have slightly changed and proceed to adding another media variable\*\*\*
3. Select **YT\_Ap\_June 🡪** apply gamma transformation decay (0.3), DOF (1), Weeks (7)
4. **\*\*\***Talk about how the RSQ & MAPE and highlight how the **(FB\_Ap\_June)** media variable has taken up a negative coefficient\*\*\*
5. **\*\*\***Remove YT data & highlight how FB has taken up a positive coefficient now but also highlight that the RSQ has decreased\*\*\*
6. Now we shall use the positive coefficient of FB as its prior and give it a 20% standard deviation & also include the YT tactic along with it.
7. Mention how the RSQ has improved from the previous iteration and both FB & YT have taken up a positive coefficient\*\*\*
8. This is one way of arriving at a coefficient and these coefficients could be validating the contribution % with the spend share % of each of the tactic and make sure that it makes business sense along with a good model fit.

**Input new transformed variable**

1. Now that we have created the new transformed variable. We shall go to the model output page and click on update model to select the new variable to test it in the platform & run the model including it
2. \*\*\*Mention how the RSQ & MAPE have changed and proceed to adding another media variable\*\*\*
3. \*\*\*Select (**Youtube\_Imp\_1**) & create a **gamma** transformed variable with parameter decay = 0.4, DOF = 1 & week/day = 3\*\*\*
4. \*\*\* repeat the process for including the new transformed variable in the model & run model\*\*\*
5. **\*\*\***Talk about how the RSQ & MAPE has improved.
6. \*\*\*Another Scenario where media variable takes -ve coefficient and how to deal with it\*\*\*
7. \*\*\*Select (**FB\_AEO\_Imp\_2**) & create a **gamma** transformed variable with parameter decay = 0.4, DOF = 1 & week/day = 5\*\*\*
8. **\*\*\***Talk about how the RSQ & MAPE and highlight how the **(FB\_AEO\_Imp\_2\_g415\_)** media variable has taken up a negative coefficient\*\*\*
9. **\*\*\***Remove (**Search\_Apple\_Imp\_g427\_ )** data & highlight how (**FB\_AEO\_Imp\_2\_g415)** has taken up a positive coefficient now but also highlight that the RSQ has decreased\*\*\*
10. Now we shall use the positive coefficient of(**FB\_AEO\_Imp\_2\_g415\_)** as its prior and give it a 10% or 20% standard deviation & also include the (**Search\_Apple\_Imp\_g427\_**)along with it.
11. \*\*\*Mention how the RSQ has improved from the previous iteration and both FB AEO has taken up a positive coefficient\*\*\*
12. This is one way of arriving at a coefficient and these coefficients could be validating the contribution % with the spend share % of each of the tactic and make sure that it makes business sense along with a good model fit.

**Step 2.2 - Model walkthrough (Response curves & Simulation)**

Now that we have run a few iterations and arrived at a result that makes both statistical and business sense, we can build a response curve for each of the media tactics and use that as a building block to run various budget optimization scenarios.

Before we can look at how we build a response curve, we shall first discuss what a response curve is and what are the key results we infer from it.

**Response curve**

1. It is a mechanism used to measure responsiveness (increase in revenue) per unit spend increase
2. It provides an optimal range of expenditure within which maximum profit is obtained
3. It has 3 key points of results
   1. **Point of current spending -** The current level of expenditure of a given tactic and its associated performance
   2. **Point of marginal peak –** The point at which the response (increase in revenue) per unit spend increase is maximum
   3. **Point of max ROI –** The point of expenditure at which maximum profit is attained
4. Point B to C is referred to as the optimal zone, which is the recommended zone of spending for a given tactic

**S-Curve transformation**

1. The Gamma transformation applied so far captures only the lag/ carry over effect that is observed in a media tactic
2. However, to consider the non-linearity and saturation aspects of the media tactics, the S-Curve transformation is applied on top of the previously existing gamma transformation.
   1. **Non-Linearity -**
   2. **Saturation -**
3. The S-Curve transformation could be applied like any other transformation, where we reach the model setup page and select the transformation type and measure of interest in “**Create transform measure**,” In this case its S-Curve & the gamma transformed media variable, respectively
4. A S-Curve transformation has 2 parameters alpha (shape) & beta (scale). To arrive at anchor points for both the parameters we make use of an excel template

**S-Curve template**

1. The template requires one inputs from a user, which could be obtained from the model output page in the platform
   1. **Transformed media data –** This could be obtained from the data export feature. Which exports an excel document contains the raw and transformed data information used in the project
2. Inputting the data mentioned would provide a default value for alpha & beta, which we shall use as an anchor point for options of S-curve transformation for a given media variable
   1. Alpha - (Mean/ SD) ^2 *(Mean and Standard Deviation of the transformed media data)*
   2. Beta - (SD^2)/Mean
   3. \*\*use **(Youtube\_Imp\_1\_g413\_)** data\*\*
3. We test for **different levels of saturation at the average current spend level** (25%, 50%, 75%) by changing alpha and beta individually for each of the saturation levels
   1. \*\*use **50% saturation level by changing alpha** in the model\*\*
4. We shall test each of these alpha and beta for the given variable and keep the transformation which provides the best fit
5. In the case of usage of priors in the original gamma transformed variable we shall make use of the “Coeff (based on curve)” as the prior for the S-Curve transformed variable, which is a scaled coefficient of the original coefficient used
6. Similarly, each of the media tactics should be tested for the various options of alpha and beta
7. The measure properties for each of those variables should be updated as well

**ROI**

1. Before building the response curves, the user is first required to generate the ROI for media tactics, which could be done in the ROI tab in the model output page
2. Where the initial step would be to upload spend information
3. We have the option of selecting the granularity at which we want to upload the spend data in (dimension & time granularity)
4. Then we shall click on “Save” in Update ROI parameter to generate the ROI for each of the media variables
5. This will display charts and a consolidated table of raw support, spend, Raw CPP, spend, efficiency & ROI

**Response curves**

1. Now we shall generate the response curves in the “Response curves” tab of the model output page
2. We have the option of generating response curves at both measure level and aggregate level
   1. Measure level is when we need curves at an individual variable level
   2. Aggregated option is used when we require a single response curve for a group of variables aggregated together
      1. This aggregation could be done in the classification section, where a group of variables, where we need an aggregated response curve for, shall be classified within a single folder
3. Final we shall click on the “Generate” button to generate the response curves
4. Post which we shall save the iteration containing the response curves

----------------------------------------------------------------------------------------------------------------------------------