**LiME**

**Documentation**

**Contents**

* Data Preprocessing
* Data Validation
* Transformation
* Model Build and Performance
* Model Tuning
* Model Result Overview
* Build Response Curves
* Scenario Planner
* Saved Scenarios

**Introduction:**

The tool is a comprehensive suite for media mix modeling offering nine intuitive tabs for various functionalities. Users can construct MMM models, analyze, optimize, and visualize spends, contributions, and scenarios across multiple media channels.

**Login Credentials**

**Username:** admin

**Password:** password

1. **Data Preprocessing**

* Upload a CSV or Excel file with appropriately labeled column names and cleaned data. Ensure there are no redundant columns, and the data has been treated for null or missing values.
* This tab involves three key steps: Renaming and Classification, Selecting the Time Variable, and Base Price Selection. In Renaming and Classification, variables are renamed for clarity. The Select Time Variable step assists in identifying the correct DateTime variable formatted accurately for temporal analysis. The Base Price Selection allows for choosing the base price variable or generating it.
  1. **Renaming and Classification**

This section allows user to modify variable names and reorganize their categorization if necessary. If any variables require renaming for clarity or consistency, simply check the box next to the variable and input the new name. The bucket is initialized intuitively based on the variable name and might require changes. To modify the bucket, choose the desired option from the dropdown menu located below the variable name. Once complete, click on the "Update Changes" button and view the changes in the table generated afterward.

Note: Substitute spaces with underscores and ensure variable names are unambiguous.

* 1. **Select Time Variable**

In datasets containing various date time variables at different levels of granularity (such as year, week, and month), selecting the appropriate time variable is crucial when constructing a model. This choice significantly impacts the accuracy and relevance of the model's outcomes. By choosing the right time variable that aligns with the desired level of data aggregation. the model can effectively capture temporal patterns and provide more accurate and insightful results.

Please choose the relevant datetime variable from the dropdown menu. This selection will be utilized for subsequent analysis, capturing crucial temporal aspects within the dataset

* 1. **Base Price Selection**

The base price holds significant importance in analysing data for Fast-Moving Consumer Goods (FMCG) and Consumer Packaged Goods (CPG) clients. It serves as a fundamental reference point for assessing pricing strategies.

If the dataset includes a base price, select it from the dropdown menu. If it's unavailable, you can generate it by choosing 'Generate Base Price.' The base price will be calculated as a 4-week moving average of the non-promotional price.

* 1. **Update Changes**

Once you've finalized your adjustments, click on 'View Changes' to review the modifications reflected in the table. Confirm these changes, then proceed by clicking 'Update Changes' to save all the modifications made to the dataset. This ensures that the refined dataset is updated and ready for further analysis or modeling

1. **Data Validation**

This tab serves as a comprehensive platform for validating and analysing both media and non-media variables through relevant plots and exploratory data analysis. It comprises three key sections: the Media Channel Analysis, facilitating the validation of media variables; the Non-Media Variables Analysis, dedicated to understanding non-media factors; and an Exploratory Data Analysis section, providing a comprehensive univariate and bivariate examination of selected variables. Together, these sections offer a thorough understanding of the dataset, enabling users to derive valuable insights and make informed decisions based on the data's patterns and relationships.

* Select Dependent Variable (will be used for all following plots for reference)
  1. **Media Channels**

Please select a media channel for analysis and choose from impressions, clicks, or spends to generate a line plot. Additionally, a summary table will be generated, including metrics like CPM or CPC based on your selection. Once the data is reviewed and confirmed as accurate and in line with expectations, click 'Validate' to confirm the validity of the variable. You have the option to validate each variable individually or use 'Validate All' to validate multiple variables simultaneously, ensuring the accuracy and reliability of the data for further analysis.

Note: If any variable remains unvalidated, a warning will persist on subsequent pages, indicating the variables that have not been validated.

* 1. **Non-Media Variables**

Similarly, select and analyze non-media variables such as price, holidays, market trends, or COVID-related factors present in the dataset, if applicable. Evaluate these variables to understand their impact and significance within the dataset.

* 1. **Exploratory Data Analysis**

In this section, users can create a correlation heatmap by selecting variables using provided checkboxes. They have the option to choose specific variables or analyze all variables collectively in the heatmap. Moreover, two buttons enable in-depth exploratory data analysis (EDA):

'Generate Profile Report': This generates a pandas profiling report, offering detailed univariate analysis for all dataset variables.

'Generate s Report': This button initiates the creation of a comprehensive bivariate analysis report using the Sweetviz package. The report provides insights into relationships between variables and can be downloaded for offline use.

1. **Transformations**

This tab provides users with the capability to apply diverse transformations to variables, Users can opt for transformations like adstock and lag. They have the flexibility to choose a global range, applying it uniformly to all variables. Alternatively, they can adjust specific values for individual variables from the dropdown menu. Afterward, they can select the desired transformations conducive to their model-building process.

Adstock: Adstock is a transformation technique used in marketing analytics to account for the prolonged impact of past advertising or exposure on consumer behavior. It applies a decay function to model the diminishing but persistent effect of previous marketing activities, creating a weighted average over time.

Lag: lag transformation involves shifting or delaying a variable's values by a specified number of periods. This technique helps capture temporal dependencies or patterns by aligning the variable's past values with its current or future values, allowing the model to account for time-based relationships and trends in the data.

**Note:** Adstock transformation is specifically applicable to media variables, lag transformation can be applied to both media and non-media variables, excluding price variables.

Adstock is applied in intervals of 0.05 within the specified range, while lag is applied in intervals of 1 within the specified range.

If user don't want to apply any transformation for variable just select 0 as input for both lag and adstock

Steps to perform Transformation:

* Select the global range for adstock and lag transformations via a slider.
* Modify individual variable ranges by selecting options from the dropdown slider if needed.
* Confirm the selections through the summary table generated below.
* Apply the chosen transformations by clicking the 'Apply Transformations' button.
* After applying transformations, users can select transformed variables for each variable based on their corresponding correlation values from the table.
* Upon completion of selections, click on "Submit Selection" to save the chosen variables.
* If no explicit selections are made, the default setting chooses the top three transformations per variable, considering their correlation with the target. Click on "Submit" to finalize the selections.

1. **Model Build and Performance**

In this tab, users can construct models based on combinations generated by transformed variables from the previous page. They have the option to assess the top 10 models, visualizing various performance metrics. These visual summaries include details such as the model summary showcasing metrics like R-squared, adjusted R-squared, and coefficients of variables. Additionally, users can view an actual versus predicted plot and examine the Variance Inflation Factor (VIF) values for the selected model. This comprehensive overview aids in evaluating model performance and understanding the influence of variables on the model's predictions.

Steps to follow:

* + Use the 'Create Combinations of Variables' button.
  + Variable combinations ensure each model iteration contains impressions or transformed clicks from every channel.
  + Choose the number of iterations and click 'Build Models.'
  + View a table displaying R2, AdjR2, and MAPE for the top 10 valid models.
  + Click on any model to view its detailed results and analysis
  + Model result is displayed for test data
  + Click on “Save this model to tune” and type a name and enter to save the model.

Note: Users can save multiple models and fine-tune them on the subsequent page.

**Valid Model:** After each iteration of model building, the model will be saved if the following criteria are met: all coefficients of the channels are positive, and at least 80% of the variables exhibit statistical significance.

**Variable Combinations:** Creating variable combinations ensures that each model iteration includes either transformed impressions or clicks from every media channel. This comprehensive approach explores various combinations of variables, specifically focusing on clicks or impressions per media channel within each iteration. This strategy enhances the model's capability to capture diverse interactions between media channels.

**Time to complete iterations:** it depends on the size of the data and combinations of variables created.

1. **Model Tuning**

In this tab, users can fine-tune the selected model from the previous page by incorporating event flags. They have the option to add specific event flags, select the added flags, and rebuild the chosen model while considering these flagged events. Additionally, users can access all the performance metrics and visualizations presented in the model build and performance page. This comprehensive approach allows users to refine the model by integrating event-related information, facilitating a deeper analysis of the model's performance in relation to these significant occurrences or events within the dataset.

**Event flags** in model tuning are instrumental in quantifying the impact of occurrences or events within a dataset. By specifying the time duration and assigning a name to these events, users create markers or indicators that highlight specific periods of interest, such as promotional campaigns, holidays, market changes, or any other significant events. These flags serve as reference points within the dataset, allowing the model to account for and analyze the influence or impact of these events on the variables and model outcomes. Incorporating event flags enables the model to better capture the effects of these occurrences, leading to more accurate and insightful predictions or analyses.

**Steps to follow:**

Select the duration of the event using the calendar and specify if it repeats annually or not from the dropdown menu, repeat annually' indicates that the flagged event recurs yearly. It suggests that the specified event's impact or occurrence repeats on an annual basis. After confirming these settings and visualizing the flag's impact through the provided visualization, then save the flag.

Users can choose a saved flag to apply to the selected model from the check box. Upon selection, clicking "Build model with selected parameters and Flags" applies the chosen settings, showcasing the results of the tuned model.

Note: No need to create event flags everytime a user selects different model for tuning

To save the tuned model, select "Save this model to tune," and provide a name for it. This action stores the model permanently, allowing easy access for further analysis or tuning without the need to reconfigure from the beginning, even if you restart the application or return later.

1. **Model Result Overview**

This tab provides an in-depth comparative analysis of spending and contributions across various channels. Users can explore these insights through multiple visualizations such as pie plots, illustrating sales and contributions, and stacked bar plots showcasing the breakdown every week. The comprehensive nature of these visualizations allows users to gain a nuanced understanding of channel performance.

1. **Build Response Curves**

This tab visualizes response curves that depict the relationship between contributed revenue and spending for each channel. Users can interactively adjust curve parameters to achieve a better fit based on their analysis of

The sigmoid curve used:

A math equation with numbers and symbols

Description automatically generated

* x: The input variable represents the independent variable.
* K: The saturation level or the maximum value that the curve approaches as x approaches infinity.
* b: Governs the steepness of the curve. Larger values of 'b' result in a steeper curve.
* a: Determines the rate of change of the curve. Higher values of 'a' mean a faster initial rise or fall in the curve.
* x0: Represents the midpoint of the curve along the x-axis. It's the value of 'x' at which the curve's slope is the steepest.

If a user wishes to adjust any parameter of the curve, they can use the respective widgets to make modifications. After tweaking, by clicking "Update Parameters," these adjusted values will be applied for optimization specifically tailored to that channel’s analysis.

1. **Scenario Planner**

In this tab, users can conduct spend optimization and simulation across multiple channels. They have the flexibility to either optimize all channels collectively or selectively choose specific channels for optimization by allocating a new budget. The optimization process relies on response curve fitting and employs an algorithm to explore various combinations, determining the most effective spending across all channels. This optimization aims to maximize sales or revenue by finding the optimal allocation of resources among the available channels, ensuring an efficient and effective utilization of the budget.

* **Enter Updated Spend Changes:** Input the revised percentage change in total spends/budget. To optimize for the same budget, enter 0. Choose the channels you wish to optimize.
* **Channel Selection**: Optimize specific channels by selection or optimize all channels with the "Optimize All Channels" widget.
* **Optimization Process**: Click the "Optimize" button to derive optimized results based on the selected spend adjustments.
* **Save Optimized Scenarios:** Save the results in the "Save Scenario" section by assigning a name to the scenario for future reference. Users have the capability to simulate multiple scenarios, allowing them to test various budget allocations or different optimization strategies. After creating and saving these scenarios, they can analyze and compare them in the dedicated "Saved Scenario" page.

**Scenario Planner Situations:**

1. Scenario Planning for same budget/ old budget: To optimize for the old budget just enter 0 as the input in the "Enter Updated Spend Changes" section

2. Fixed budget allocation for the selected channel(s) while optimizing the remaining channels: Uncheck the preferred channels you wish to maintain with a fixed budget, then select the remaining channels for optimization. Click 'Optimize' to proceed with the selected settings.

3. Reduced Budget: If users intend to optimize for a reduced budget, they can input the reduced budget amount using negative percentage values. For instance, if the new budget is 10% less than the current budget, they would input "-10" as the value.

1. **Saved Scenario**

In this tab, users can both view and download saved scenarios, accessing comprehensive details such as NRPU, ROI, MROI, and spends per NRPU. Multiple scenarios, tailored to different budgets, can be saved on the scenario planner and reviewed individually on this page for thorough analysis.