**RTM/OSS Common Functions**

The folder structure and function:

1. **Data** folder. Model datasets used for validation purpose.

* Testing\_Data.xlsx
* SKU-standardisation.xlsx

1. **Code** folder

* **Feature\_Creation.** The following functions are standalone functions and can be used on any dataset with respective required columns.

1. (stats): Function to get statistics (max, min, mean, etc.) of input variables from the original dataset giving output as a new dataframe.
2. (time\_trend\_ratio): Function to add time trend ratio feature to original dataset. The time trend ratio implies the ratio of current week variable and user input previous week variable.

* **EDA.** The following functions can be used for analysis of data using visualizations. Different functions output different plots based on user inputs.

1. (categorical\_dist): Function to get period level categorical distribution plots based on user input.
2. (chron): Function to get the number of instances from the dataset where there are continuous days of data missing and total number of days missing between start date and end date for each item.
3. (missing\_analysis): Function to get the summary of missing values, -ve values, or zero values for the variable input by the user. The output is the missing percentage value for each item in the granularity. e.g. missing sales values as percentage for each Corporate Segment.

* **Data\_Cleaning**. The following are standalone functions.

1. (outliers): Outlier removal function from the dataset based on user input threshold. The output is the original dataset without outliers.
2. (sku\_standardize): Function to get top three matches for first string list from the second string list based on fuzzy scores. The user inputs two lists of strings and get output as a dataframe with top three matches for each string in list one.
3. **Output** folder

* **Feature\_Creation** folder
* **EDA** folder
* **Data\_Cleaning** folder