|  |
| --- |
| TIGER ANALYTICS |
| Incubation Project:  Market Mix Modeling for a US Multinational Company |
|  |

# Background:



Our client (Company A) is an American multinational food, snack, and beverage corporation. It oversees the manufacturing, distribution, and marketing of its products

The client has the following objectives:

* Understand which media outreach has been most successful for their campaigns across different media (Digital/TV/Out of Home) in a given region/location.
* Further, within each media, analyse and identify the most effective platforms. For example, within Digital media, which platform (YouTube, TikTok, etc.) was the most effective campaign.

# Data:



Company A has shared the following data for the project:

* [Media Data](https://docs.google.com/spreadsheets/d/1FKZv-1sFitSfAvOlHEZFAHlSBSgFSKrP/edit?usp=drive_web&ouid=109077973923084780405&rtpof=true)

Media execution data across different vehicles like TV, Digital, and OOH for a Brand B1 of Company A

* [Macro Data](https://docs.google.com/spreadsheets/d/1Z_hKGVXmKJkZozSOoKOTdIBVy8UINFTu/edit#gid=1314853264)

Macroeconomic & Covid-related data

* [Nielsen Data](https://docs.google.com/spreadsheets/d/1jGx6nqcJ9sBqY1vsOFnK-PrCcnWkSLMv/edit?usp=drive_web&ouid=109077973923084780405&rtpof=true)

Sales & Shipment (Internal trade data) for Brand B1

All the files have a clear and detailed description of the dataset (including all the fields present).

**Deliverables:**



Data Preparation

Demonstrates:

* Provide a report for data sufficiency, sparsity, and anomalies in each data source if any

Data exploration and feature engineering

Demonstrates:

* Merge the required data sources to build a master dataset
* Understand the trend of sales, volumes, impressions, price, distributions and Y-O-Y change of all drivers.

Build Optimal models to identify which media stream contributes to the growth

Demonstrates:

* Perform appropriate data transformation
* Identify the right metrics to interpret the model and select the right variables
* Justify the models based on
  + Model Performance
  + Drivers
* Model 1: Identify drivers of the sales/volume of a given brand
  + DV: Volume for each media across all the stores in a given region
  + IDV: price, distributions, discounts, impressions, GRP, etc.
  + Model Specification: Linear Regression: Drivers of the sales/volume of a given brand
  + Analysis of Drivers: Trend and YOY change w.r.t all drivers
  + ROI (Return of investment): Aggregated at platforms, e.g. digital or TV or out of home
  + Outcome – ROI: Effective ROI where ROI is high-   
    volume\*price = Revenue; Revenue/ Spends = ROI
* Model 2:
  + DV: Aggregated volume(sales) contributions of each media from across stores from the primary model.
  + IDV: Media variables depending on the platform (Example: GRP for TV, impressions for digital)
  + Model Specification: Zero intercept Model
  + Outcome – ROI: Split across different platforms within a given media-   
    volume\*price = Revenue; Revenue/ Spends = ROI

Calculate ROI for the media

* ***Outcome 1***: Calculate the ROI (return of investment) for each aggregated media used in the model. Example: Digital or TV or out of the home.
* ***Outcome 2***: Calculate the effective ROI (return of Investment) split across different platforms within a given media.

**Note**: Volume\*Price = Revenue; Revenue/ Spends = ROI

## 

## Deliverable Best Practices

* Structured code base with a few tests
* The submitted code should satisfy coding standards and the `QC` tests should all pass
* Share your final codes along with EDA and modeling outputs generated with mentors

# High Level Project Plan:



* Week 1:
  + Concept understanding: Linear Regression/Zero - intercept model
  + R basics
* Week 2 & 3:
  + R advanced
  + Market Mix modelling: s-curve transformation / ad-stock / ROI / Elasticity calculation
* Week 4 & 5:
  + Data Understanding and data preparation data
  + Data cleaning
  + Exploratory data analysis
* Week 6 – 9:
  + Model 1: Feature engineering / Model building
  + Model 2: Feature engineering / Model building
  + ROI (Return of Investment) calculation
* Week 10 - 11:
  + Result generation
  + Results presentation

# Useful Resources:



**R:**

* Udemy course(mandatory): <https://tigeranalytics.udemy.com/course/r-programming/>
* R basics: <https://iqss.github.io/dss-workshops/R/Rintro/base-r-cheat-sheet.pdf>
* Cheat sheet: <https://www.rstudio.com/resources/cheatsheets/>
* data table: <http://www.trutschnig.net/Slides_Data_Table_Steinmassl_Eibl_Burgstaller.pdf>
* dplyr: <https://cran.r-project.org/web/packages/dplyr/vignettes/dplyr.html>
* Dates: <https://www.r-bloggers.com/2013/08/date-formats-in-r/>
* String: <https://bookdown.org/f_lennert/introduction-to-r/stringregex.html>

**Zero intercept model:**

* <https://rpubs.com/aaronsc32/regression-through-the-origin>

**Understanding of Market Mix Models (MMX): [please look into the Resources folder]**

* MMX overview
* Contributions calculation
* Nielsen Data
* Ad-stock