Background

Pernalonga, a leading supermarket chain of over 400 stores in Lunitunia, sells over 10 thousand products in over 400 categories.  Pernalonga regularly partners with suppliers to fund promotions and derives about 30% of its sales on promotions.  Until recent experimentation with personalized promotions, most of Pernalonga’s promotions are chain-wide promotions.

The beer category is promoted regularly in Pernalonga’s stores.  Mahou San Miguel, a Spanish brewer, sells three San Miguel beer products in Pernalonga’s stores and regularly partners with Pernalonga to promote its products via weekly flyers and in-store displays.  Mahou San Miguel also employs other marketing vehicles such as email, web (display banners and paid search), and traditional media (TV and Radio).

Problem

Your analytics consulting firm was selected by Mahou San Miguel (the client) to verify the effectiveness of its promotions and marketing partnership with Pernalonga.  The client is interested in identifying promotion and marketing activities that drive significant incremental sales for continuation into 2021.  The timing is perfect, your Data Science team recently concluded a recommender systems project and a pricing project with Pernalonga, so you are well-versed with Pernalonga’s data.

Your Data Science team with its experience in marketing mix modeling in the Lunitunia market also have the following insights:

* + Seasonality of the Beer category
  + National holidays in Lunitunia in 2016 and 2017
  + TV advertisements have an 8-week half-life, and GRPs map into target audience 2+ Reach via the following formula

Reach=0.95(1−e−0.020GRP)Reach=0.95(1−e−0.020GRP)

* + Radio advertisements have a 4-week half-life, and GRPs map into target audience 2+ Reach via the following formula

Reach=0.90(1−e−0.025GRP)Reach=0.90(1−e−0.025GRP)

With the above insights and supplemental transaction and product data from Pernalonga, you are ready to demonstrate your capabilities in marketing analytics and help the client identify promotion and marketing vehicles that drive significant results.

Available Data

The file [Pernalonga.zip](https://canvas.emory.edu/courses/84073/files/5135847/download) and [SanMiguel.zip](https://canvas.emory.edu/courses/84073/files/5531002/download) contains the following tables:

1. transaction\_table.csv contains transaction history in 2016 and 2017 for close to 8,000 customers
   * cust\_id – Customer ID
   * tran\_id – Transaction ID
   * tran\_dt – Transaction Date
   * store\_id – Store ID
   * prod\_id – Product ID
   * prod\_unit – Product unit of measure: CT for count and KG for kilograms
   * prod\_unit\_price – Unit price of the product
   * tran\_prod\_sale\_qty – Quantity/units of the product in the transaction
   * tran\_prod\_sale\_amt – Sales amount for the product before discounts in the transaction
   * tran\_prod\_discount\_amt – Total amount of discounts applied to the product in the transaction
   * tran\_prod\_offer\_cts – Total number of offers on the product resulting in the total amount of discounts in the transaction
   * tran\_prod\_paid\_amt – Amount paid for the product after discounts are applied in the transaction
2. product\_table.csv contains the product to subcategory and category mapping and descriptions for about 11,000 products
   * prod\_id – Product ID
   * subcategory\_id – Subcategory ID
   * category\_id – Category ID
   * sub\_category\_desc – Subcategory name (in Portuguese)
   * category\_desc – Category name (in Portuguese)
   * category\_desc\_eng – Category name (in English)
   * brand\_desc – Brand of the product, including NO LABEL and PRIVATE LABEL
3. transaction\_table\_supp.csv contains supplementary transaction history for San Miguel products in 2016 and 2017 with same fields as transaction\_table.csv
4. product\_table\_supp.csv contains supplementary records for San Miguel products with same fields as product\_table.csv
5. promo\_ad.csv contains the promotion and advertising activity records
   * tran\_wk – the date of the Sunday representing the week
   * vehicle – the promo or advertising vehicle
   * amount – the amount of advertising in units; a “1” means the promotion (Flyer or Store Display) is on for the week
   * unit – the unit indicated by the amount
   * prod\_assoc – the Product ID of the product featured in the promo or advertisement; “ALL” means all San Miguel products
6. seasonality.csv contains the seasonality index for products in the Beer Category
   * tran\_wk – the date of the Sunday representing the week
   * seas\_index – a number indicating seasonality index for the week
7. holiday.csv contains a list of national holidays celebrated in Lunitunia in 2016 and 2017
   * tran\_wk – the date of the Sunday representing the week
   * holiday – the name of the holiday; a “Pr” prefix before a holiday indicates a week before the holiday

Note that customer, store, product and promo and marketing information beyond what is available above are not provided.

Grading

Professional data scientists are expected to be domain experts implementing sound mathematical models using robust and reusable computer programs.  Your work will be graded according to the following criteria:

* Integration of domain knowledge/practicality into solution (20%)
* Creativity and mathematically sound application/execution of chosen technique/model (25%)
* Robustness and efficiency of solution/code (25%)
* Report and presentation flow, delivery, and defense (20%)
* Peer evaluation (10%)

Reports

The project written report must include:

1. model diagnostics to justify the validity of the models
2. decomposition of weekly sales volume for each San Miguel product into the following components:
   * Base (includes Seasonality and Holidays)
   * Shelf Price
   * Discount
   * Flyer
   * Store Display
   * Email
   * Web Display
   * Paid Search
   * TV
   * Radio
3. recommendations on which promo (on which product) and marketing activities to continue in 2021