Walmart Labs – Data Challenge

Thank you for your interest in Walmart Labs.

As we move forward with your first round of interview, we would like you to share your thoughts and approach to any of the following retail problems:

1. How does customer Engagement change over time?
   a. Do customers spend more or less
   b. What categories are growing or shrinking with customers
   c. How might we reengage customers within the store
2. Which Demographic factors appear to affect customer engagement? Especially in certain categories.

You may choose any of the above topics or may choose to address your own problem statement.

Please provide a presentation/Summary (ppt, pdf), along with your code, covering:

- EDA Findings, Algorithmic techniques/Data Logic
- Think around business constraints that you would consider
- Clearly state business assumptions made (if any)
- Lastly comment on what would your next steps could be

We'd prefer to have your response by **5th February'20 12:00 pm**. However, let us know if you have any constraints and we can work around them.

We will carefully review your submission and come back to you with the decision of the data challenge, which is the first round. The shortlisted data challenges will proceed to the web interview round with panelists, where we may discuss the submission and technical details in the application of the algorithm/logic.

We are excited to talk to you and look forward to receiving your response.

Please reach out to us for any questions that you may have.

Data Set Description:

This Data set contains 100 households.

Following are the datasets:

1. Household Demographics

Variable	Description
HSHD_NUM	Uniquely Identifies each Household
Loyalty_Flag	Indicates if household is loyal to a brand
Age_Range	Age range of the buyer in that household
Marital	Marital status of the buyer
Income_range	The household earnings per annum
Homeowner	Indicates if the buyer owns or rents a house
Hshd_composition	Details of people in the house
HH_size	Number of people in the house
Children	Number of children in the household

2. Transaction Data

Variable	Description
Basket_Num	Uniquely identifies a purchase occasion
HSHD_NUM	Uniquely Identifies each Household
Purchase_Date	Date on which purchase was made
Product_Num	Uniquely identifies each Product
Spend_Amount	Amount retailer receives from purchase
Units	Number of products purchased during that transaction
Region	Region in which store is located
Week_num	Week number of the purchase date
Year	Year in which purchase was made

3. Product Data

Variable	Description
Product_Num	Uniquely identifies each Product
Department	Groups similar products together
Commodity	Groups similar products at a more granular level
Brand	Indicates if product brand is private or national
NA	Flagged Y if item is Natural/Organic and N otherwise

The Data Set Overview:

HOUSEHOLDS

Hshd_num
Loyalty_flag
Age_Range
Marital_Status
Income_range
Homeowner_desc
Hshd_composition
Hh_size
Children

TRANSACTIONS

Hshd_num
Basket_num
Date
Product_num
Spend
Units
Store_region
Week_num
Year

PRODUCTS

Product_num
Department
Commodity
Brand_type (Private/National)
Natural/Organic Flag