Initial list of KPI to be considered for model building

Lets suppose today is 1st oct'23, approx. 1 month before Diwali'23, when this modelling pipeline will run every year, and generate the list of categories and products under it that needs to promoted, and also the type of promotions that should be used to maximize sales, revenue and profit i.e. achieve all business targets.

Recency analysis framework

Timeline for be considered for recency KPI creation:

1ST Jan'23

Time period for which data is considered for calculating the KPIs

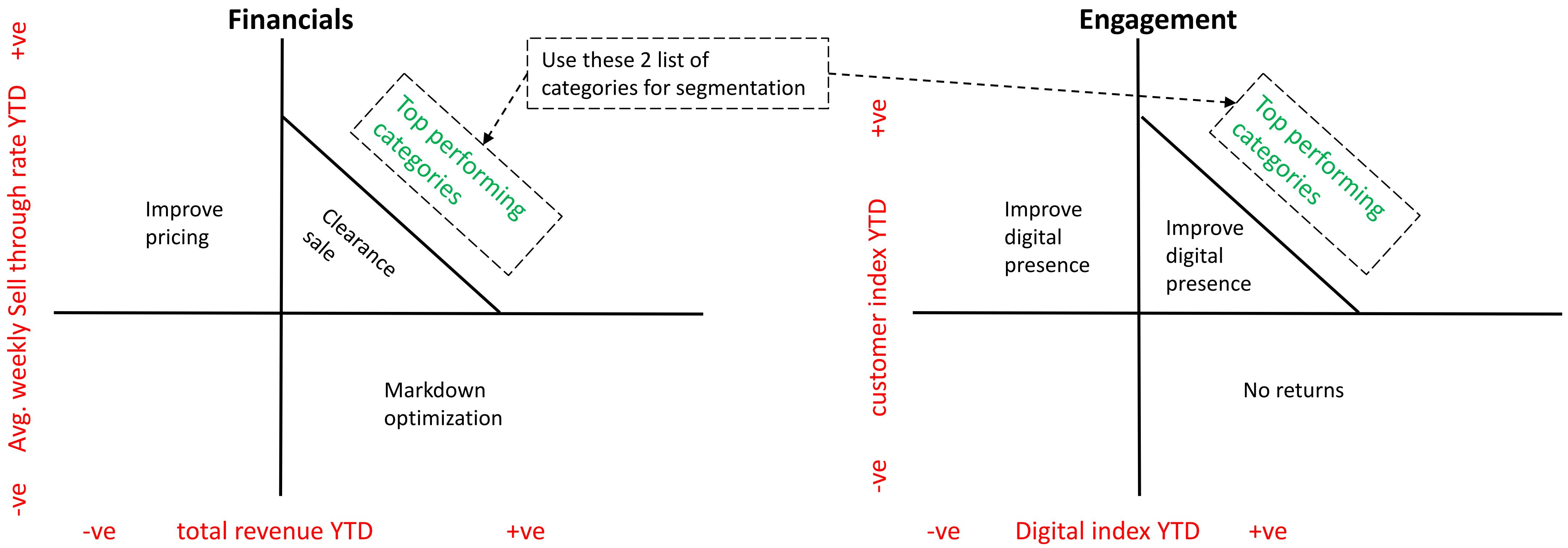
1ST Oct'23

Financial KPIs	Sales KPIs	Customer KPIs	Digital KPIs
• **Total revenue** = sum(price*quantity) • **AOV** = total revenue/no. of transactions • **Gross profit margin** = [(revenue – COGs)/revenue]	• **Total sales** = sum(units sold) • **Return rate** = (no. of returns/no. of units sold)*100 • **Inventory turnover rate** = COGs/[(BOP+EOP inv.)/2] i.e. COGs/avg. inventory • **Sell through rate** = [Total units sold/(total units sold + EOP inventory)] • **Avg. weekly sales** = Avg.(week wise sales)	• **Conversion rate** = (No. of conversions/No. of visitors)*100 • **Click through rate** = (No. of clicks/no. of impressions)*100 • **Customer ratings** = Sum of all ratings/No. of ratings • **Repeat purchase rate** = (No. of repeat customers/Total no. of customers)*100	• **Page viewed** = sum(page visits) • **Time spend** = avg(seconds) • **Social media reach rate** = (No. of shares/No. of posts)*100 • **Social media engage rate** = (No. of likes/No. of posts)

- ***All these KPIs will be calculated at category level at YTD (year to date) [for recency only]***
- ** Considering digital presence of the business exists so creating the customer and digital KPIs. These type of data (digital KPIs) is generally gathered from Adobe analytics and Google analytics clickstream data.
- *Would be mostly **considering revenue (and not profit margins)** since during the Diwali time period, business will be heavily running campaigns and promotions*

Customer index (CI) = w1*Conversion rate + w2*click through rate + w3*customer ratings
Digital index (DI) = w1*page viewed + w2*reach rate +w3*engage rate

Category filtering & bucketing through recency category behaviors & performance



Analyzed and bucketed ***'top performing categories' area list of categories*** and create the segments.

<u>List of categories that are overlapping:</u>

Strong recency financials and engagement: are the top performing categories as they have high sell through rate incremental, high revenue incremental, high digital presence incremental and high customer index incremental for consecutive 2 years

<u>List of categories that are present in either of them:</u>

Strong recency financials (only) : high sell through rate incremental, high revenue incremental for consecutive 2 years but doesn't have high / has low engagements

Strong recency engagement (only): high digital presence incremental and high customer index incremental for consecutive 2 years but doesn't have high / has low financial

Calculating historical KPIs & metrics [pre Diwali and in Diwali time periods]

Sales KPIs	Digital KPIs	Customer KPIs

- Calculate all the KPIs in the same fashion i.e. same formula (refer last slide) at <u>***week level***</u>
- ***Only use sell through rate, revenue from sales and financial KPIs and consider all KPIs for digital and customer***.
- Change the timelines for data consideration for KPI calc. as defined below

Calculate incremental change for all the KPIs for every category at week level on past 2 years data

<u>***Incremental = [Diwali week KPI value – Avg.(weekly values till -8 weeks from Diwali)]/Diwali week KPI value***</u>

-8 weeks from Diwali'22

a1 Time period for which data is considered for calculating the KPIs.

-1 week from Diwali'22

Diwali'22

b1 Calc. all of these KPIs of the Diwali'22 week

-8 weeks from Diwali'21

a2 Time period for which data is considered for calculating the KPIs.

-1 week from Diwali'21

Diwali'21

b2 Calc. all of these KPIs of the Diwali'21 week

Category	STR'22	STR'22'Diwali	STR'21	STR'21'Diwali	**STR Increment'22**	**STR Incremnet'21**
X	Avg. weekly STR (a1)	STR (b1) Diwali week	Avg. weekly STR (a2)	STR (b2) Diwali week	**(a1-b1)/a1**	**(a2-b2)/a2**

Category	Revenue'22	Revenue'22' Diwali	Revenue'21	Revenue'21' Diwali	**Revenue Increment'22**	**Revenue Incremnet'21**
X	Avg. weekly revenue(a1)	revenue(b1) Diwali week	Avg. weekly revenue(a2)	revenue(b2) Diwali week	**(a1-b1)/a1**	**(a2-b2)/a2**

Note: STR is sell through rate

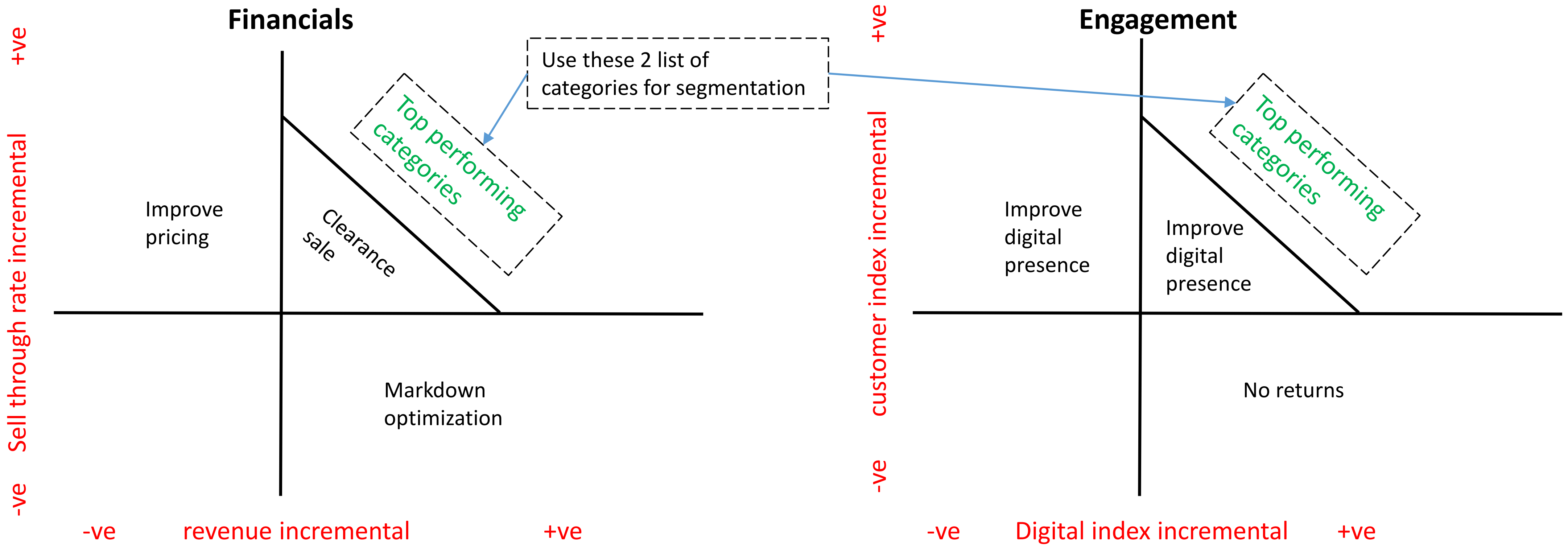
<u>***Category X: STR incremental = (STR incremental'22 + STR incremental'21) / 2***</u>

Similarly, avg. of the revenue incremental

Category	CI'22	C1'22' Diwali	CI'21	CI'21' Diwali	**CI'22**	**CI'21**
X	Avg. weekly CI (a1)	CI (b1) Diwali week	Avg. weekly CI (a2)	CI (b2) Diwali week	**(a1-b1)/a1**	**(a2-b2)/a2**

Similarly ***calc. for Digital index (DI)***

Category filtering & bucketing through historical category behaviors & performance



Only choose the analyzed and bucketed ***'top performing categories' area list of categories*** and create the segments. [2 years historical data taken]

List of categories that are overlapping:

Strong historical financials and engagement : List of categories that are overlapping at the top performing categories as they have high sell through rate incremental, high revenue incremental, high digital presence incremental and high customer index incremental for consecutive 2 years

List of categories that are present in either of them:

Strong historical financials (only): high sell through rate incremental, high revenue incremental for consecutive 2 years but doesn't have high / has low engagements

Strong historical engagement (only): high digital presence incremental and high customer index incremental for consecutive 2 years but doesn't have high / has low financial

Category segmentation for recommendation and analysis

Category	Category Recency bucket	Category Historical bucket	Category Segment
Category	strong recency financials and engagement, strong recency financials, strong recency engagement	strong historical financials and engagement, strong historical financials, strong historical engagement	A, B, C
Can be any one of these values or segments under every column of recency bucket and historical bucket			

Segmentation using business logics and heuristics for better explain ability and for making easy changes later

Segment A: (strong recency financials and engagement) && (strong historical financials and engagement)
This segment has the highest performing categories both historically (during Diwali time period) and also has been performing very consistently in the recent past.

Segment B: (strong recency financials) && (strong historical financials)
Segment B: (strong recency engagement) && (strong historical financials)
This segment although has high performing categories, has a risk of not performing if not promoted properly as in through the marketing channels to reach the right customer base. As, the engagement factor is missing in this segment.

Segment C: (strong recency engagement) && (strong historical engagement)
Segment C: (strong recency financials) && (strong historical engagement)
Customer in these segment of categories do indulge in lot, but business isn't able to squeeze the maximum juice out of this due to its financial instability. This means, business needs to plan better on pricing strategies and markdowns.

Category choosing summary & product filtering kick-off

<u>Segment A</u> tells that all the categories in it, ***performed the best*** in the ***Diwali week for the past 2 years(historical analysis),*** when compared to other weeks (till -8 weeks before Diwali of every year) ***both in terms of financials and engagement***, meaning, it has higher revenue generated that week, higher sell through and higher customer and digital engagements.
Also***, recency analysis*** tell that these set of categories show ***strong financial and engagement levels***. Thus taking into the consideration the current market trends as well.

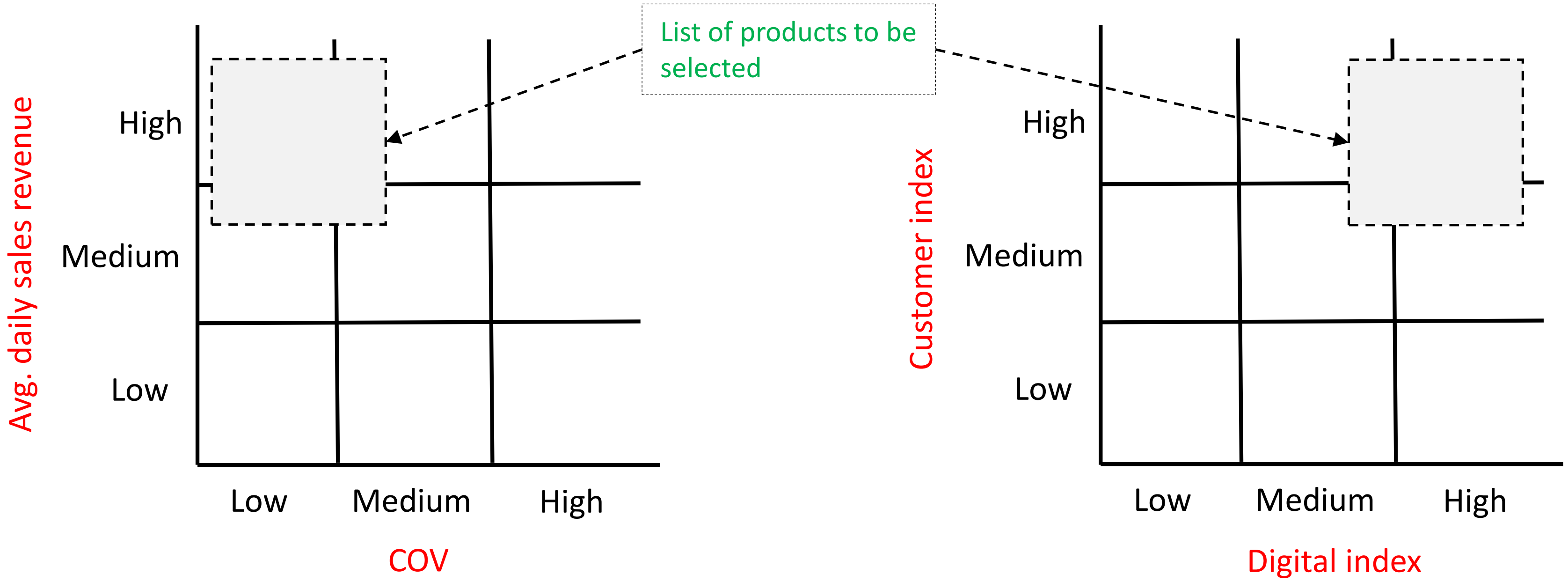
Filter out the list of products under all the categories, analyze the set of products that are to be promoted.

The category analysis and filtering was done at a weekly level (comparing other weeks with the Diwali week); now to analyze products under those top performing categories in the Diwali week, ***calc. all metrics at daily level*** (in the Diwali week – historical; current year recent daily level metrics of products)

Product level KPIs to be calculated

<table>
<tr><th>Financial KPIs</th><th>Sales KPIs</th><th>Customer KPIs</th><th>Digital KPIs</th></tr>
<tr>
<td>• Avg Daily revenue =sum (price*quantity)</td>
<td>• Avg. daily sales = Avg.(day wise sales)
• Pareto analysis = flag top 20% products that contributes to 80% sales revenue in that Diwali week
• COV (demand variability) = standard deviation of demand/Avg. daily demand</td>
<td>• Conversion rate = (No. of conversions/No. of visitors)*100
• Click through rate = (No. of clicks/no. of impressions)*100
• Customer ratings = Sum of all ratings/No. of ratings</td>
<td>• Page viewed = sum(page visits)
• Time spend = avg(seconds)
• DAU = Count (Distinct NewUsers)
• Bounce rate = % of users leaving after viewing one page
• Social media reach rate = (No. of shares/No. of posts)*100
• Social media engage rate = (No. of likes/No. of posts)</td>
</tr>
<tr><td colspan="4">• All these KPIs will be calculated at product-daily level
• Customer and Digital KPIs would remain, as business would be heavily running promotions and campaigns.</td></tr>
</table>

Product filtering & bucketing through historical product behaviors & performance



Note: Bucketing of these values and high, medium, low can be done through various statistical ways

In the last year (historical) Diwali week, **'List of products to be selected' area** are the set of products that were high performers.

List of products that are overlapping:

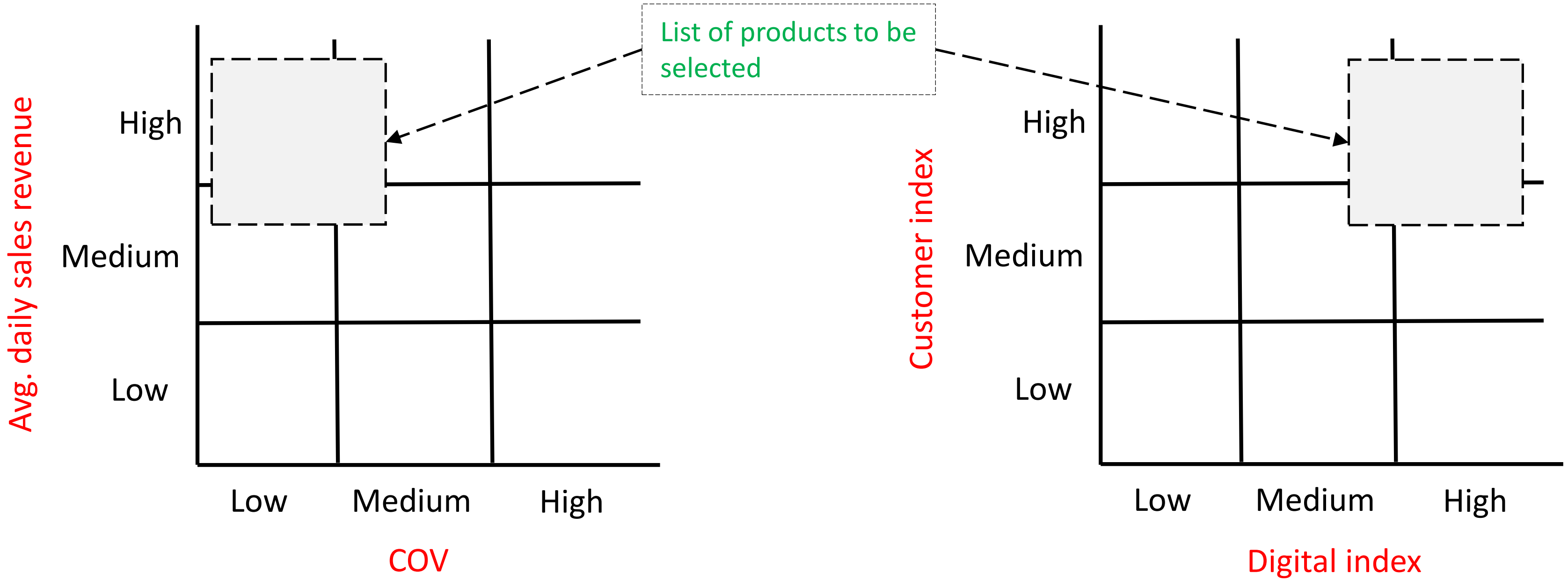
Strong historical sales and engagement: are the top performing products as they had high avg. daily sales revenue, consistently sold through out the Diwali week, high digital presence and customer engagements

List of products that are present in either of them:

Strong historical sales (only): had high avg. daily sales revenue and consistently sold through out the Diwali week [even with low customer engagements]

Strong historical engagement (only) : had high customer engagements, but low sales [means people had needs, may be didn't got fulfilled due to multiple factors like pricing, out of stocks or delivery issues etc.]

Product filtering & bucketing through recency product behaviors & performance



Note: Bucketing of these values and high, medium, low can be done through various statistical ways

In the recent last 14 days from running this pipeline , **'List of products to be selected' area** are the set of products that were **highly performing** thus **capturing recent upward trending products**.

List of products that are overlapping:

Strong recent sales and engagement: are the top performing products as they are showing high avg. daily sales revenue, consistent sales recently, as well as, high digital presence and customer engagements recently

List of products that are present in either of them:

Strong recent sales (only): having high avg. daily sales revenue and consistent sales recently [even with low customer engagements]

Strong recent engagement (only) : having high customer engagements, but low sales [means people had needs, may be didn't got fulfilled due to multiple factors like pricing, out of stocks or delivery issues etc.]

Category segmentation for recommendation and analysis

Product	Product Recency bucket	Product Historical bucket	Product Segment
Product ID	strong recency sales and engagement, strong recency sales, strong recency engagement	strong historical sales and engagement, strong historical sales, strong historical engagement	A, B, C
Can be any one of these values or segments under every column of recency bucket and historical bucket and accordingly final segment will be decided			

Segmentation using business logics and heuristics for better explain ability and for making easy changes later

Segment A: (strong recency sales and engagement) && (strong historical sales and engagement)
This segment has the highest performing products both historically (during Diwali week) and also has been performing very consistently in the recent past.

Segment B: (strong recency sales) && (strong historical sales)
Segment B: (strong recency engagement) && (strong historical financials)
This segment although has high performing products, has a risk of not performing if not promoted properly as in through the marketing channels to reach the right customer base. As, the engagement factor is missing in this segment.

Segment C: (strong recency engagement) && (strong historical engagement)
Segment C: (strong recency sales) && (strong historical engagement)
Customer in these products do indulge in lot, but business isn’t able to squeeze the maximum juice out of this due to its financial instability. This means, business needs to plan better on pricing strategies and markdowns. Since it is picking up pace in the recent past, can be change that some products can perform.

Final category wise product filtering and recommendation for campaign, promotions and analysis

Category	Product	Category Recency bucket	Category Historical bucket	Product Recency bucket	Product Historical bucket	Category segment	Product Segment
Category	Product ID	strong recency financials and engagement, strong recency financials, strong recency engagement	strong historical financials and engagement, strong historical financials, strong historical engagement	strong recency sales and engagement, strong recency sales, strong recency engagement	strong historical sales and engagement, strong historical sales, strong historical engagement	A, B, C	A, B, C
Can be any one of these values or segments under every column of recency bucket and historical bucket and accordingly final segment will be decided							

List of products that falls under **category segment A** and **product segment A** are highly recommended as they are ***high revenue generating, high sell through rate, high average daily sales, consistent selling, high customer and digital presence and engagements both historically and also recently***.

If the list of products are high in numbers, run a pareto anlaysis and chart the top 20% products that contributes to the 80% revenue in this category and product segment A. We can further run a market basket score and halo effect score on the products to better club and generate tailor made recommendations for promotions.

Causal inference model for selecting the best promotions for promoting the category-product

Most general approach - ***Fit a linear model using the following features***, get the coeffs., Y-variable is revenue. For every category, **filter category wise data and built separate models**; *choose the categories that fall under segment A*. Take ***historical 3 years of data at daily date level*** with all these features and value, mentioned below as examples.
Multiply the coefficients with the X feature values and get the incremental (from the base value).

Promotion-related Features:

- ***Promotion Type***: Identify the type of promotion **(e.g., discount, BOGO, free shipping, 20% discount, 50% discount etc. as columns)** [Values – 1, 0 flags as on that day was one of these promotions running or not; multiple promotions can also run, so accordingly multiple column values can have 1, 0]
- ***Promotion Duration:*** The length of time the promotion is active. [Values – since it a daily level data, insert 1, 0 accordingly, so if it runs for 3 days continuously, insert 3 1's for consecutive 3 days on the respective dates]
- ***Promotion Intensity:*** The extent of the discount or the attractiveness of the promotion. [Values – H {>=50%}, M {20-50%}, L {<20%}]

Marketing and Advertising:

- ***Advertising Spending:*** Amount spent on advertising the promotion(e.g. facebook spend, affiliates spend etc. as columns). [Values – 1000, 15000…. Numeric spend figures]
- ***Marketing Channel:*** The platform or channel used for marketing (online, offline, social media). [Values – facebook, hoardings, affiliates etc.]

Product-related Features:

- ***Category average price:*** The average price of products in the category [Values – 200, 300…numeric values; avg. price of inventory available that day]
- ***Product Availability:*** Stock levels and product availability during the promotion [Values – 100, 150…numeric values; quantity of inventory available]

Temporal Features:

- ***Seasonality:*** Consider if the promotion aligns with seasonal trends. [Values – high, medium, low, neutral alignment with season]

Customer Behavior:

- ***Historical Purchase Data:*** Previous purchases by customers in that category in the last 7, 14, 21 days from that day of promotion [Values – numerical values…revenue generated by customers from all transactions 7, 14, 21 days before, from that date ; basically creating exogenous features]
- ***Customer Loyalty:*** Loyalty program participation or repeat purchases. [Values – numerical values…no. of purchases under loyalty program]

Causal inference model for selecting the best promotions for promoting the category-product

External Factors:

- ***Economic Conditions***: Economic indicators that might affect purchasing power (e.g. GDP, unemployment rate, CPI index as columns) [Values – downloaded numerical data from govt. repo. If available]
- ***Competitor Activity:*** Actions taken by competitors during the promotion period. [Values – competitors spend numerical values from third parties]

Online Presence:

- ***Website Traffic:*** Number of visitors to the website during the promotion. [Values – numerical values as total footfalls on the website for that category]
- ***Social Media Engagement:*** Likes, shares, and comments on social media related to the promotion.

Customer Feedback:

- ***Customer Reviews:*** Sentiment analysis of customer reviews during the promotion. [Values – can run a basic NLP model and get the sentiments on everyday basis as per the reviews, comments, support tickets raised at daily level]
- ***Customer Ratings***: ratings collected through purchases. [Values – numerical value as avg. customer rating for that category on that day]

Regression model : log (Y) = m1*log(x1) + m2*log(x2) + m3*log(x3)…… use log to bring all column values under a single range and then fit the model

Multiple the m(coeff.) values of the promotion features, with the promotions x1, x2… values and sum it up i.e. m1*(sum of all values of x1). Check the (incremental/decrement) contribution of that promotion i.e. {[sum of actual x1 promotion values - m1*(sum of all values of x1)]/ sum of actual x1 promotion values}. Similarly do this for all the promotions and then RANK the promotions.

Repeat this exercise on only the Diwali month filtered data for past 5 years. Do PCA since size of data for model fitting is less.
Check the (incremental/decrement) contribution of the promotions and the RANK the promotions.

Promotion features/promotion types that ranks as the top in terms of contribution scores for every category, use that type of promotions with the list of products from product segment A that falls under the category segment A model.

Promotion & campaign planning (data science & analytical) model pipeline high level overview