ADVISAI CONSULTING
from Data to Action

Apparel brand
Sales Analysis Assignment

Advisai Consulting

By

-Kinsuk Ghatak

-9038426200

-kinsukghatak@gmail.com

-https://www.linkedin.com/in/kinsukghatak/

Sec 1 : Exploratory analysis of the data

-Sales Trend & growth analysis

-Category analysis

-SKU level deep dives

-Product basket analysis

-Repeat Purchase behaviour analysis

Methodology & steps followed for data preparation :

Raw data set

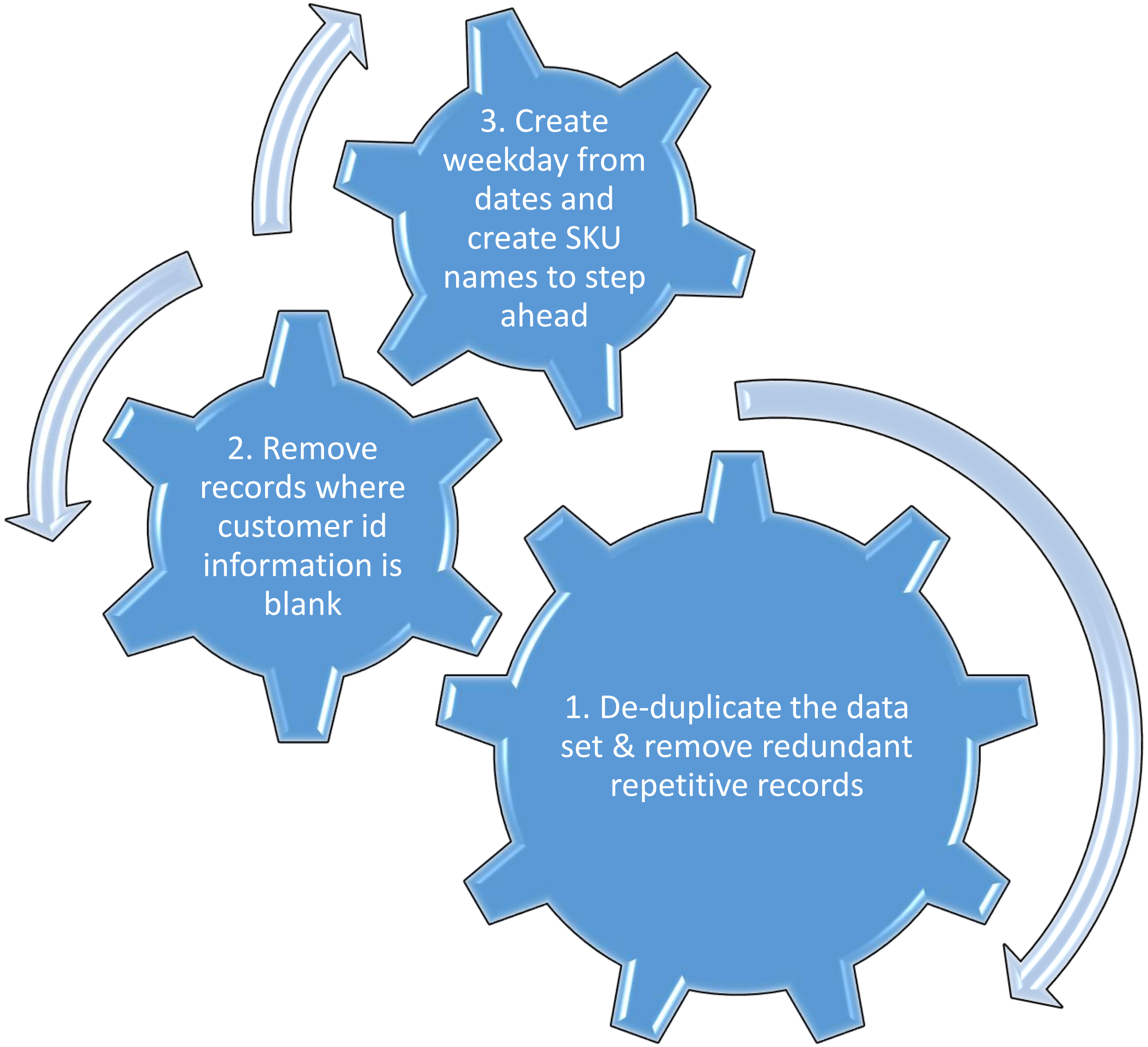
- # of rows : 8,11,773

Remove duplicate rows

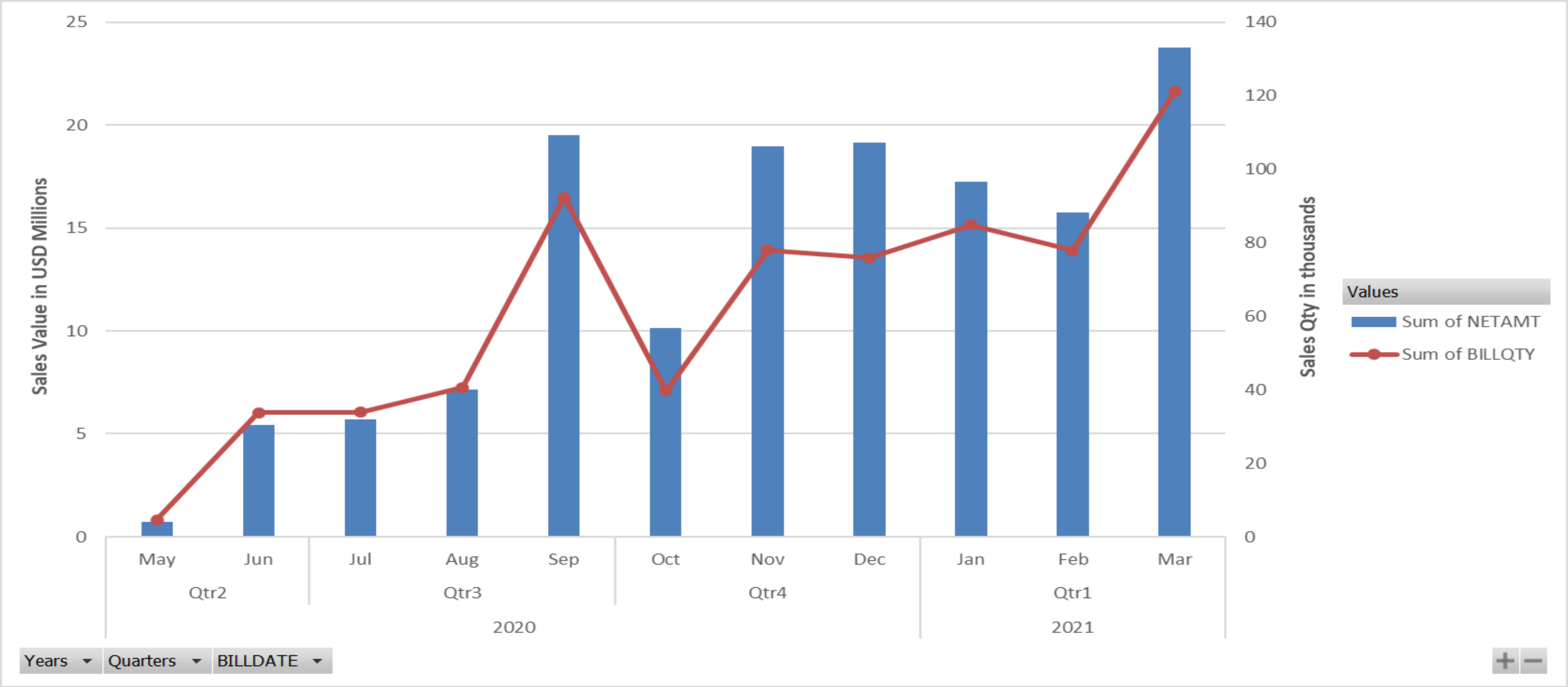
- # of rows : 8,10,951

Remove records with missing custid

- # of rows : 6,47,178



The apparel brand has seen a 74% CAGR Growth on Q-o-Q

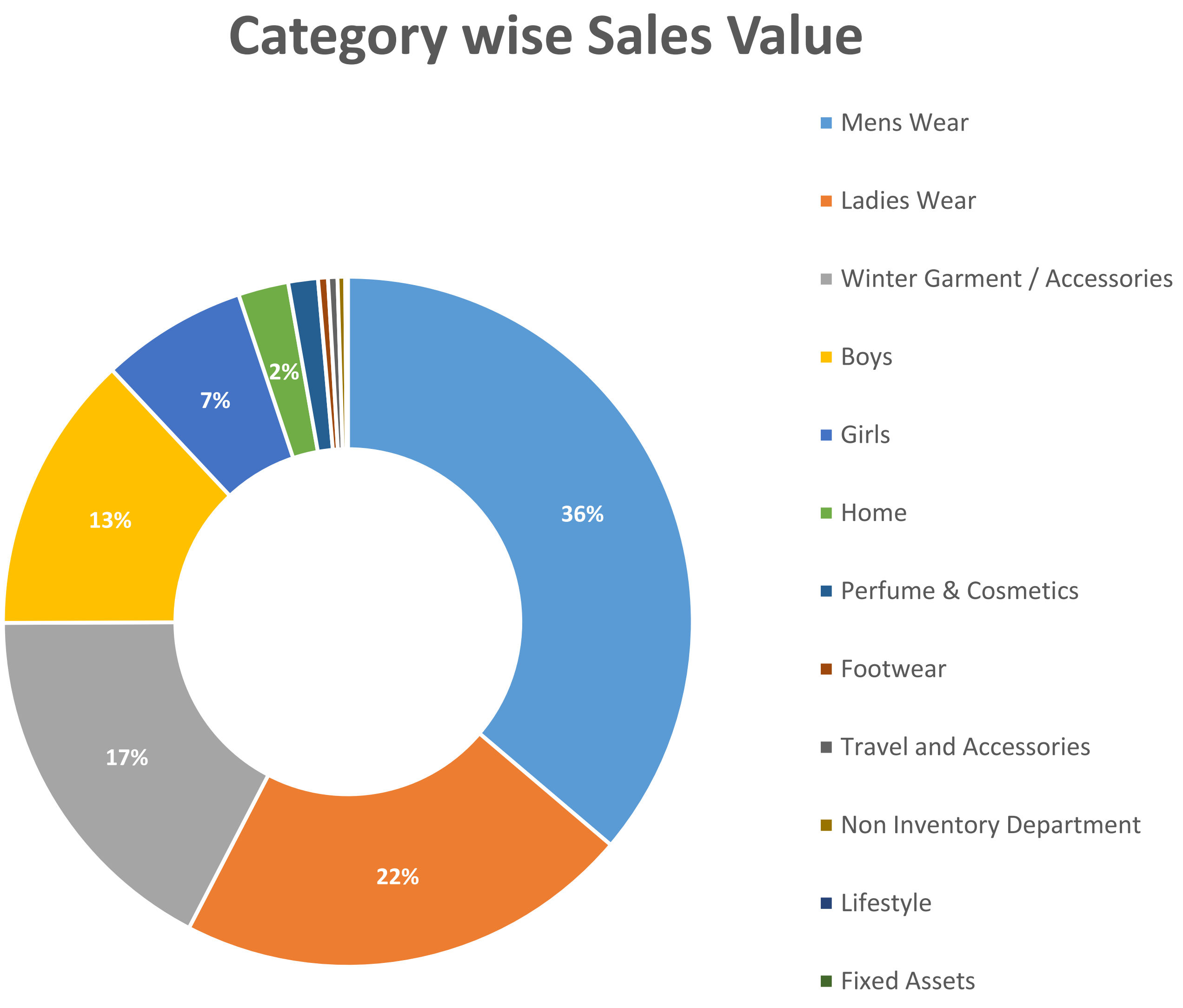


Quarters	Sales Value	Sales Qty	Value Growth	volume Growth
2020-Q2	61,29,745	38,456		
2020-Q3	3,23,49,476	1,66,822	427.7%	334%
2020-Q4	4,82,66,018	1,93,596	49.2%	16%
2021-Q1	5,67,69,066	2,83,747	17.6%	47%
CAGR	74%	65%		

Category wise Sales Value analysis

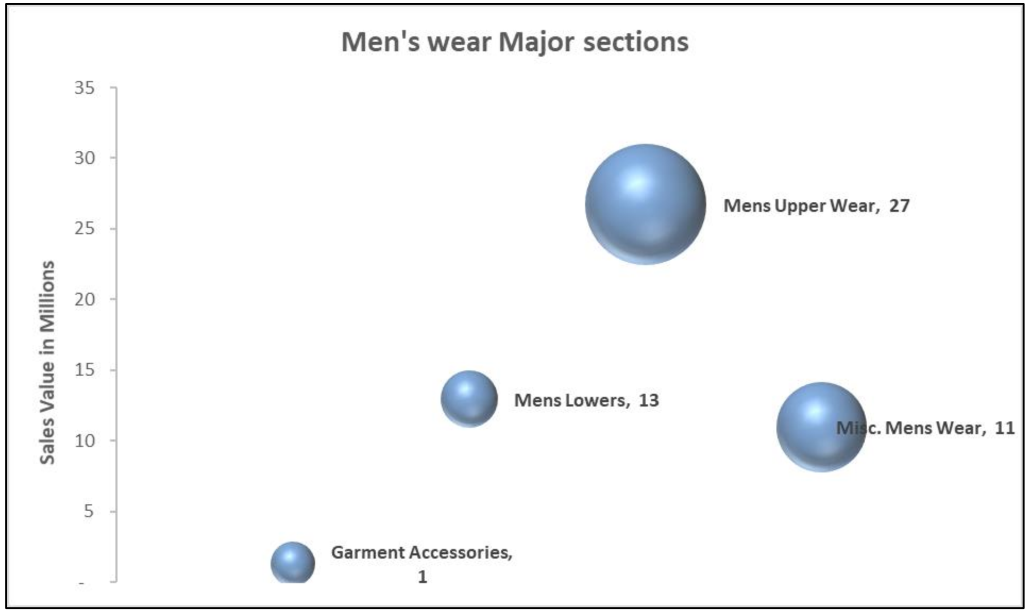
Category	Sales Value
Mens Wear	5,19,50,771
Ladies Wear	3,07,41,896
Winter Garment / Accessories	2,48,66,427
Boys	1,87,57,382
Girls	98,35,873
Home	33,78,633
Perfume & Cosmetics	19,94,057
Footwear	6,64,853
Travel and Accessories	6,35,110
Non Inventory Department	5,13,559
Lifestyle	1,75,676
Fixed Assets	70

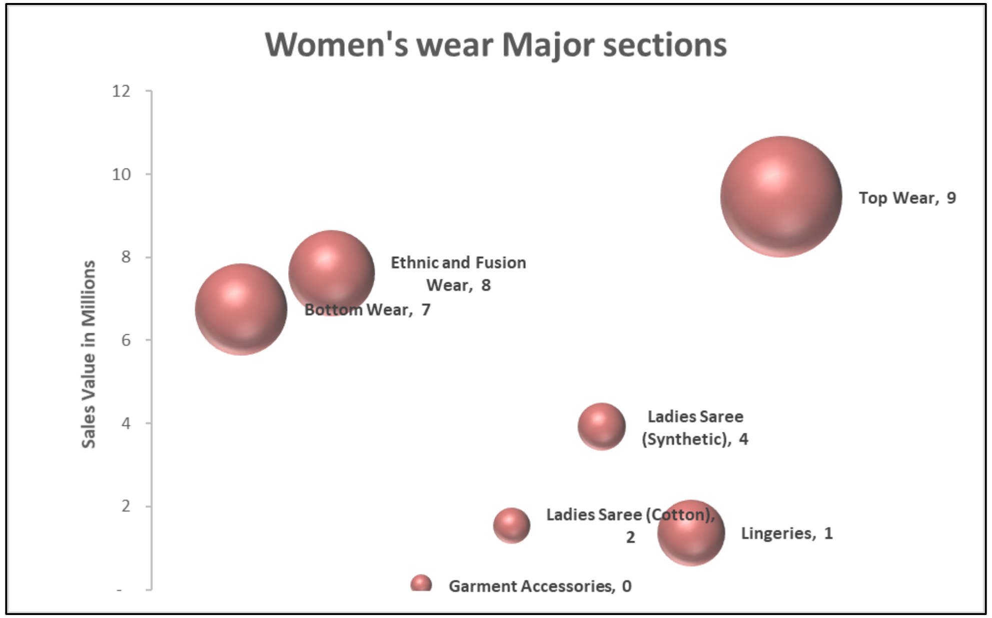
- Almost 60% of the sales come from Men's and women's wear (Ladies wear) categories



Major sections under men's and women's wear categories

ADVISAI CONSULTING
from Data to Action





- ❖ Top wear or upper wear are the major sub segments in both men's and women's categories.
- ❖ The size of the bubbles represent sales volume

Mens' Wear Sections	Sales Value	Sales Volume
Garment Accessories	12,98,765	12,781
Mens Lowers	1,29,61,434	20,773
Mens Upper Wear	2,66,94,140	97,886
Misc. Mens Wear	1,09,96,432	54,142

Ladies' Wear Sections	Sales Value	Sales Volume
Bottom Wear	67,49,663	21,517
Ethnic and Fusion Wear	76,14,354	18,577
Garment Accessories	98,586	1,098
Ladies Saree (Cotton)	15,42,500	3,328
Ladies Saree (Synthetic)	39,16,850	5,774
Lingeries	13,63,196	11,391
Top Wear	94,56,747	37,715

- ❖ The top 3 SKUs on a month on month basis in terms of Sales Value have been extracted and analysed. The details of the same can be found in the attached spreadsheet.

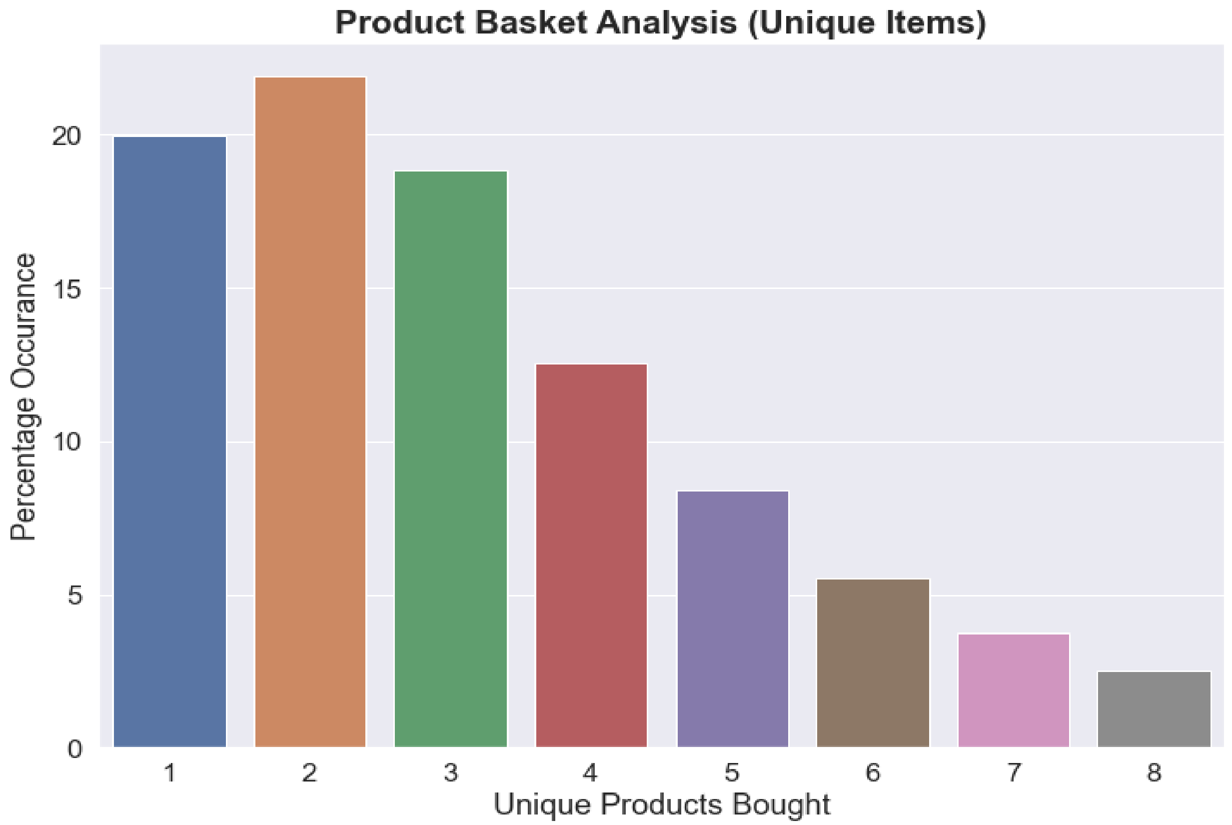
Microsoft Excel Worksheet

Weekday wise sales analysis : Impulse buying is inevitable

- Since the brand under consideration is mostly apparel segment an impulse buying behaviour is naturally expected.

- As usual , the buying behaviour can be easily replicated in terms of more Dollar sales value conversion happening towards the ends of the week in last 4 quarters

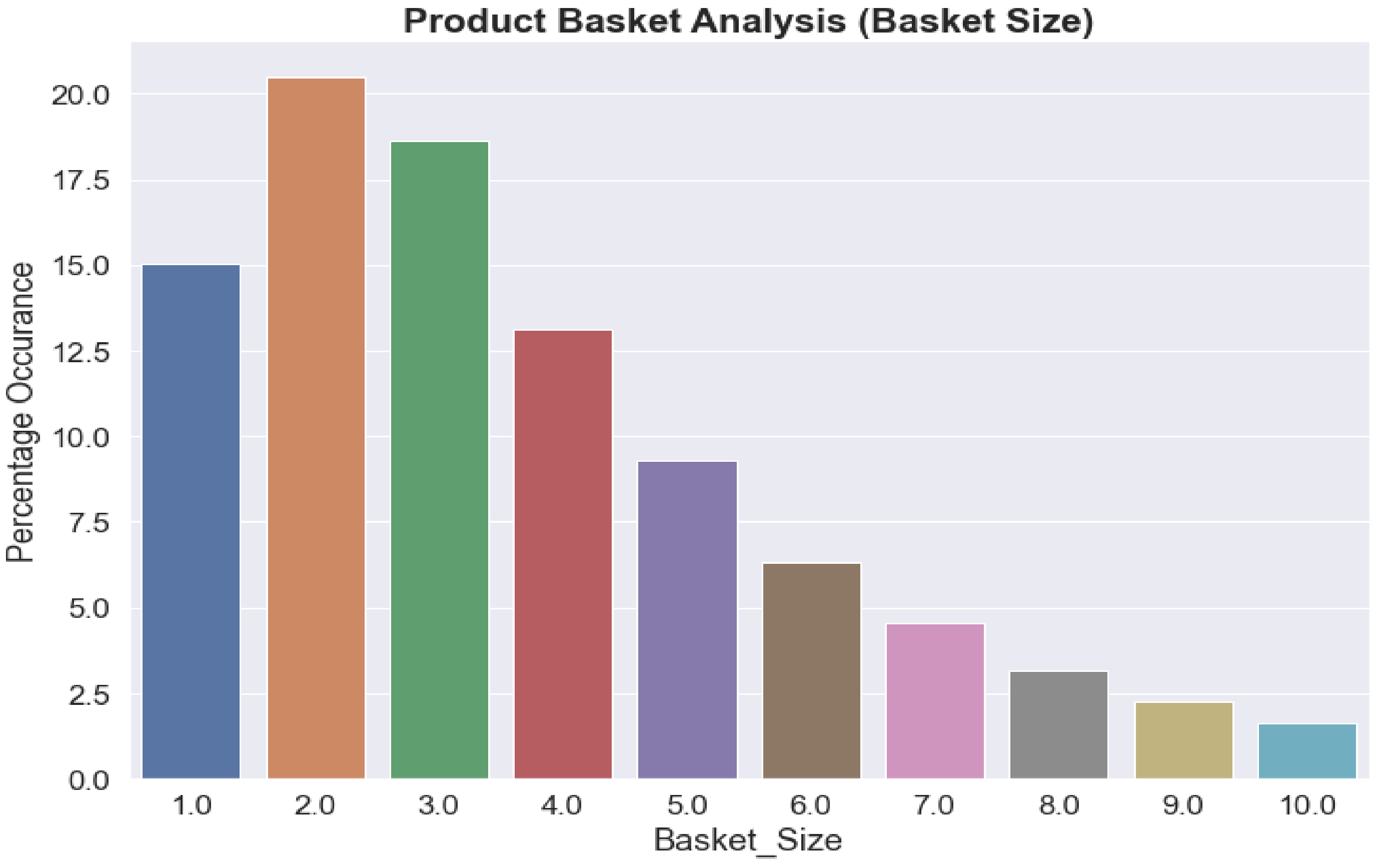
Basket analysis in terms of unique products bought



Unique Products Bought	Total_occurance	Percentage Occurance	Cumulative_Occurance
2	37984	21.9	21.9
1	34578	19.94	41.84
3	32672	18.84	60.68
4	21786	12.56	73.24
5	14591	8.41	81.65
6	9616	5.55	87.2

- ❑ Customers have almost equal distribution of instances where they are buying either 1 / 2/ 3 unique products.
- ❑ The analysis has been carried on the 'BARCODE' information provided in the data
- ❑ In almost 90% of the cases the maximum unique products customers have bought are 6 ; with heavy weight towards 2 unique products(in almost22% of the cases)

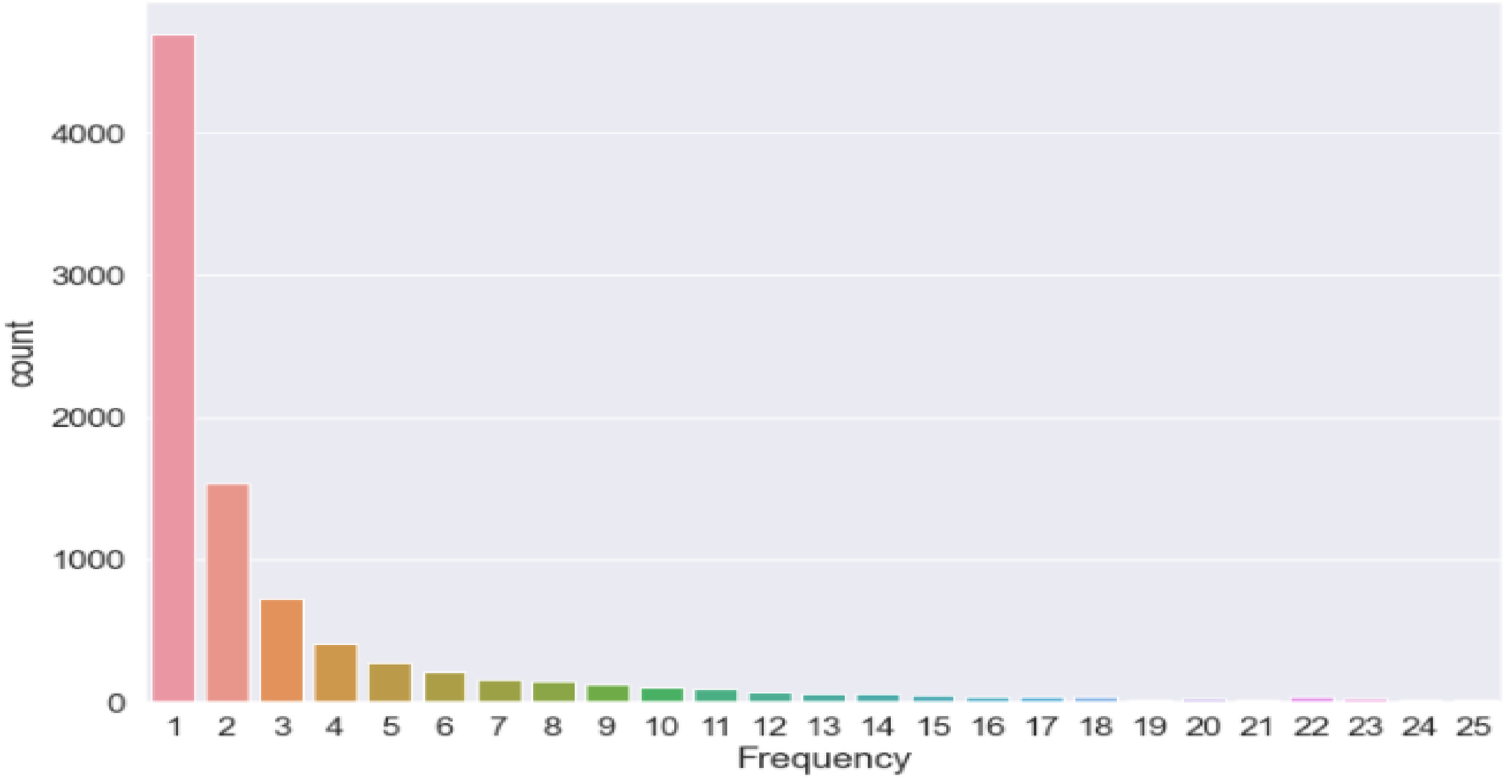
Basket size analysis



Basket_Size	Total_occurance	Percentage Occurance	Cumulative_Occurance
2	33471	20.5	20.5
3	30429	18.63	39.13
1	24577	15.05	54.18
4	21433	13.13	67.31
5	15165	9.29	76.6
6	10313	6.32	82.92
7	7422	4.55	87.47
8	5134	3.14	90.61

- In almost 40% of the cases the average basket size is either 2 or 3
- From the distribution it looks like customers prefer buying 2 products at one go in almost 20% of the instances which is significant
- Further analysis like Market basket comparison using a-priory can be explored in order to understand the buying behaviour of the customers and to analyse and understand which are the products that are getting bought together

Customer repeat purchase analysis :



Frequency of repeat / single purchase	Count_of_instances	Percentage Occurance	Cumulative_Occurance
1	4691	48.1	48.1
2	1534	15.73	63.83
3	724	7.42	71.25
4	411	4.21	75.46
5	281	2.88	78.34
6	212	2.17	80.51
7	154	1.58	82.09
8	143	1.47	83.56
9	123	1.26	84.82
10	102	1.05	85.87

- For a significant number of times customers have come back and made a repeat purchase over different billing dates

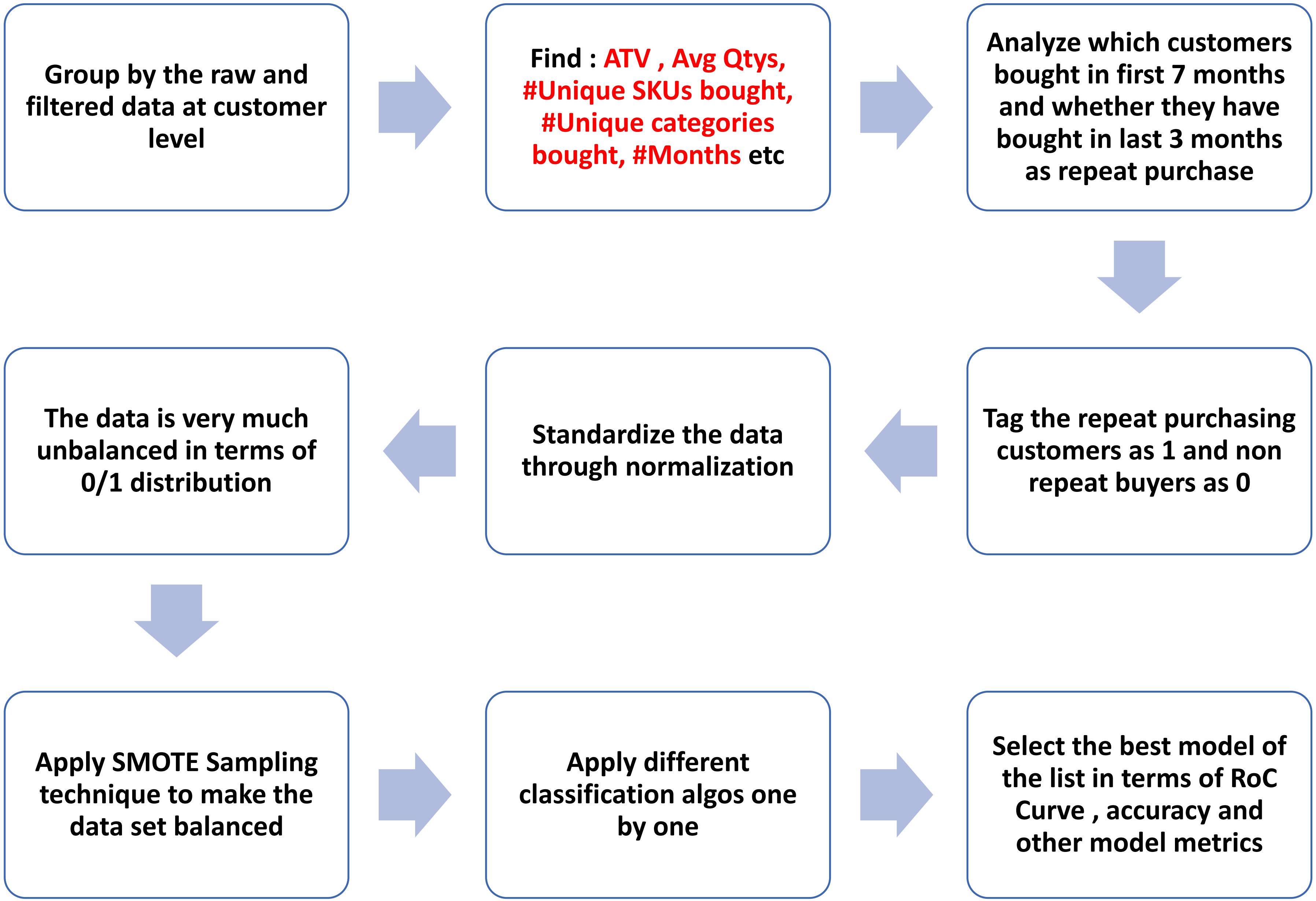
- For almost 52% of the cases customers have made a repeat purchase through comeback

- This hints that the brand nurtures a good loyalty maybe either through promo campaigns , customer service , activation methods and so on

Sec 2 : Classification model

-Analysis on probability of repeat buying

Classification model steps (Data prep and model building) :



Final Classification model , performance and metrics :

ADVISAI CONSULTING
from Data to Action

Confusion Matrix :

		TRUE	
		0	**1**
Predicted	**0**	413	228
	1	344	966

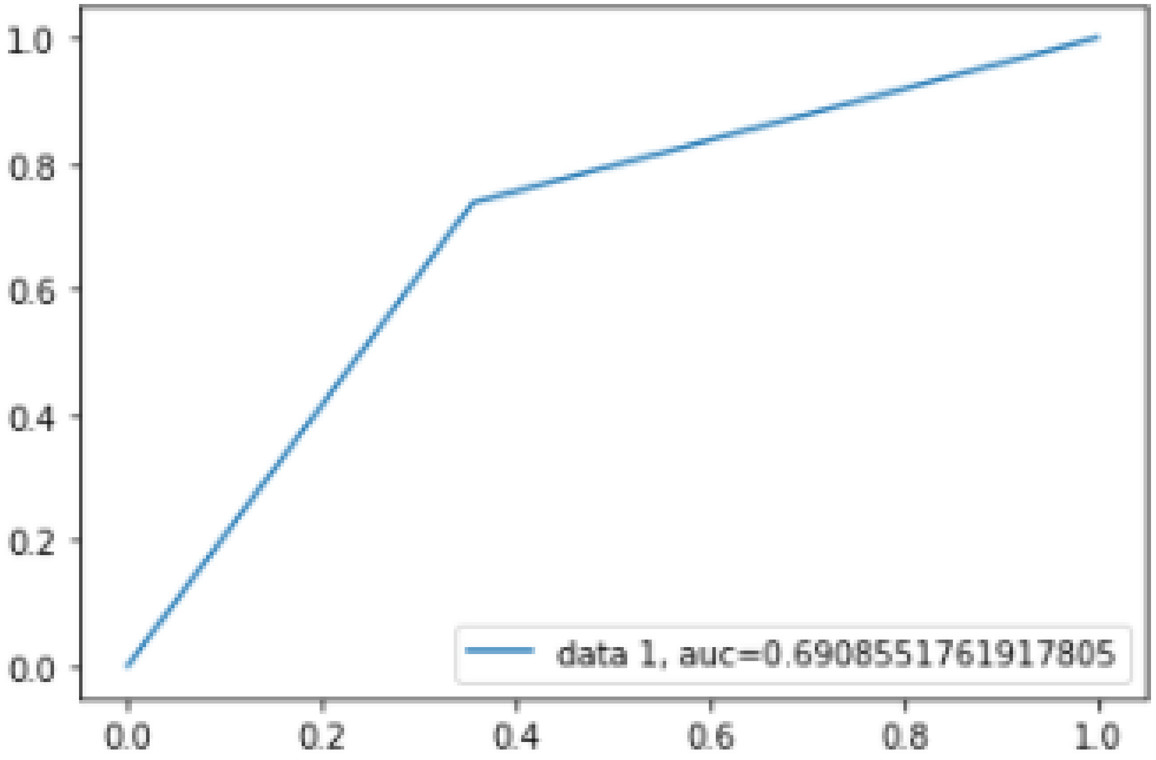
Model performance metrics:

	Training	Testing
Accuracy Score	81.30%	70.10%
Recall Score	80.70%	73.70%
Gini Score	63.30%	38.20%

- ❑ sensitivity is : 0.73
- ❑ specificity is : 0.64
- ❑ accuracy is : 0.71
- ❑ balanced accuracy is : 0.69

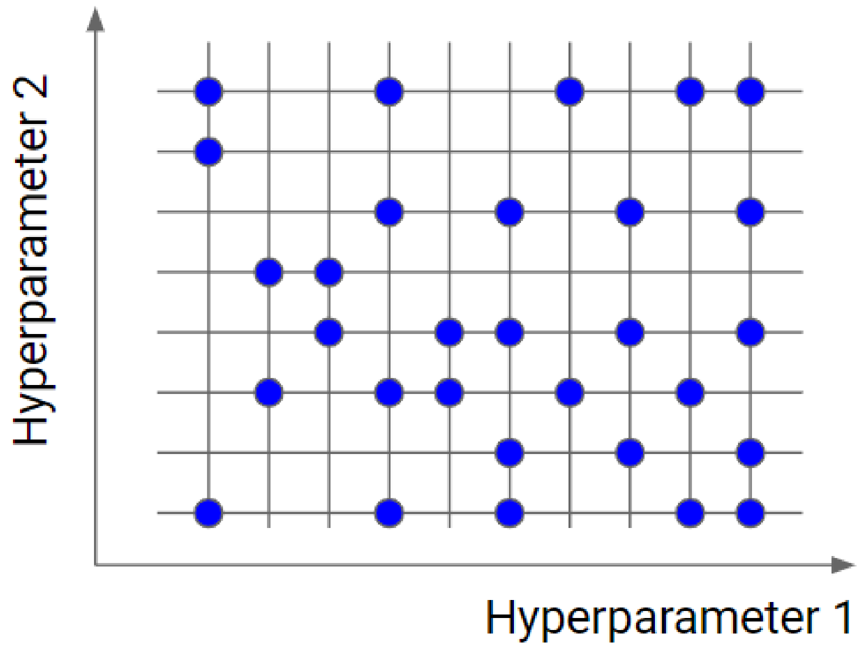
Final model :
XGBoost classifier with SMOTE oversampling under ML pipeline method

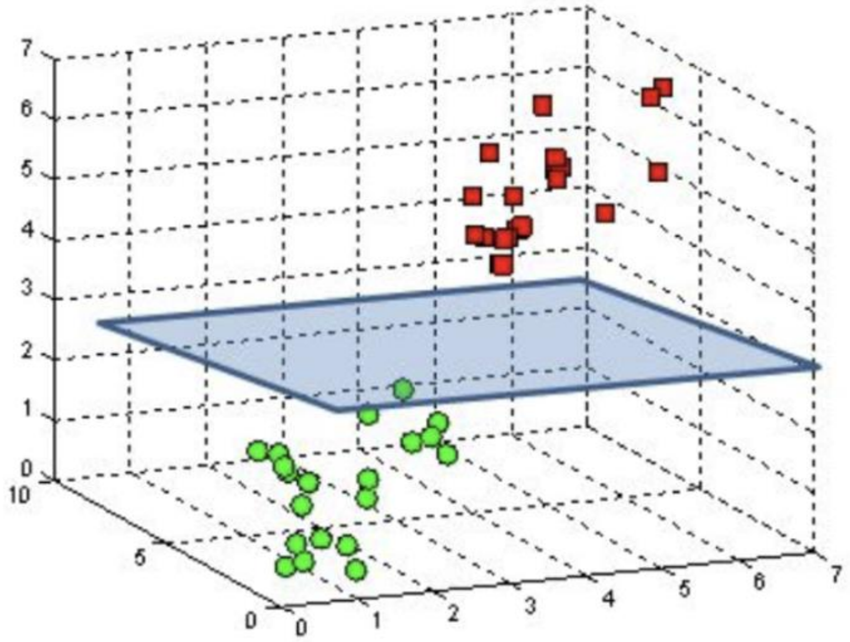
RoC Curve :



Classification model further scopes of improvements :

- ❑ Application of GRID search technique in Random Forest for more better hyper parameter tuning and selecting the best model out of that
- ❑ Further exploration and tuning of modelling parameters under XG Boost to improve the performance
- ❑ Further exploration with Support Vector machine algorithm





Sec 3 : Customer segmentation

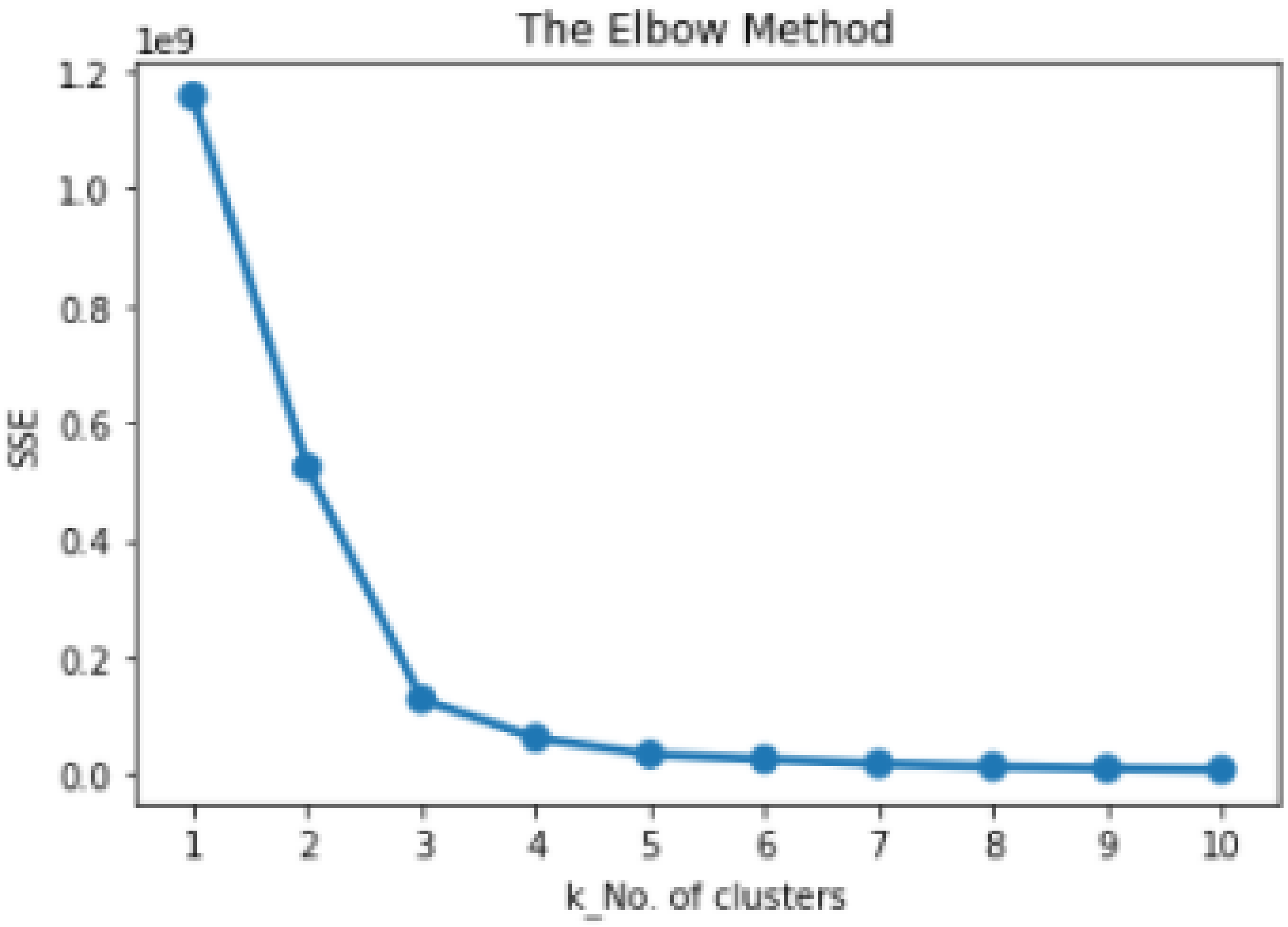
-Clustering customers into different buckets based on their buying behaviour

Customer Segmentation using k-means : Steps followed :

ADVISAI CONSULTING
from Data to Action

1. Filter customers based on frequency of transactions (greater than 1)
2. Take customer base who have transacted more than once
3. Calculate their total monetary engagement , frequency of purchases (in terms of unique dates) and How recent they have purchases (Recency) . Apart from that also calculate number of unique SKUs bought , no of Categories bought & no. of unique months they bought in
4. Normalize the data and identify optimal number of clusters using k-means scree plot
5. Apply the k-means model on the normalized data set
6. Merge the cluster or segment information to the original data set
7. Group by the segments and find summary of few metrics
8. Create profile of different customer segments

k-means : Optimum no. of clusters:



- ❑ Though the marginal decrease of SSE goes down significantly after k=3 ; still we have taken k=4 as the final number of clusters here in this case.

Customer segmentation : profiles and different segments

ADVISAI CONSULTING
from Data to Action

kmeans_cluster	Count of customers	Total_Sales_Value	Total_Qty	#Bills	ATV/ Qty	ATV/Bill	Avg recency	Avg #Categories bought	Avg of #Unique_Months	Avg of #Unique_SKU bought	Profile/Segment Name
1	4,724	3,73,29,229	1,72,642	43,348	216	861	153.75	5.17	3.67	29.5	Champions
2	100	3,56,52,133	1,70,260	41,699	209	855	109.54	10.66	10.87	1120.72	Loyals
3	405	6,17,97,173	2,96,028	75,637	209	817	109.69	10.03	10.51	519.95	Potentials
4	1	43,71,535	23,910	8,210	183	532	109	11	11	9555	Outlier