

**INDIAN INSTITUTE OF TECHNOLOGY
(IIT) MADRAS**

**BACHELOR OF SCIENCE (BS)
IN
DATA SCIENCE & APPLICATIONS**

**BUSINESS DATA MANAGEMENT
(BDM) CAPSTONE PROJECT**

FINAL SUBMISSION

SUBMITTED BY

ANSH KUSHWAHA

Roll No.: 21f1006019

21f1006019@student.onlinedegree.iitm.ac.in

INDEX

<u>S. NO.</u>	<u>TOPIC</u>	<u>PAGE NO.</u>
1	Title	2
2	Executive Summary	2-3
3	Problem Objectives	3
4	Methodologies	3
5	Data Collection	4
6	Analysis Process / Method	4-7
7	Observations	7-8
8	Results & Findings	9-18
9	Interpretation of Result & Recommendations	19-20

TITLE

Solving Cosmetics Wholesale Business Problems Using Data Analytics

EXECUTIVE SUMMARY

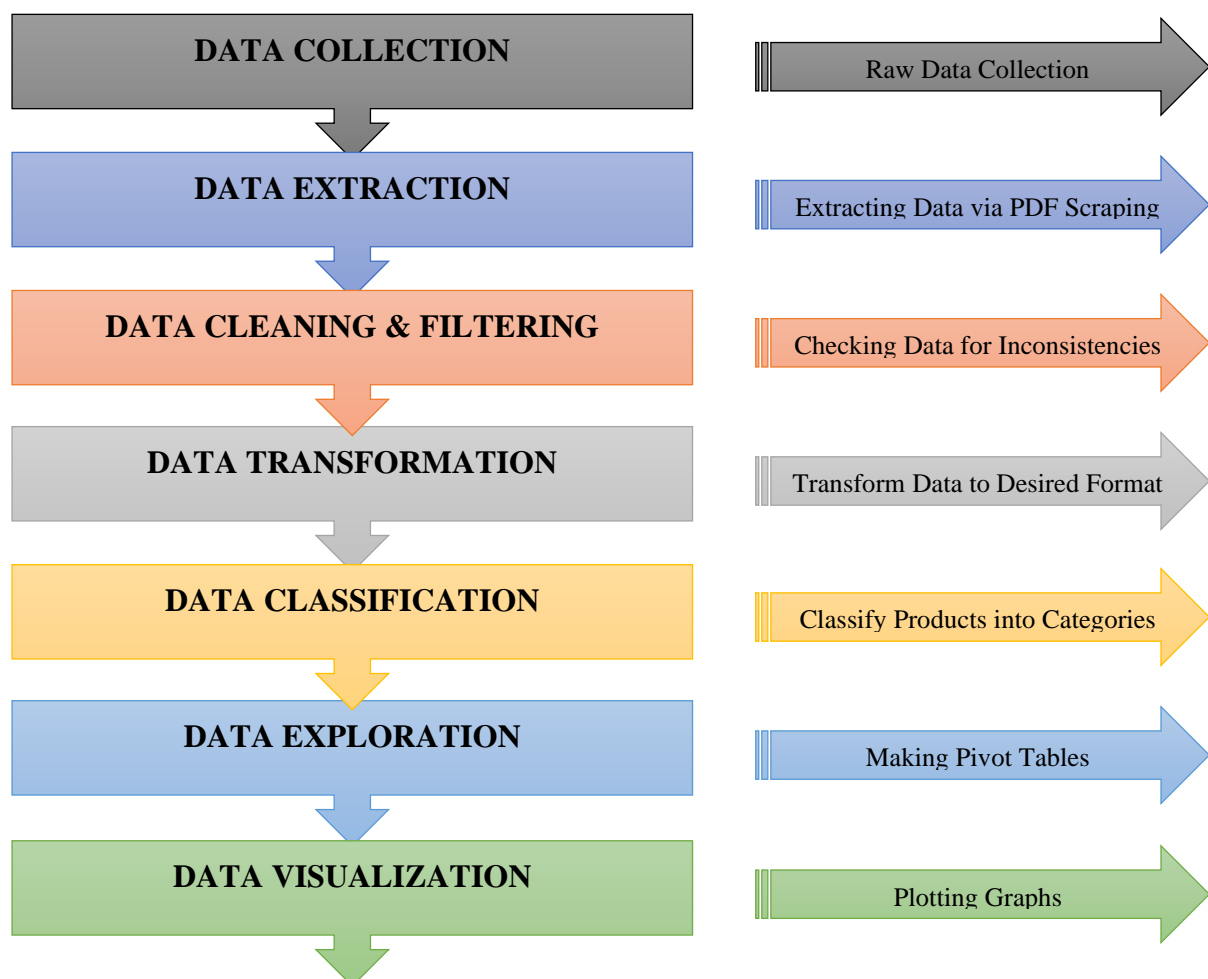
The company mentioned here falls under the B2B (business-to-business) category. It's a small wholesale cosmetics business with one owner and two employees that started operations six months ago. This type of business faces many challenges, including losses due to expired products in stock; finding retailers to work with; earning the trust of their traders, etc. When the business is in its infancy, it is also a struggle to learn about the market and how things work. These problems make it difficult for a new business to position itself in the market and grow. New businesses get stuck between what to do and what not to do.

Data analytics is currently the best way to guide these startups and help them grow. The data generated by these companies can help them respond to the changing needs and preferences of their customers. It will also help to learn about changes in market trends and trading trends. The solution created after analyzing the data can give comprehensive advice on what to buy and what not to buy and in what quantities. It can also give detailed recommendations to businesses on what to stop doing, what to start doing, and what to keep doing.

PROBLEM OBJECTIVES

- ❖ To find ways to get rid of the loss caused by expired products in stock.
- ❖ To find the steps to be taken, in that fixed capital, to increase profit and grow the business.

METHODOLOGIES



DATA COLLECTION

To find the above-stated problem objectives, I need to have complete data on the purchase and sales of all products, including their quantities, discount, tax, etc.

So, I **collected purchase and sales invoices from the previous months** to get all these details in one place. I got xerox copy of purchase invoice & PDF copy of sales invoice.

ANALYSIS PROCESS / METHOD

Analysis Tool Used: → Microsoft Excel

❖ **Extracting the Data via PDF Scrapping**

Since the Sales Invoices I collected are in PDF format, I extracted the data from the PDF using **PDF Scrapping**. I used the **Tabula Windows Application Software** for Scrapping purposes. The extracted data was then converted into an Excel Sheet “Sales” in the format shown below:

Retailer	Date	Product	HSN	Quantity	Unit	MRP	Rate	Discount	GST	Amount
----------	------	---------	-----	----------	------	-----	------	----------	-----	--------

❖ **Entering the Data into Excel Sheet**

Since the Purchase Invoices I collected were in Paper format, I manually entered the data into an Excel Sheet “Purchases” in the format shown below:

Date	Product	HSN	Quantity	Unit	MRP	Rate	Discount	GST	Amount
------	---------	-----	----------	------	-----	------	----------	-----	--------

❖ **Checking the Data for any Inconsistency**

The data is checked for errors and inconsistencies as follows:

- Extra whitespace in any cell will be removed using Microsoft Excel's split() function.
- Each cell is checked for any misspelt words and corrected.

- Each cell is formatted to the same type as the data it contains. Like the cells containing numerical values are formatted to be of type number & similar with date, currency, percentage, etc.

❖ Transforming the Data into Desired Format

These purchase and sales sheets are combined and filtered to form three Excel sheets. Unwanted features like HSN, Unit, etc. are removed.

The desired format of the sheets is shown below:

Sheet – 1: Calculations

Product	HSN	Purchase Discount	Purchase Rate	Purchase Rate (After Discount)	Sales Discount	Sales Rate (After Discount)
CGST	SGST	Total GST	Net Purchase Rate (Including Tax)	Net Sales Rate (Including Tax)	Net Profit	Maximum Profit to Retailer

Sheet – 2: Purchase

PRODUCT	DATE	QUANTITY	MRP	RATE	AMOUNT
---------	------	----------	-----	------	--------

Sheet – 3: Sales

PRODUCT	RETAILER	DATE	QUANTITY	MRP	RATE	AMOUNT
---------	----------	------	----------	-----	------	--------

❖ Creating Ledger of Stock

To solve the above stated problems, we have to create an inventory ledger. The stock ledger is a tabular collection of the daily opening and closing amounts of each product. This helps identify fast-moving and slow-moving products, and then planning which products to buy and how much.

The filtered purchase and sales sheets are then combined to form an inventory register.
The desired format is shown below:

Sheet – 4: Inventory

PRODUCT	TREND (SPARKLINES)	...	[DATE OF PURCHASE / SALES]	...	RATE	STOCK
---------	-----------------------	-----	---------------------------------	-----	------	-------

❖ **Classifying Products into Different Categories**

Since there are so many products, I have classified all the products into different categories based on their prices and inventory trends. This will help me understand the data better.

Classification based on Price (M.R.P.) of the Product:

<u>PRICE (M.R.P.)</u>	<u>CATEGORY</u>
Below 100	0-100
Between 101 & 200	101-200
Between 201 & 300	201-300
Between 301 & 400	301-400
Between 401 & 500	401-500
Between 501 & 600	501-600
Between 601 & 700	601-700
Between 701 & 800	701-800
Between 801 & 900	801-900
Between 901 & 1000	901-1000
Above 1000	1000+

Classification based on Inventory Trends of the Products:

<u>CATEGORY</u>	<u>CRITERIA</u>
1	Products which are never sold during the observation period
2	Products which are first sold are but not sold after its second purchase
3	Products which have several sales - purchase & still available in stock
4	Products which get stock out & purchased again
5	Products which get stock out but not purchased again

❖ Making Pivot Tables

A pivot table is a table of values grouped to group the individual elements of a larger table into one or more discrete categories. It helps us to focus on a small portion of the data and analyze what things are lagging behind.

❖ Plotting Graphs

Graphs and charts are used to visually depict data and make it easy for us to find trends and trend breaks in data points. It clearly shows what factors are causing these problems & helps us to analyze the data efficiently.

OBSERVATIONS

- ❖ The products were purchased three times from the company during the observation period.
- ❖ These products are sold to three different retailers. Some retailers buy a couple of times a month, while others buy all at once.
- ❖ The products are purchased at a much lower price than MRP and has an additional 12 % discount. According to government regulations, 18 % GST (CGST and SGST) applies to purchases.
- ❖ The products are sold at a 32.2 % discount off the MRP, plus 18 % GST.

Calculation of Final Bill Amount

$$\begin{aligned} X &= \text{Rate} * \text{Quantity} \\ Y &= X (1 - (\text{Discount} / 100)) \\ \text{Amount} &= Y (1 + (\text{Tax} / 100)) \end{aligned}$$

In case of Purchase,

Rate	-	A Price lower than MRP
Discount	-	12
Tax	-	18

In case of Sales,

Rate	-	MRP
------	---	-----

Discount	-	32.2
Tax	-	18

❖ Total number of Products under observation are: 132

Number of Products in each categories:

➤ Based on Price (M.R.P.) of the Products

<u>CATEGORY</u>	<u>NUMBER OF PRODUCTS</u>
0-100	1
101-200	20
201-300	22
301-400	31
401-500	29
501-600	7
601-700	5
701-800	14
801-900	1
901-1000	0
1000+	2

➤ Based on Inventory Trends of the Products

<u>CATEGORY</u>	<u>CRITERIA</u>
1	Products which are never sold during the observation period
2	Products which are first sold are but not sold after its second purchase
3	Products which have several sales - purchase & still available in stock
4	Products which get stock out & purchased again
5	Products which get stock out but not purchased again

<u>CATEGORY</u>	<u>NUMBER OF PRODUCTS</u>
1	63
2	10
3	54
4	4
5	1

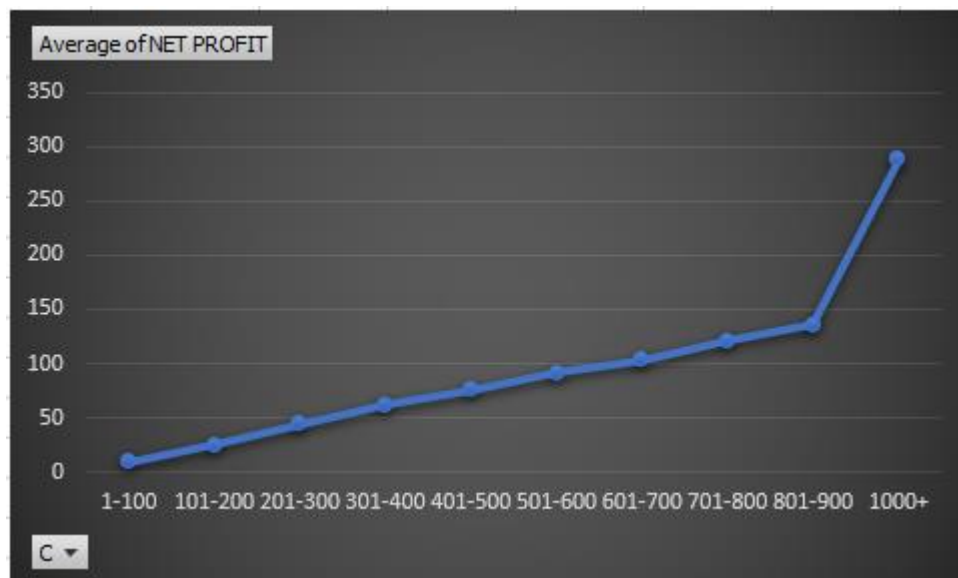
RESULTS & FINDINGS

❖ MRP Ranges v/s Net Profit per Product

➤ Pivot Table

Row Labels	Average of NET PROFIT
1-100	8.80044
101-200	24.5397048
201-300	44.55315273
301-400	62.08118555
401-500	75.42346786
501-600	90.90868343
601-700	103.1808048
701-800	120.5378429
801-900	135.974704
1000+	287.868316

➤ Pivot Chart



➤ Observation

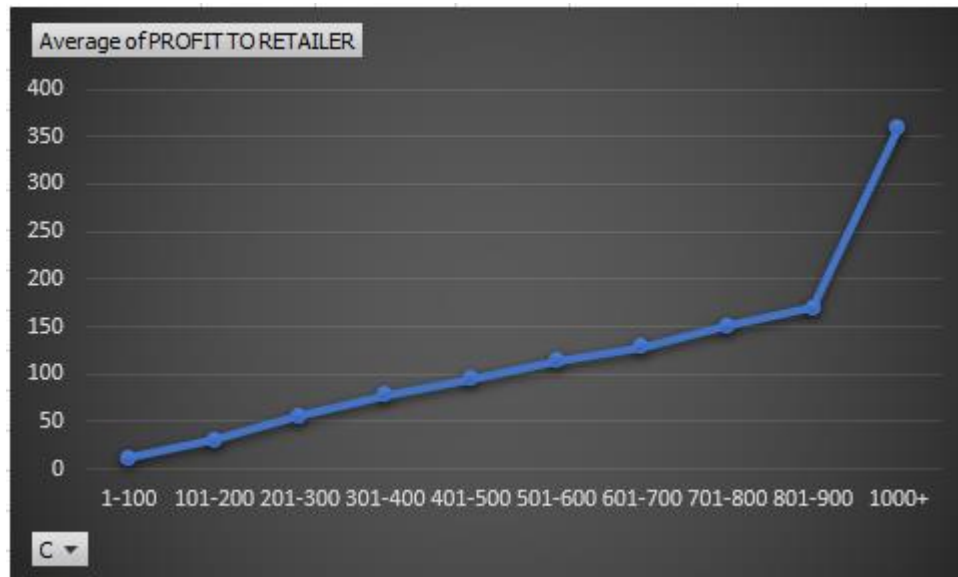
The average net profit of the products in various price (M.R.P.) ranges shows a sudden jump when the product's price exceeds the amount of 1000 Rs.

❖ MRP Ranges v/s Profit to Retailer per Product

➤ Pivot Table

Row Labels	Average of PROFIT TO RETAILER
1-100	10.9978
101-200	30.673864
201-300	55.64341455
301-400	77.59738065
401-500	94.27769241
501-600	113.6344114
601-700	128.9742
701-800	150.66986
801-900	169.966
1000+	359.82802

➤ Pivot Chart



➤ Observation

The average profit to retailer per product in various price (M.R.P.) ranges shows a sudden jump when the product's price exceeds the amount of 1000 Rs.

❖ MRP Ranges v/s Average Bill Amount of Products Purchased

➤ Pivot Table

Row Labels	Average of AMOUNT
1-100	1267.26336
101-200	2096.867822
201-300	2523.256395
301-400	4693.510443
401-500	3341.375635
501-600	3264.094099
601-700	5492.598628
701-800	3851.486682
801-900	6528.711552
901-1000	0
1000+	6813.523904

➤ Pivot Chart



➤ Observation

The Products are in the price ranges 301-400, 601-700, 801-900 & 1000+ are purchase more in comparison to others.

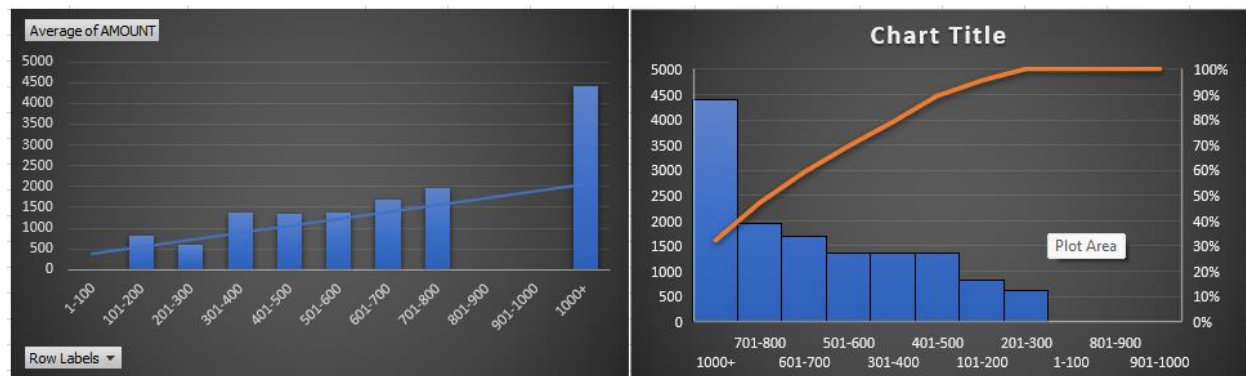
No products are purchased in the price range 901-1000.

❖ MRP Ranges v/s Average Bill Amount of Products Sold

➤ Pivot Table

Row Labels	Average of AMOUNT
1-100	0
101-200	826.941345
201-300	606.0303
301-400	1358.236797
401-500	1346.154261
501-600	1360.182291
601-700	1690.0845
701-800	1950.0975
801-900	0
901-1000	0
1000+	4397.28652

➤ Pivot Chart



➤ Observation

The Products in the price ranges 101-200, 301-400 & 1000+ are more in comparison to others.

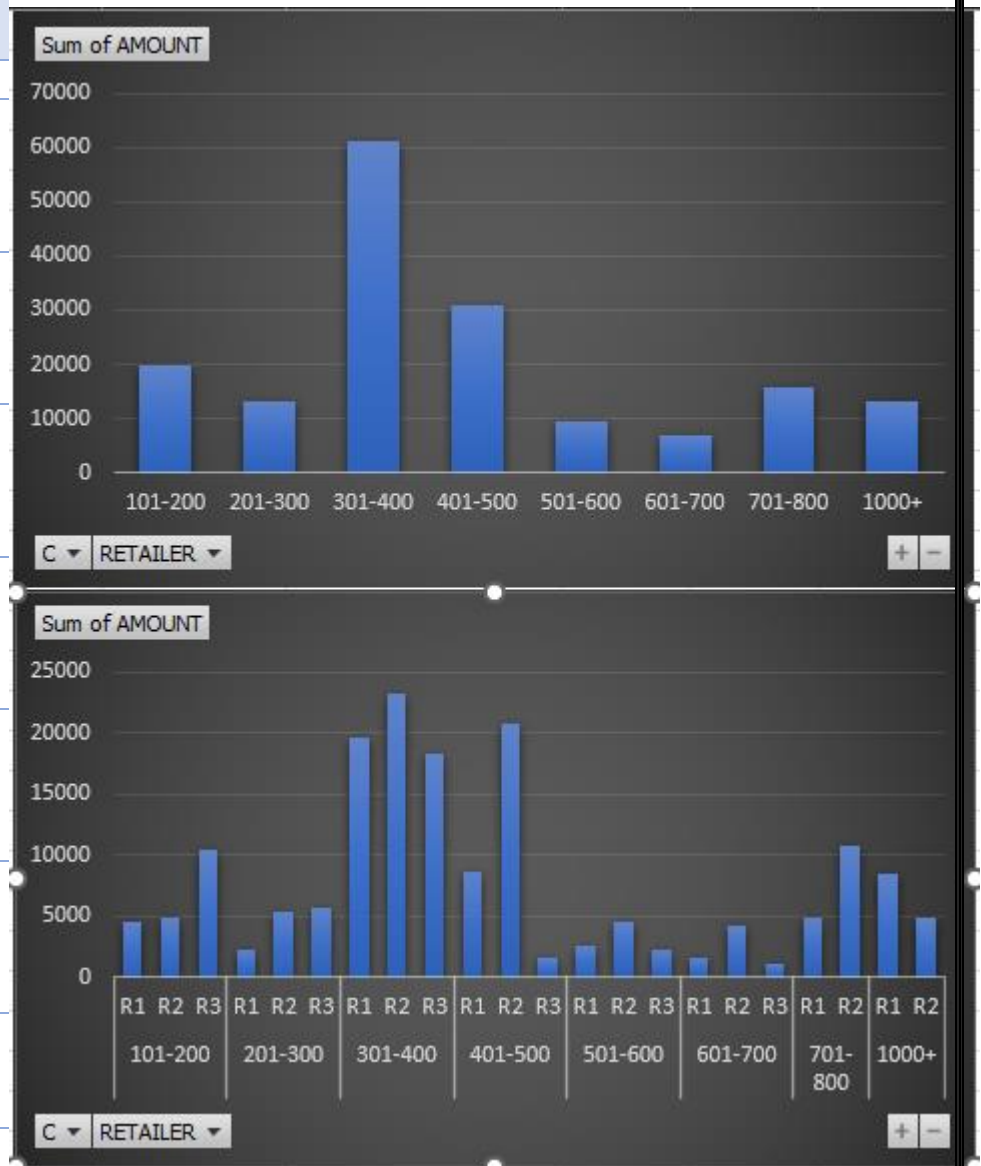
No products are sold in the price ranges 1-100, 801-900 & 901-1000.

❖ Price Ranges v/s Total Revenue per Retailer

➤ Pivot Table

Row Labels	Sum of AMOUNT
101-200	19846.59228
R1	4620.231
R2	4797.83988
R3	10428.5214
201-300	13332.6666
R1	2274.51372
R2	5313.86568
R3	5744.2872
301-400	61120.65588
R1	19560.17796
R2	23277.1638
R3	18283.31412
401-500	30961.548
R1	8700.435
R2	20741.037
R3	1520.076
501-600	9521.27604
R1	2637.73188
R2	4621.83108
R3	2261.71308
601-700	6760.338
R1	1560.078
R2	4160.208
R3	1040.052
701-800	15600.78
R1	4800.24
R2	10800.54
1000+	13191.85956
R1	8394.81972
R2	4797.03984
Grand Total	170335.7164

➤ Pivot Chart



➤ Observation

Retailer R3 is not buying costly products (MRP > 700). In fact, R3 is only interested in products with MRP < 400.

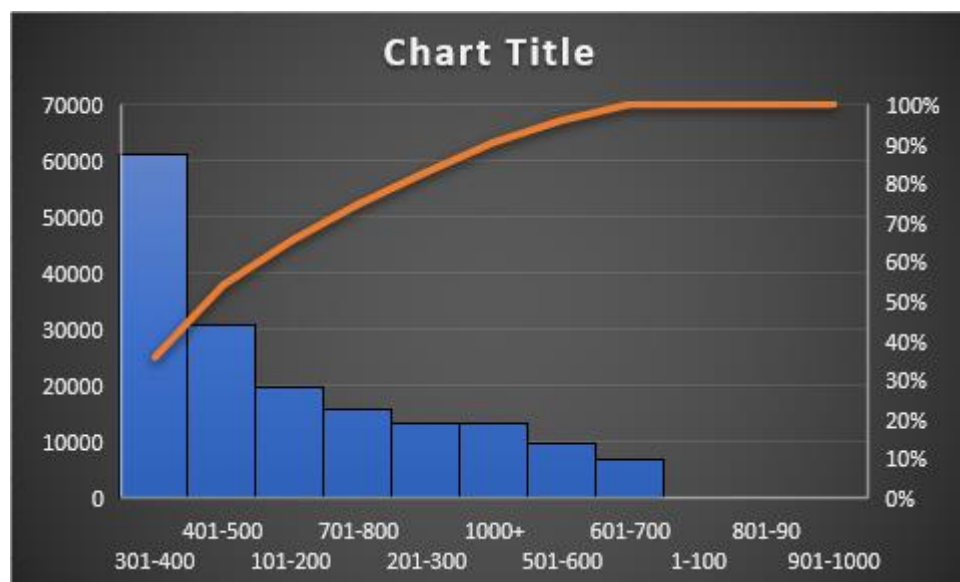
Retailers R1 & R2 are also comparatively less interested in products of range 501 – 700

❖ Price Ranges v/s Total Revenue (Pareto Analysis)

➤ Pivot Table

Row Labels	Sum of AMOUNT
1-100	0
101-200	19846.59228
201-300	13332.6666
301-400	61120.65588
401-500	30961.548
501-600	9521.27604
601-700	6760.338
701-800	15600.78
801-90	0
901-1000	0
1000+	13191.85956
Grand Total	170335.7164

➤ Pivot Table



➤ Observation

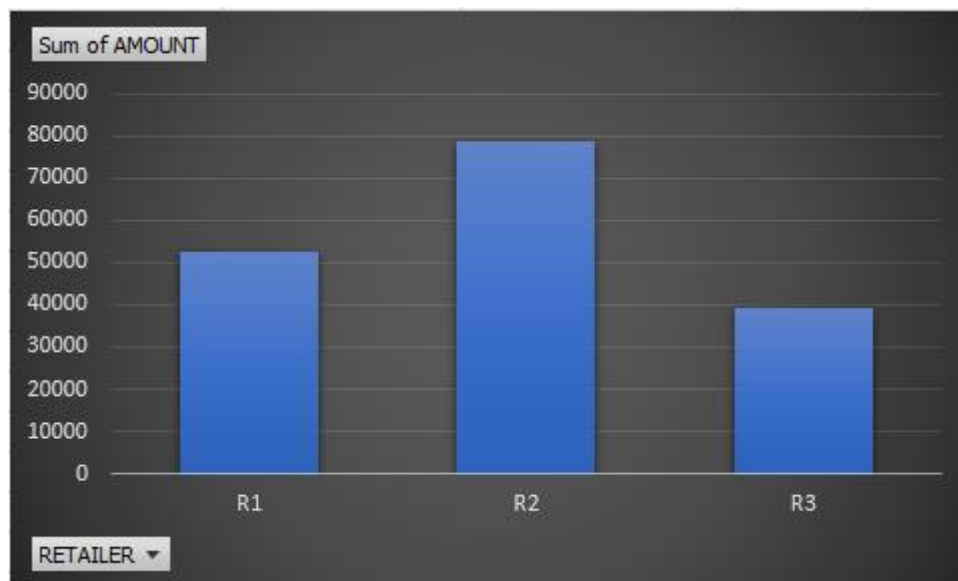
It does not follow Pareto Principle (80 : 20) but still approx. 54 % of the revenue is generated by the products in the price range 301-400 & 401-500.

❖ Retailers v/s Total Revenue

➤ Pivot Table

Row Labels	Sum of AMOUNT
R1	52548.22728
R2	78509.52528
R3	39277.9638
Grand Total	170335.7164

➤ Pivot Chart



➤ Observation

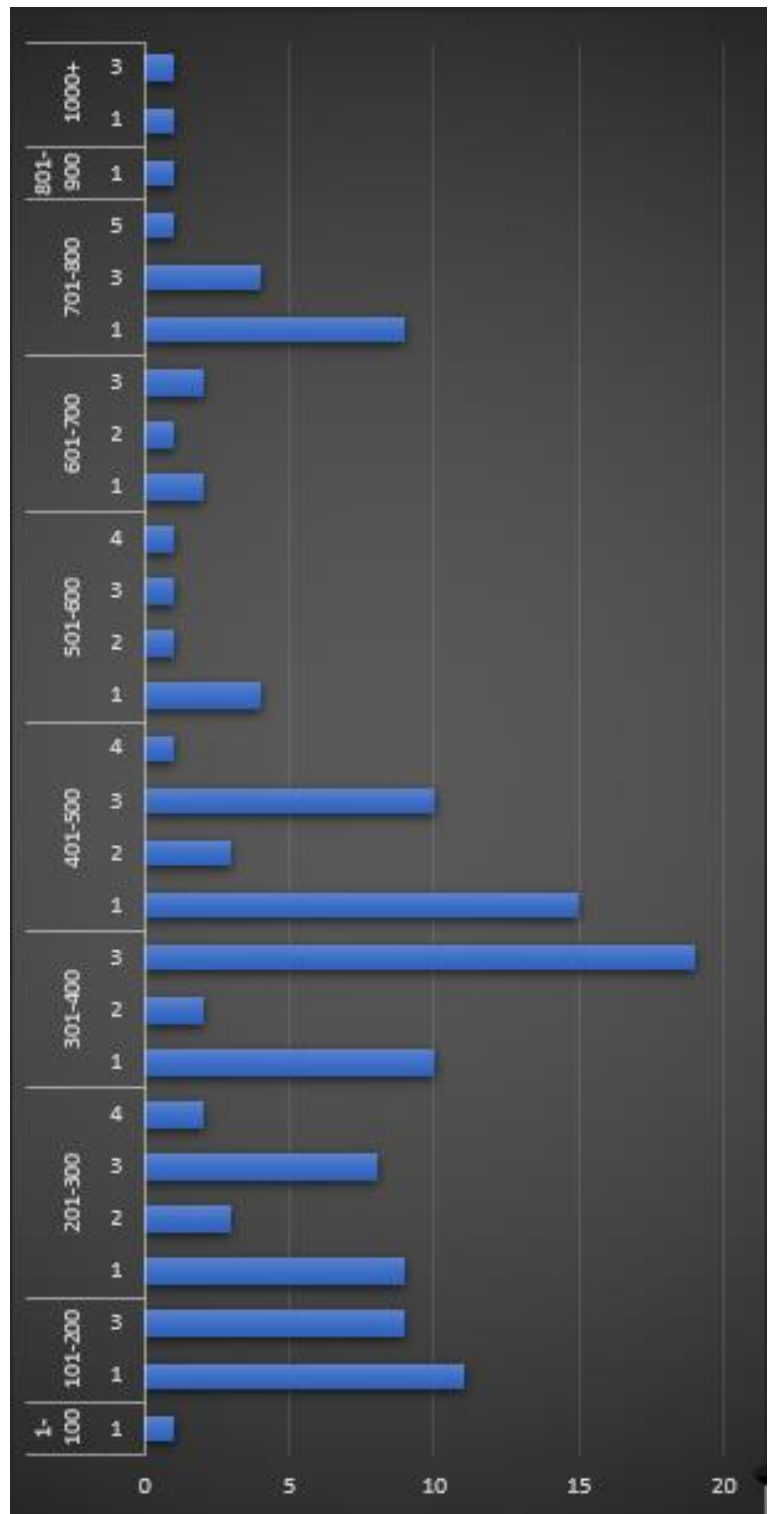
Retailer R2 generates the Maximum Revenue. It generates approximately 46 % of the total revenue.

❖ MRP Range v/s Category of Product in Stock

➤ Pivot Table

Row Labels	Count of X
1-100	1
1	1
101-200	20
1	11
3	9
201-300	22
1	9
2	3
3	8
4	2
301-400	31
1	10
2	2
3	19
401-500	29
1	15
2	3
3	10
4	1
501-600	7
1	4
2	1
3	1
4	1
601-700	5
1	2
2	1
3	2
701-800	14
1	9
3	4
5	1
801-900	1
1	1
1000+	2
1	1
3	1
Grand Total	132

➤ Pivot Chart



➤ Categories based on Inventory Trends

<u>CATEGORY</u>	<u>CRITERIA</u>
1	Products which are never sold during the observation period
2	Products which are first sold are but not sold after its second purchase
3	Products which have several sales - purchase & still available in stock
4	Products which get stock out & purchased again
5	Products which get stock out but not purchased again

➤ Observations

- 📊 In price ranges 1-100 & 801-900 products are not sold during observation period.
- 📊 In every price ranges category 1 products are dominant.
- 📊 Only range 701-800 contains a product of category 5.
- 📊 Category 2 products are in ranges from 201 to 700.
- 📊 Category 4 products are only in ranges 201-300, 401-500 & 501-600.

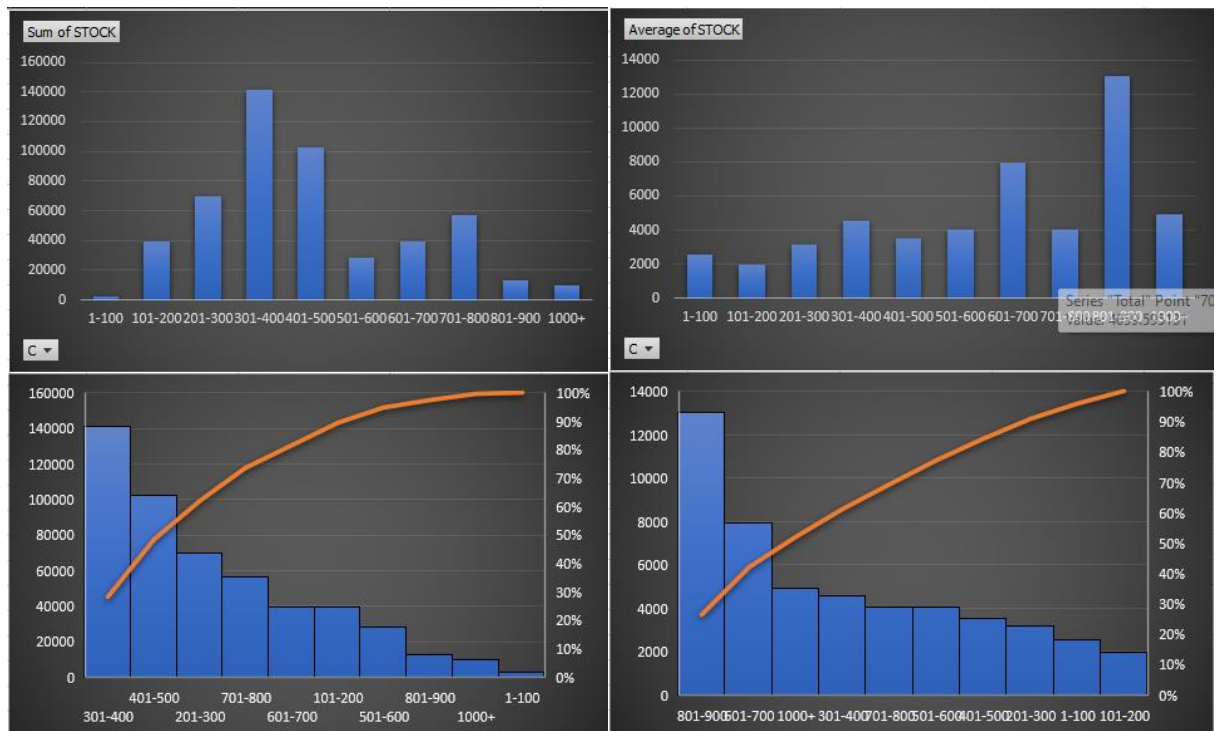
❖ **MRP Ranges v/s Price of available Products in Stock**

➤ Pivot Table

Row Labels	Sum of STOCK
1-100	2534.52672
101-200	39341.23776
201-300	69709.16269
301-400	141494.9385
401-500	102479.5818
501-600	28187.9495
601-700	39780.33558
701-800	56834.33184
801-900	13057.4231
901-1000	0
1000+	9789.20448

Row Labels	Average of STOCK
1-100	2534.52672
101-200	1967.061888
201-300	3168.598304
301-400	4564.352854
401-500	3533.778682
501-600	4026.849929
601-700	7956.067117
701-800	4059.595131
801-900	13057.4231
901-1000	0
1000+	4894.60224

➤ Pivot Chart



➤ Observations

Products in the price range 301-400 acquire maximum amount of stock. In fact products in price ranges 301-400 & 401-500 acquire approximately 48 % amount of stock.

INTERPRETATION OF RESULTS & RECOMMENDATIONS

- ❖ Try to promote and increase sales of products with MRP above Rs 1000 (i.e. Category 1000+). These products have comparatively high profit percentage for both wholesalers and retailers.
- ❖ Guidelines for purchasing products from different price ranges:

<u>CATEGORY</u>	<u>WHAT TO DO</u>
0-100	Stop Purchasing
101-200	Continue Purchasing in Same Quantity
201-300	Decrease Purchase Quantity
301-400	Increase Purchase Quantity
401-500	Increase Purchase Quantity
501-600	Continue Purchasing in Same Quantity
601-700	Decrease Purchase Quantity
701-800	Continue Purchasing in Same Quantity
801-900	Stop Purchasing
901-1000	Start Purchasing
1000+	Increase Purchase Quantity

- ❖ Retailer wise recommendations:

R1	<ul style="list-style-type: none">➤ Promote products from MRP range 201-300 & 501 to 700.➤ Provide some discount to further increase the sales of products from MRP range 301 to 500.
R2	<ul style="list-style-type: none">➤ Promote products from MRP range 501 to 700.➤ Provide some discount to further increase the sales of products from MRP range 101 to 300 & 1000+.➤ As R2 generates highest revenue, provide some offers to build better relations.
R3	<ul style="list-style-type: none">➤ Promote products whose MRP is greater than 700.➤ Provide some discount to further increase the sales of products from MRP range 401 to 700.➤ As R3 generates lowest revenue, try some measures to build trust & increase orders.

❖ Guidelines for different categories of products based on their Inventory trends:

<u>CATEGORY</u>	<u>CRITERIA</u>
1	Products which are never sold during the observation period
2	Products which are first sold are but not sold after its second purchase
3	Products which have several sales - purchase & still available in stock
4	Products which get stock out & purchased again
5	Products which get stock out but not purchased again

<u>CATEGORY</u>	<u>WHAT TO DO</u>
1	Stop Purchasing until the current stock gets sold
2	Stop Purchasing until the current stock gets sold
3	Decrease the purchase of products that acquires large portion in stock & Increase for the products which acquires less
4	Purchase again but not in large quantity
5	Start Purchasing & see the trend for next month

Due to the large number of products under observation, I only mention product categories here according to their inventory trends. For exact product names along with their categories and inventory trendlines, please refer to the "Inventory" Sheet in my "BDM Capstone Project Data" Workbook.

Link to my “BDM Capstone Project Data” Workbook on Google Sheets:

<https://docs.google.com/spreadsheets/d/1n55GdZwYka3N-FhIhTyLPTGDUHI96G4VcdTBuRkRga4/edit?usp=sharing>

THANK YOU