MARKETING & RETAIL ANALYTICS PROJECT

BY NANCY GUPTA

Part A:

Problem Statement:

An automobile parts manufacturing company has collected data on transactions for 3 years. They do not have any in-house data science team, thus they have hired you as their consultant. Your job is to use your data science skills to find the underlying buying patterns of the customers, provide the company with suitable insights about their customers, and recommend customized marketing strategies for different segments of customers.

Agenda & Executive Summary of the data ->

- The purpose and scope of the data is to get suitable insights about the customers and their needs
- Data Dictionary for Sales Dataset :-

Column Name	Description
ORDERNUMBER	This column represents the unique identification number assigned to each order.
QUANTITYORDERED	It indicates the number of items ordered in each order.
PRICEEACH	This column specifies the price of each item in the order.
ORDERLINENUMBER	It represents the line number of each item within an order.
SALES	This column denotes the total sales amount for each order, which is calculated by multiplying the quantity ordered by the price of each item.
ORDERDATE	It denotes the date on which the order was placed.
DAYS_SINCE_LASTORDER	This column represents the number of days that have passed since the last order for each customer. It can be used to analyze customer purchasing patterns.
STATUS	It indicates the status of the order, such as "Shipped," "In Process," "Cancelled," "Disputed," "On Hold," or "Resolved"
PRODUCTLINE	This column specifies the product line categories to which each item belongs.
MSRP	It stands for Manufacturer's Suggested Retail Price and represents the suggested selling price for each item.
PRODUCTCODE	This column represents the unique code assigned to each product.
CUSTOMERNAME	It denotes the name of the customer who placed the order.
PHONE	This column contains the contact phone number for the customer.
ADDRESSLINE1	It represents the first line of the customer's address.
CITY	This column specifies the city where the customer is located.
POSTALCODE	It denotes the postal code or ZIP code associated with the customer's address.
COUNTRY	This column indicates the country where the customer is located.
CONTACTLASTNAME	It represents the last name of the contact person associated with the customer.
CONTACTFIRSTNAME	This column denotes the first name of the contact person associated with the customer.
DEALSIZE	It indicates the size of the deal or order, which are the categories "Small," "Medium," or "Large."

About Data (Info, Shape, Summary Stats, your assumptions about data) ->

First five rows of the data :-

	ORDERNUMBER	QUANTITYORDERED	PRICEEACH	ORDERLINENUMBER	SALES	ORDERDATE	DAYS_SINCE_LASTORDER	STATUS	PRODUCTLINE	MSRP	PRODUCTCODE	CUSTOMERNAME
0	10107	30	95.70	2	2871.00	43155	828	Shipped	Motorcycles	95	S10_1678	Land of Toys Inc.
1	10121	34	81.35	5	2765.90	43227	757	Shipped	Motorcycles	95	S10_1678	Reims Collectables
2	10134	41	94.74	2	3884.34	43282	703	Shipped	Motorcycles	95	S10_1678	Lyon Souveniers
3	10145	45	83.26	6	3746.70	43337	649	Shipped	Motorcycles	95	S10_1678	Toys4GrownUps.com
4	10168	36	96.66	1	3479.76	43401	586	Shipped	Motorcycles	95	S10_1678	Technics Stores Inc.

CUSTOMERNAME	PHONE	ADDRESSLINE1	CITY	POSTALCODE	COUNTRY	CONTACTLASTNAME	CONTACTFIRSTNAME	DEALSIZE
Land of Toys Inc.	2125557818	897 Long Airport Avenue	NYC	10022	USA	Yu	Kwai	Small
Reims Collectables	26.47.1555	59 rue de l'Abbaye	Reims	51100	France	Henriot	Paul	Small
Lyon Souveniers	+33 1 46 62 7555	27 rue du Colonel Pierre Avia	Paris	75508	France	Da Cunha	Daniel	Medium
Toys4GrownUps.com	6265557265	78934 Hillside Dr.	Pasadena	90003	USA	Young	Julie	Medium
Technics Stores Inc.	6505556809	9408 Furth Circle	Burlingame	94217	USA	Hirano	Juri	Medium

Information of the data :-

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2747 entries, 0 to 2746
Data columns (total 20 columns):
 #   Column                Non-Null Count  Dtype  
---  ------                --------------  -----  
 0   ORDERNUMBER           2747 non-null   int64  
 1   QUANTITYORDERED       2747 non-null   int64  
 2   PRICEEACH             2747 non-null   float64
 3   ORDERLINENUMBER       2747 non-null   int64  
 4   SALES                 2747 non-null   float64
 5   ORDERDATE             2747 non-null   int64  
 6   DAYS_SINCE_LASTORDER  2747 non-null   int64  
 7   STATUS                2747 non-null   object 
 8   PRODUCTLINE           2747 non-null   object 
 9   MSRP                  2747 non-null   int64  
 10  PRODUCTCODE           2747 non-null   object 
 11  CUSTOMERNAME          2747 non-null   object 
 12  PHONE                 2747 non-null   object 
 13  ADDRESSLINE1          2747 non-null   object 
 14  CITY                  2747 non-null   object 
 15  POSTALCODE            2747 non-null   object 
 16  COUNTRY               2747 non-null   object 
 17  CONTACTLASTNAME       2747 non-null   object 
 18  CONTACTFIRSTNAME      2747 non-null   object 
 19  DEALSIZE              2747 non-null   object 
dtypes: float64(2), int64(6), object(12)
memory usage: 429.3+ KB
```

Inferences:-

- If we check the information of the data, we will find that the data has 2747 entries with a total of 20 columns.
- Out of the 20 columns , 2 columns are of float type, 6 columns are od integer type and the rest of the 12 columns are object type.
- It takes up a memory of 429.3 KB

- The below picture clearly defines that there is no null value in the data.
- There are no duplicated rows in the data as well.

```
ORDERNUMBER             0
QUANTITYORDERED         0
PRICEEACH               0
ORDERLINENUMBER         0
SALES                   0
ORDERDATE               0
DAYS_SINCE_LASTORDER    0
STATUS                  0
PRODUCTLINE             0
MSRP                    0
PRODUCTCODE             0
CUSTOMERNAME            0
PHONE                   0
ADDRESSLINE1            0
CITY                    0
POSTALCODE              0
COUNTRY                 0
CONTACTLASTNAME         0
CONTACTFIRSTNAME        0
DEALSIZE                0
dtype: int64
```

```
Number of duplicate rows = 0
```

- Below is the picture depicting a brief description of the data in this dataset.

	ORDERNUMBER	QUANTITYORDERED	PRICEEACH	ORDERLINENUMBER	SALES	ORDERDATE	DAYS_SINCE_LASTORDER	MSRP
count	2747.000000	2747.000000	2747.000000	2747.000000	2747.000000	2747.000000	2747.000000	2747.000000
mean	10259.761558	35.103021	101.098951	6.491081	3553.047583	43598.914088	1757.085912	100.691664
std	91.877521	9.762135	42.042548	4.230544	1838.953901	230.231295	819.280576	40.114802
min	10100.000000	6.000000	26.880000	1.000000	482.130000	43106.000000	42.000000	33.000000
25%	10181.000000	27.000000	68.745000	3.000000	2204.350000	43412.000000	1077.000000	68.000000
50%	10264.000000	35.000000	95.550000	6.000000	3184.800000	43640.000000	1761.000000	99.000000
75%	10334.500000	43.000000	127.100000	9.000000	4503.095000	43786.000000	2436.500000	124.000000
max	10425.000000	97.000000	252.870000	18.000000	14082.800000	43982.000000	3562.000000	214.000000

Exploratory Analysis and Inferences :-

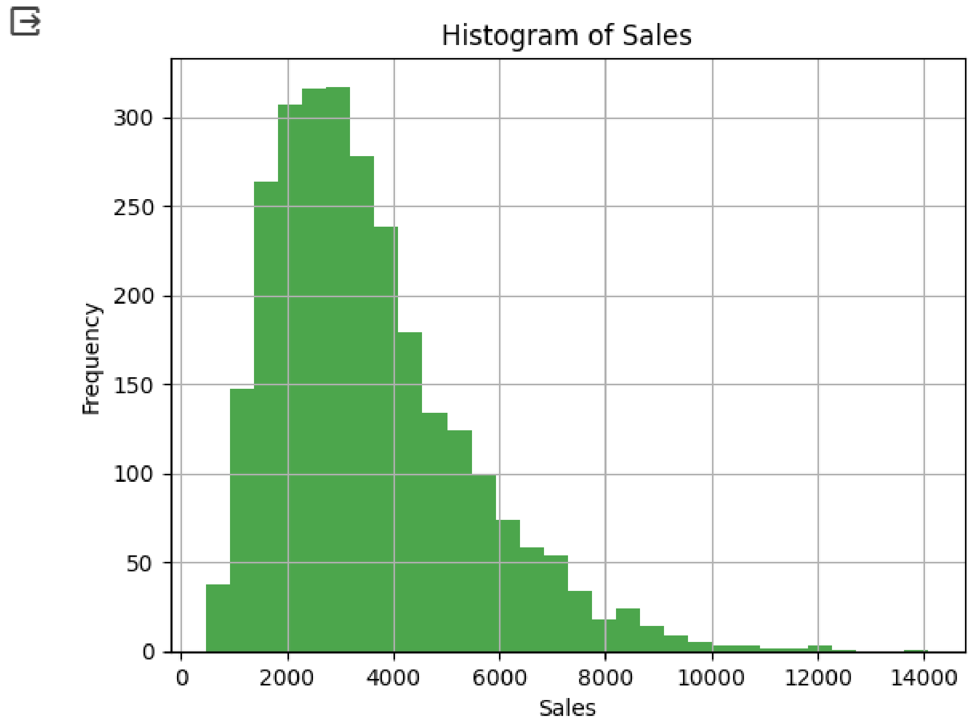


Fig 1:1

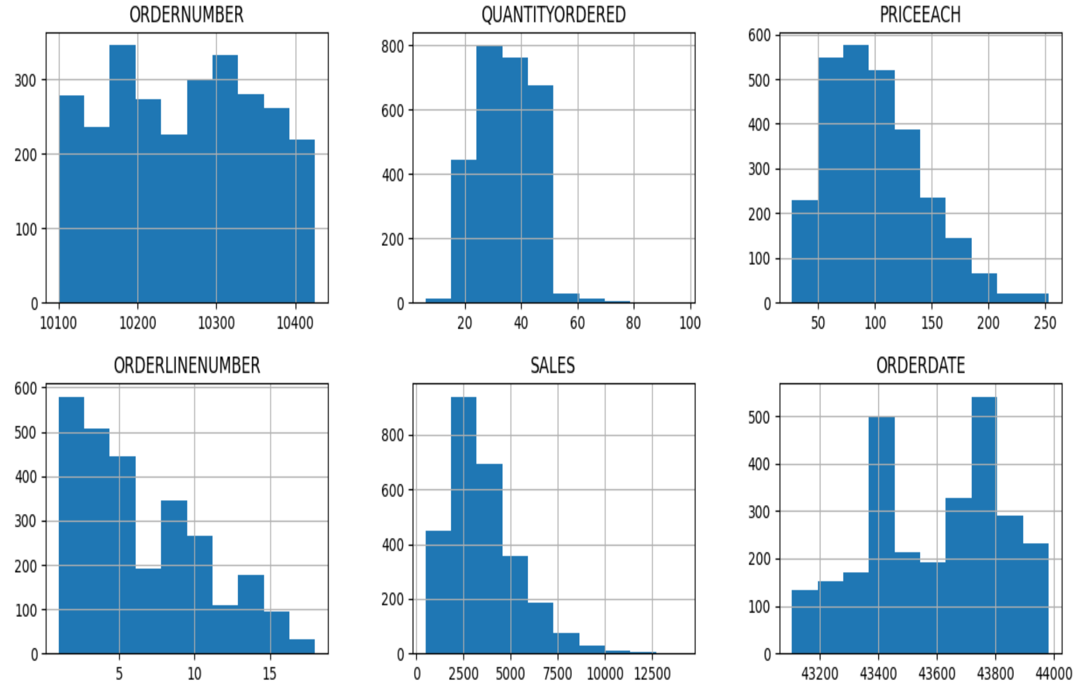


Fig 1:2

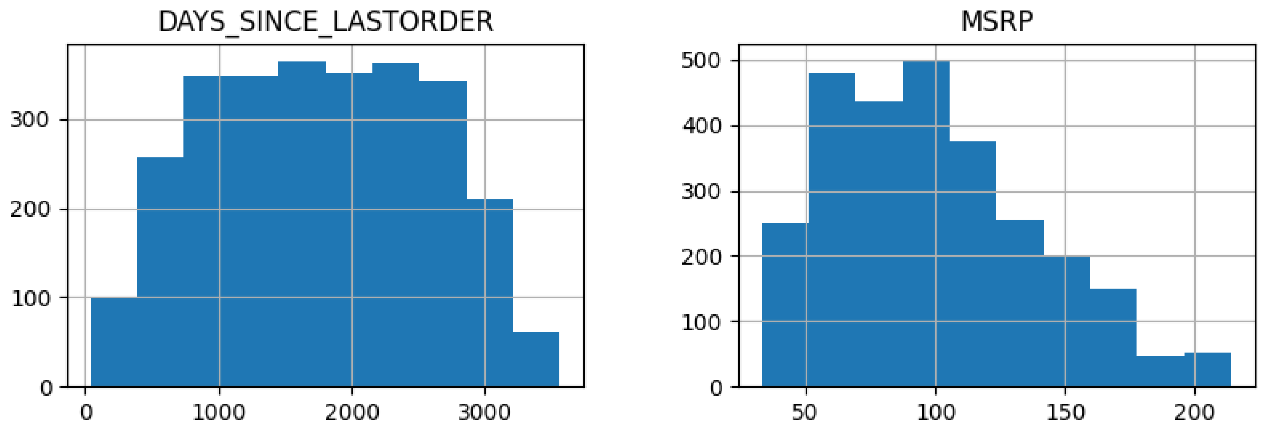
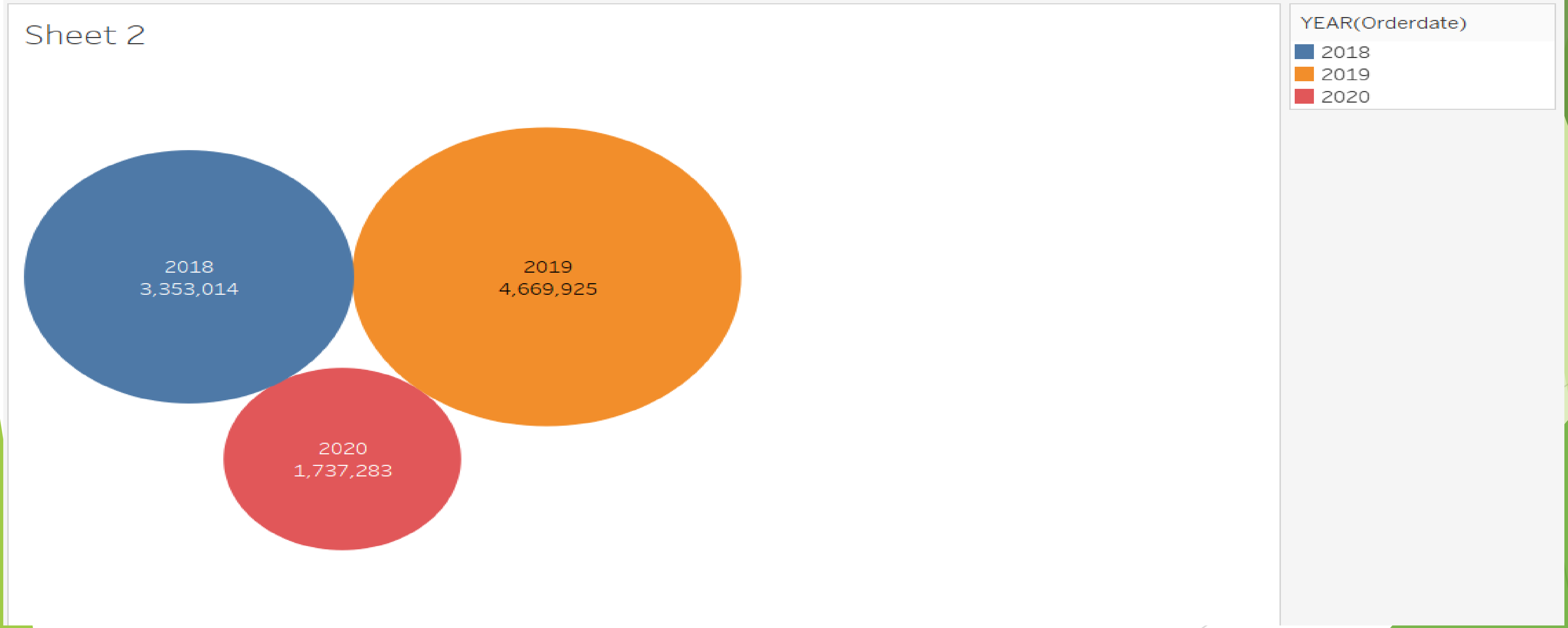


Fig 1:3

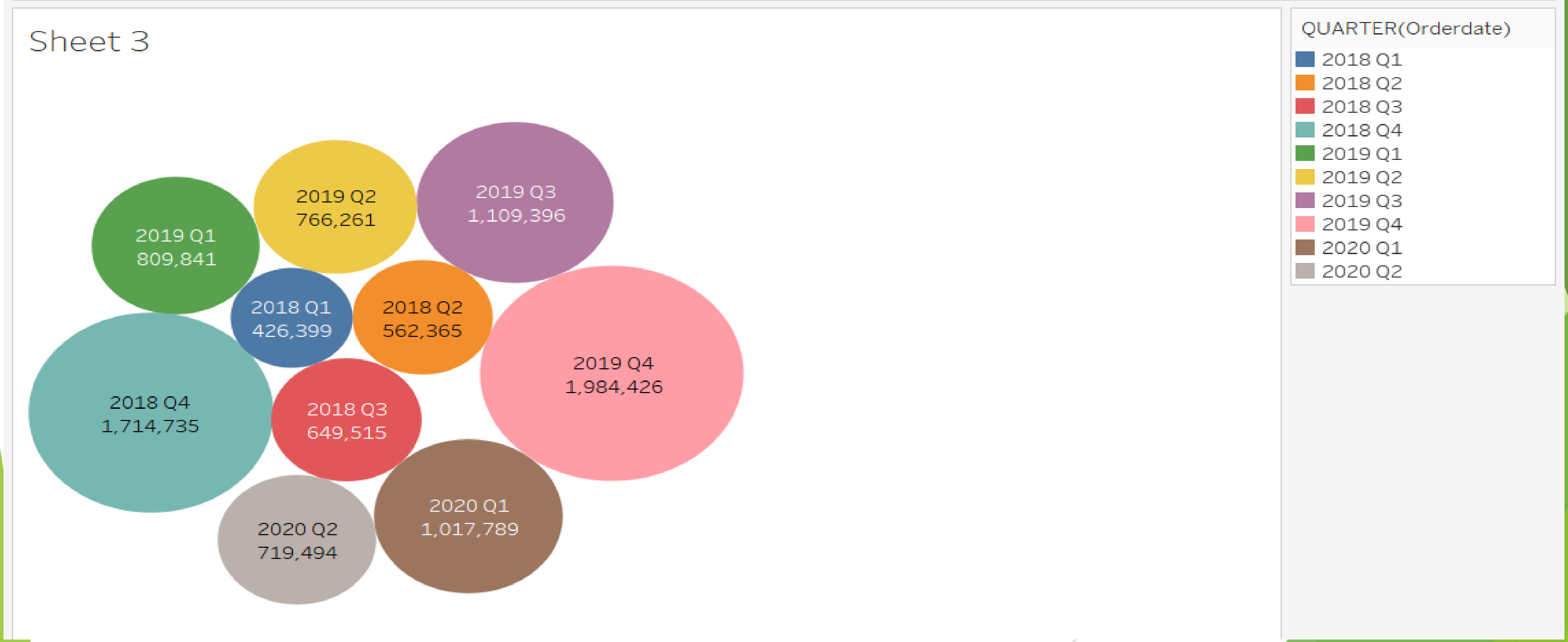
- Fig 1:1 is the histogram of sales particularly according to which the sales are high between 2000 to 4000
- 50% of the sales is in the bracket between 2000 to 4000
- Sales are least between 12000 to 14000
- Above 75% the sales are less
- Fig 1:2 and 1:3 depicts the details of all the 8 numeric variables.

Visual representation of total number of sales in 2018, 2019 and 2020 :

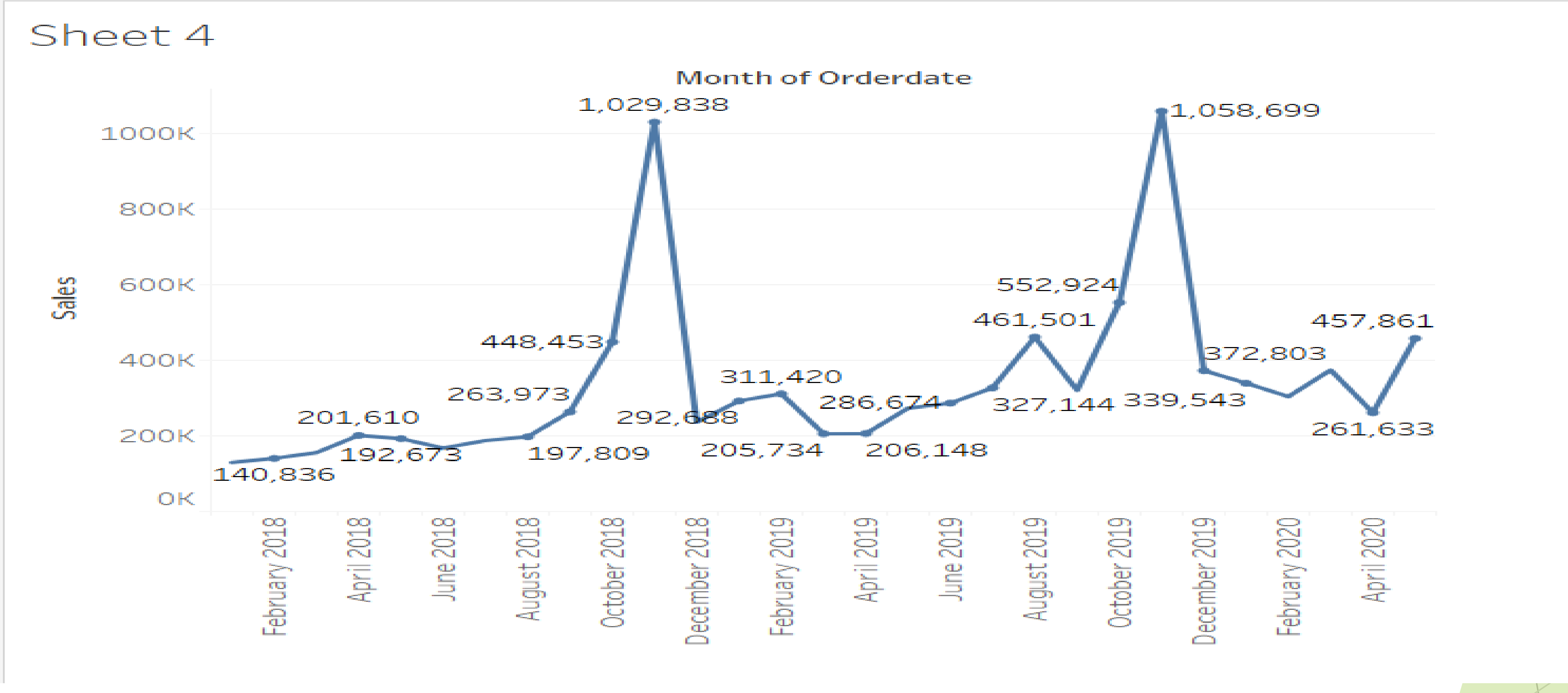


- The above picture depicts the total number of sales for 2018, 2019, 2020.
- The total number of sales for 2018 is 3353014
- The total number of sales for 2019 is 4669925
- The total number of sales for 2020 is 1737283
- The total Sales for 3 years altogether are 9760222

Visual representation of Quarterly Sales:

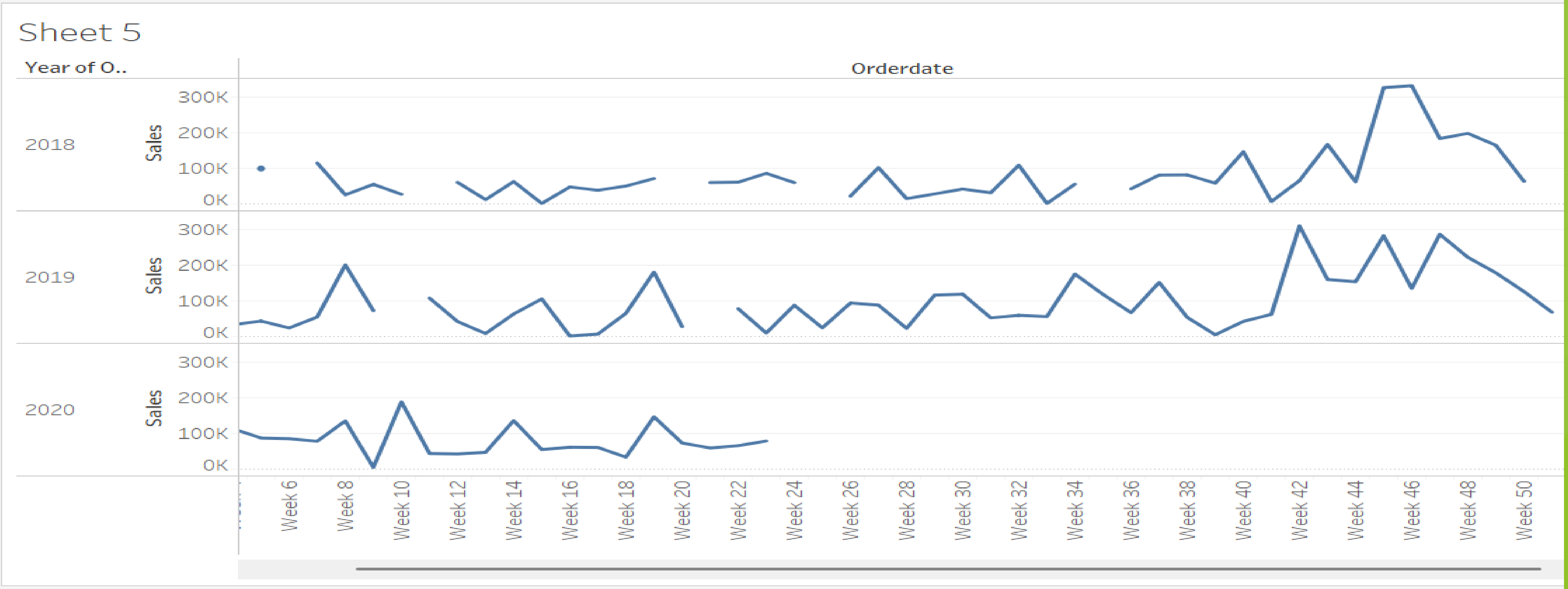


Visual representation of Monthly Sales over 3 years :



- The maximum sale we find in the month of November for both the years 2018 and 2019.
- October month for both the years also gives moderate sale however a sudden drop of sale is seen in the month of December.
- From January to August we find the sales between 200K to 250K
- In the month of May 2020 we do find positive trend on sales

Visual representation of Weekly sales over the 3 years :



- 10th week of 2020 shows a positive trend on sales however there has been barely any sales in the previous 2 years hence it is an improvement.
- If we see the sale trend till week 23 of all the 3 years we see a significant improvement for the year 2020 as the sales are continuous. There has been ups and downs but there is a constant inflow but in the year 2018 and 2019 we do find discontinuity.

Visual representation of Sales Vs Dealsize :

- ➢ Medium size orders have given the most number of sales.
- ➢ Large size orders have given least number of sales.
- ➢ There are 9,019,094 amount of order in the shipped status which is a matter of concern or a factor to look into.
- ➢ The other important look out will be the On Hold status which is 178,979.
- ➢ We should also look into the cancelled orders which is amounting to be 194,487 and need to find out as to why the orders got cancelled.

Summary on Sales Vs Status :

Sheet 7

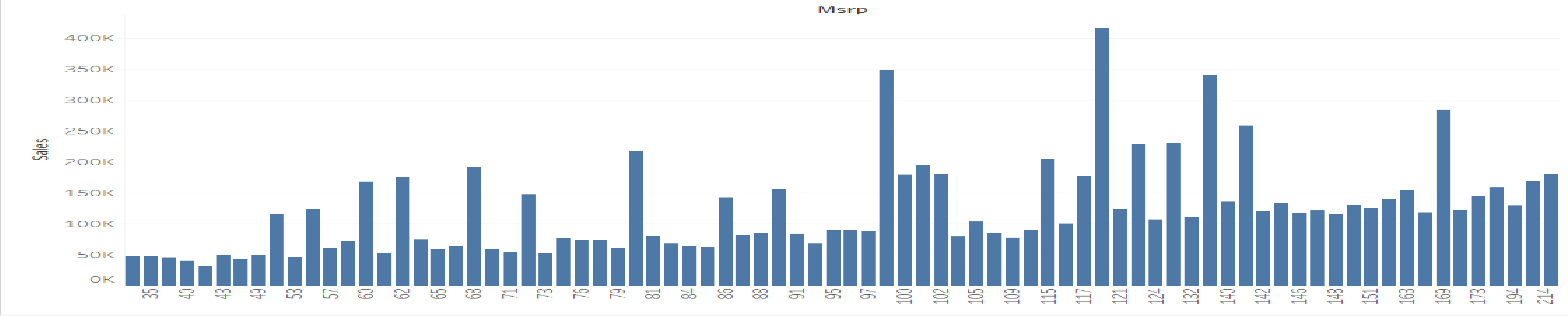
Status	
Cancelled	194,487
Disputed	72,213
In Process	144,730
On Hold	178,979
Resolved	150,718
Shipped	9,019,094

SUM(Sales)

72,213 9,019,094

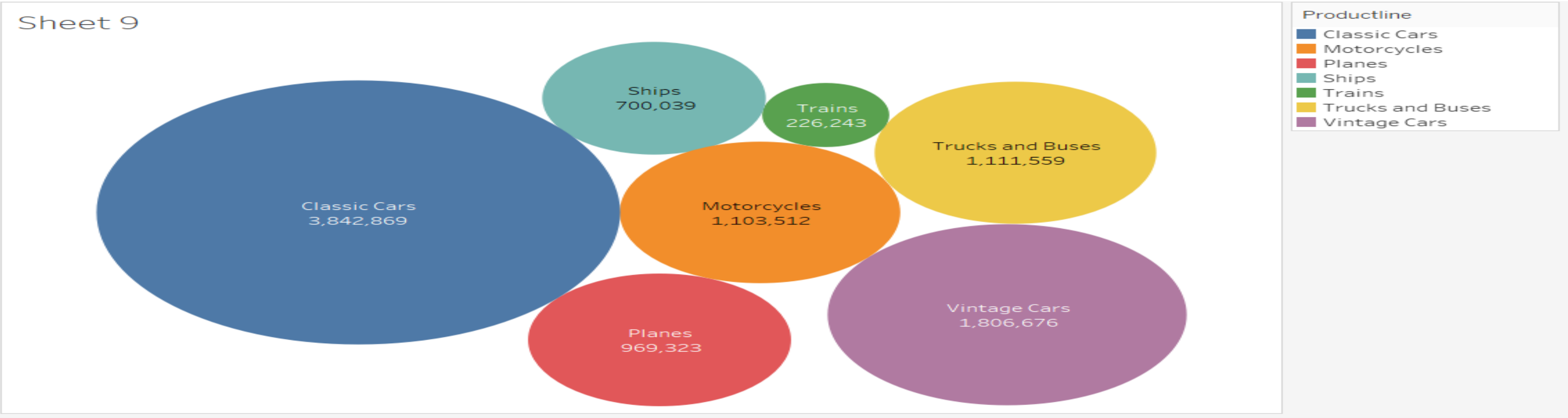
Visual representation of MSRP Vs Sales :

Sheet 8



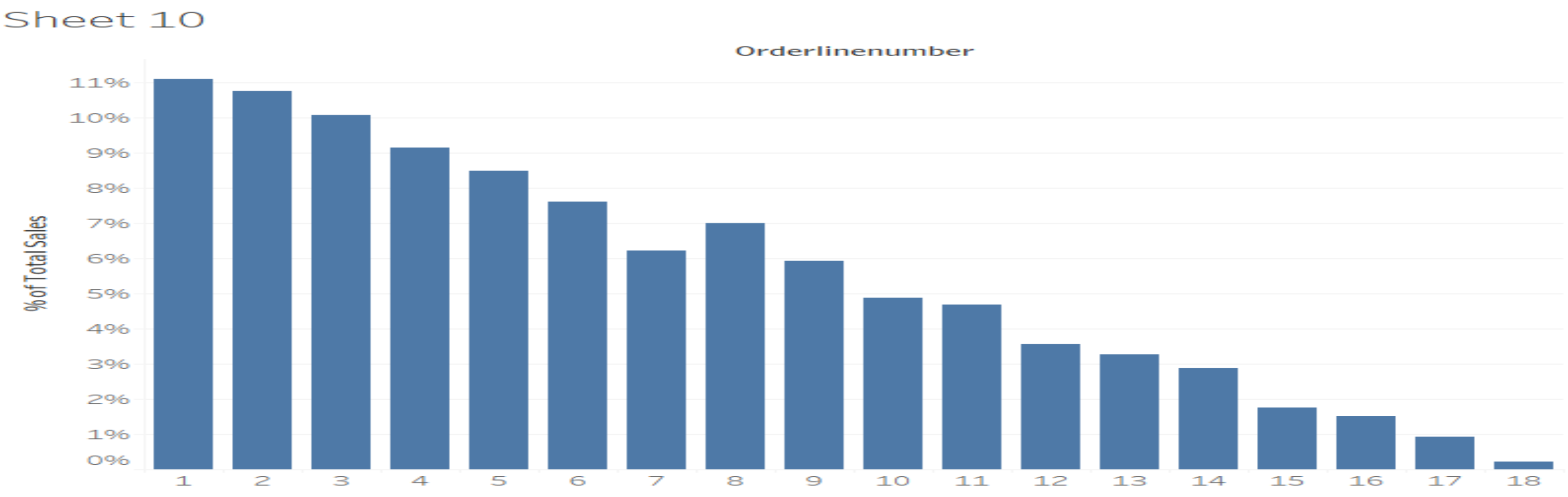
- The below plot shows the sales across the various product lines.
- Classic cars are seen to generate most revenue.
- Trains are seen to generate the least revenue.

Visual representation of Product line Vs Sales:

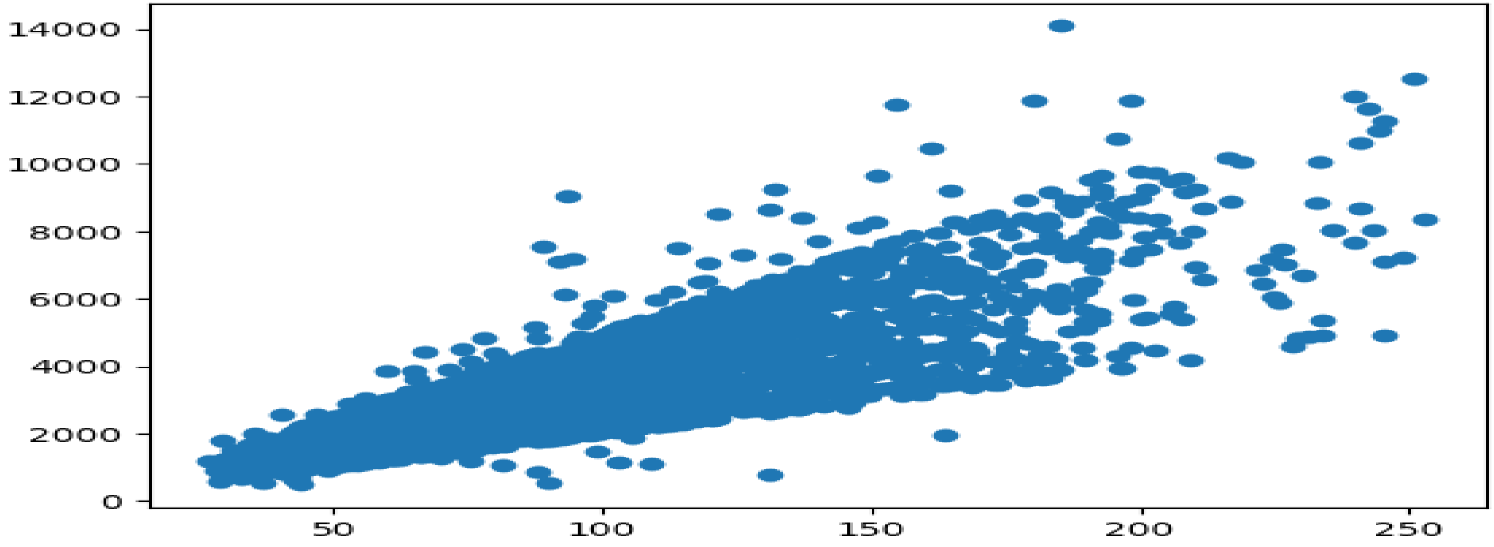


- After Classic cars vintage cars are in demand.
- Trucks and Buses and Motorcycles have very little difference among them.
- Order line number labelled as 1 has the high sales percentage followed by number 2 in the below mentioned figure.
- Number 18 has got the least percentage of total sales.

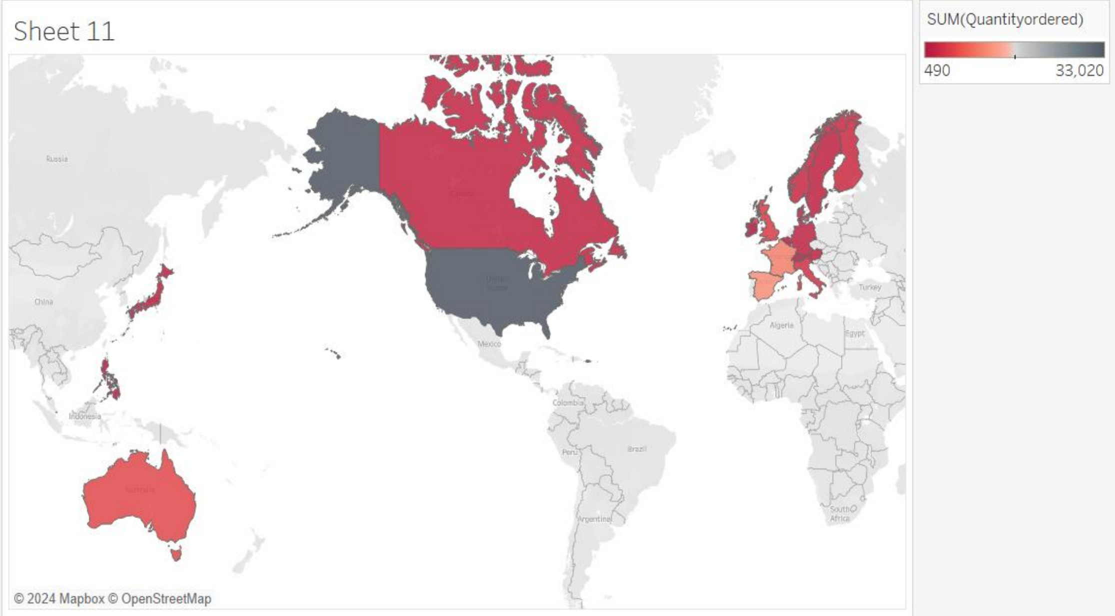
Visual representation of Order line number Vs percentage of total Sales :

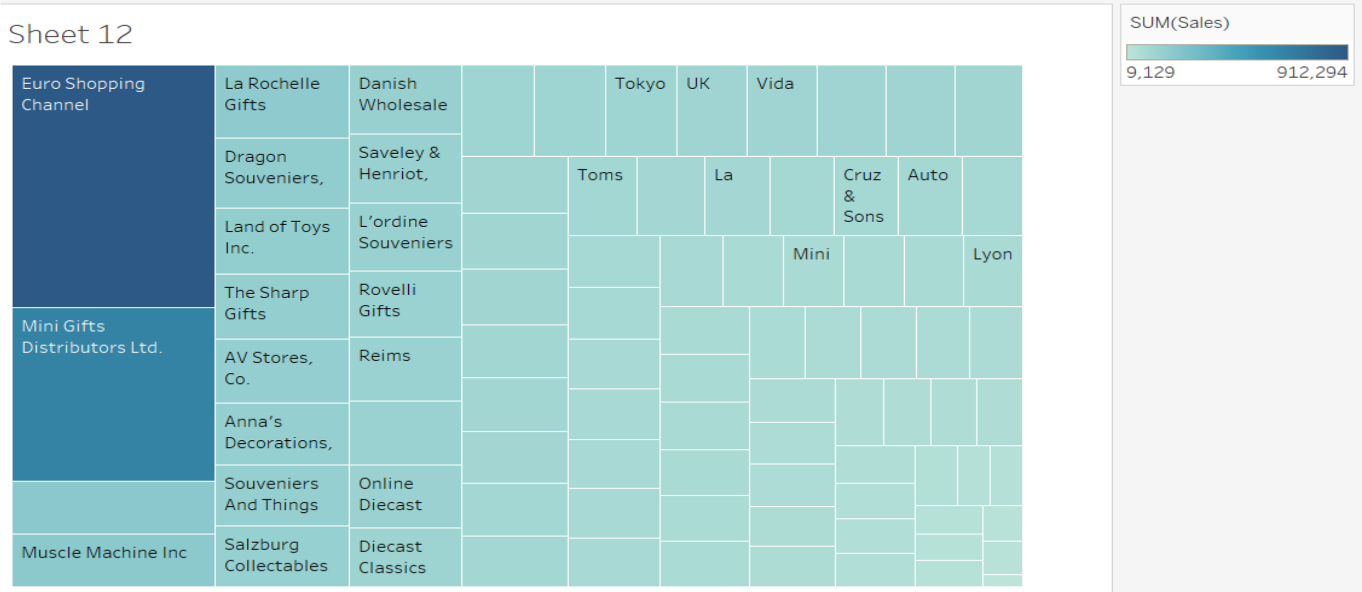


<matplotlib.collections.PathCollection at 0x7f41e072ffd0>

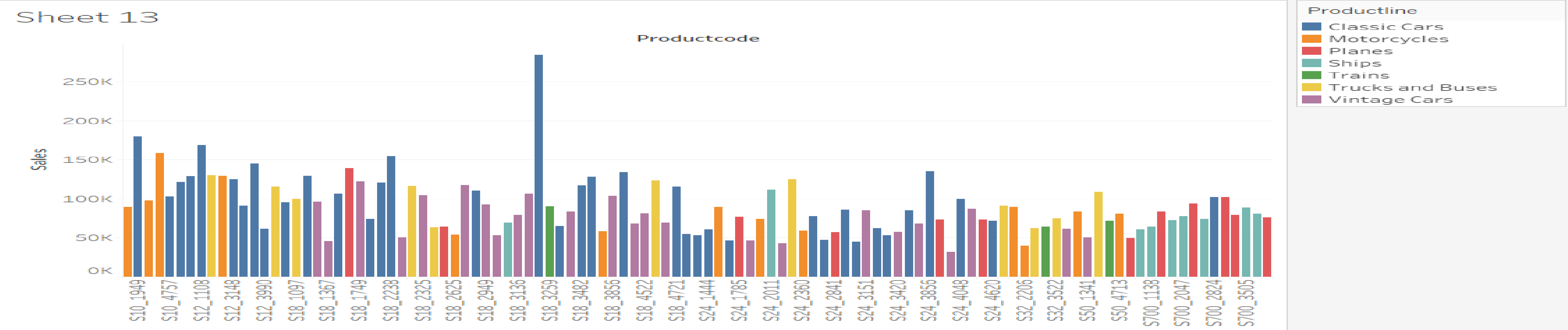


- The above plot shows the scattered data between price of each item in the order with their total number of sales.
- Price of items between 20 to 200 has high demand.

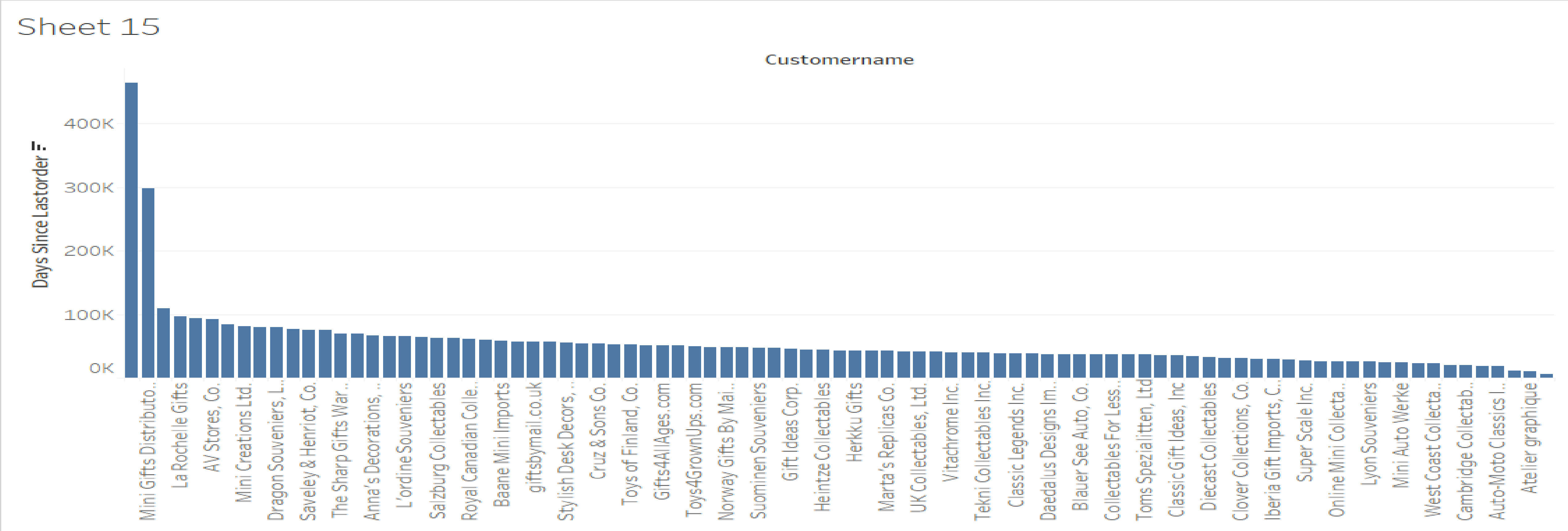




- ➢ The above two graphs represent the country wise sale distribution and number of sales with respect to the customer's name.
- ➢ Euro Shopping Channel is the most loyal customer followed by Mini gifts distributions limited.
- ➢ USA has the most number of quantity ordered that is 33,020.

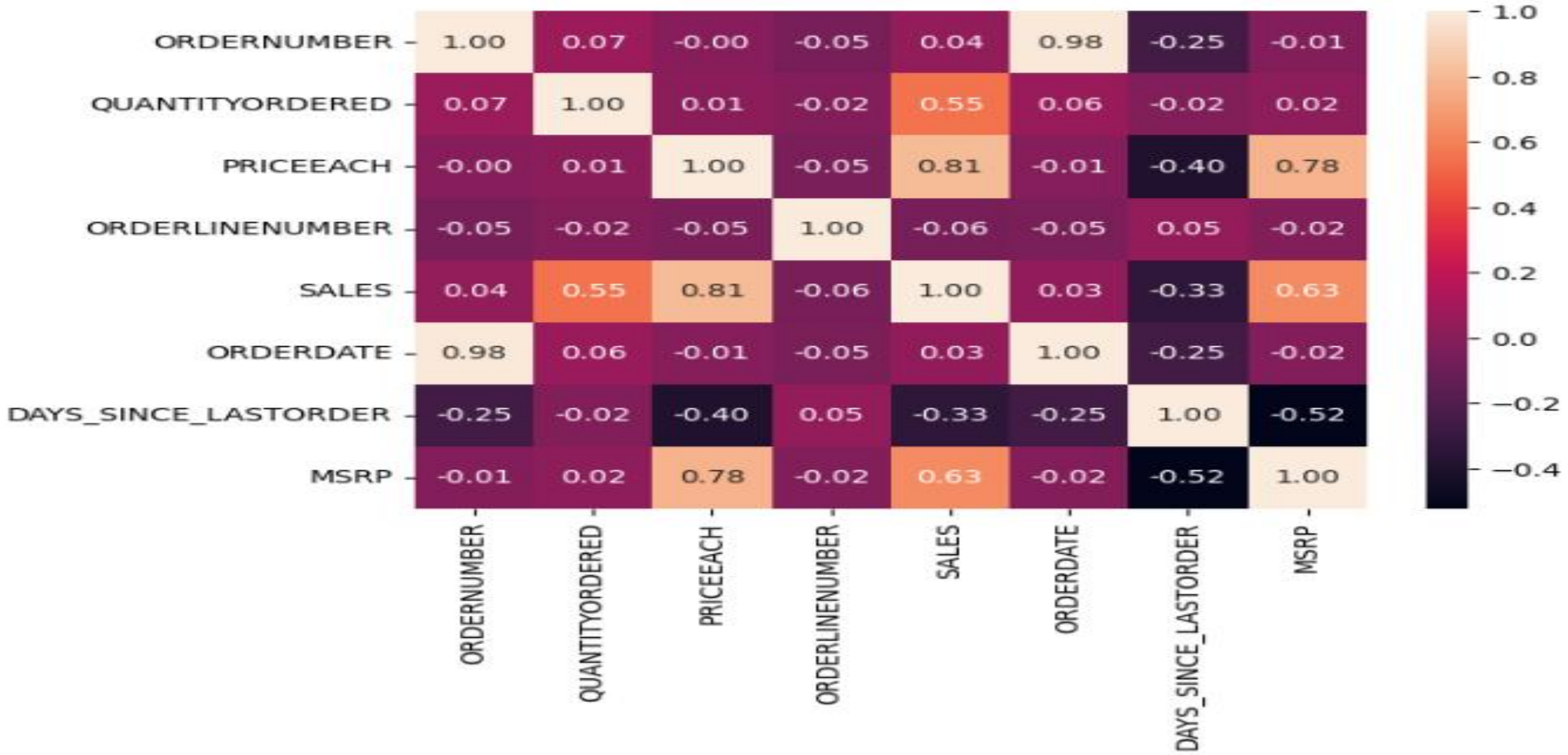


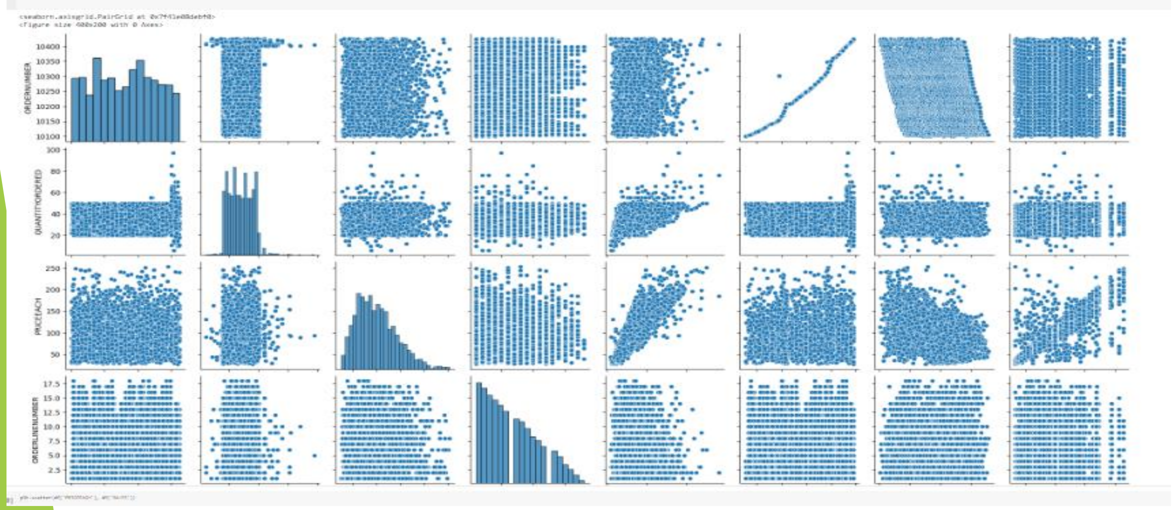
- The above graph shows the sales on each product code.
- S18_3232 product code has the highest sale of 284249
- S24_3969 product code has the lowest sale of 31739
- This plot displays different colors on each product line of this graph.

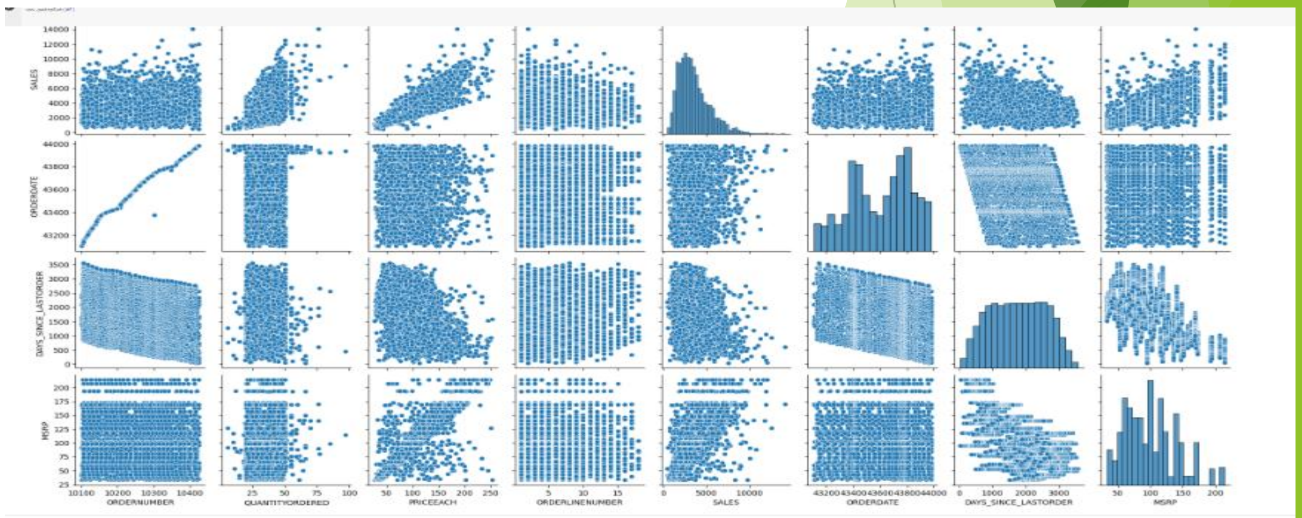


- The above graph represents the customer’s name against the days since last ordered.

- The above graph represents the customer's name against the days since last ordered.







Inferences:-

- There is a high sale on large items as per the graph on sale and deal size.
- We also see that Sales calculated in these countries ie. USA, Canada, Australia, Ireland, UK, Spain, Italy, France, Switzerland, Austria, Belgium, Germany, Denmark, Norway, Sweden, Finland, Singapore, Philippines, Japan.
- In the above graph shows a trend in the year of sales.
- The sale was good in the year 2018 and then went high in the year 2019 but then the sale dropped in the year 2020.
- There is also a graph plotted above on sales with country and city which clearly shows the high rate of sale in USA and the low rate of sale in Belgium.
- The Price on each variable is highest on the shipped Status.
- Under the Status variable the category shipped is high on Sales.
- Classic cars are sold more in number and high sales.
- Sales are high on quantity ordered between 20 and 50 items.
- The above graphs show how the Sales variable differ accordingly with other variables.
- The 3-year-old data shows how frequent customers have come to purchase these products or parts of the automobiles.
- A comparison has been done with variable mostly with sales so that a clear interpretation can be made as to where the sales going low and why some customers are not coming back to purchase the products.
- There is a high sale on large items as per the graph on sale and deal size.

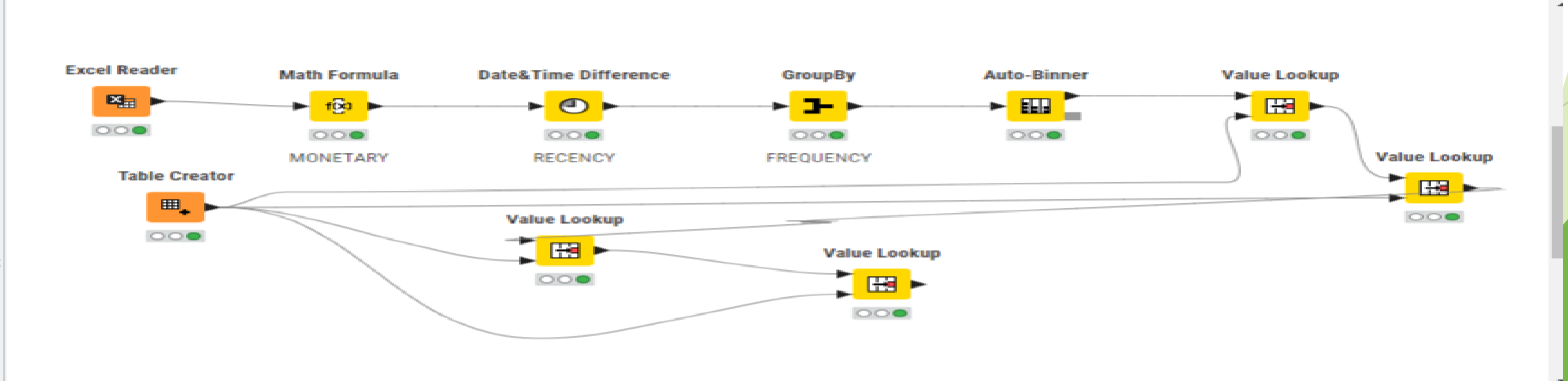
RFM:-

RFM stands for **Recency**, **Frequency**, and **Monetary** value. It is a method used by businesses for analyzing customer value. Here's what each component represents:

1. **Recency**: This refers to how recently a customer has made a purchase. Customers who have made purchases more recently are often considered more valuable because they are more likely to make repeat purchases.
2. **Frequency**: This refers to how often a customer makes purchases. Customers who make frequent purchases are often considered more valuable because they demonstrate loyalty and ongoing engagement with the brand.
3. **Monetary** : This refers to the total amount of money a customer has spent on purchases. Customers who have spent more money are often considered more valuable because they contribute more directly to the revenue of the business.

By analyzing these three factors together, businesses can segment their customer base into different groups based on their value to the company. This allows businesses to tailor their marketing strategies and customer engagement efforts more effectively, focusing on retaining and growing their most valuable customers.

Below is showcased the Knime Workflow Image:



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
1	ORDER	QUANT	PRICEE	ORDER	SALES	ORDER	DAYS_S	MONET	RECEN(	MSRP	PRODU	DEALSI	CUSTO	QUANT	SALES [	MONET	RECEN(	Monet(	Frequer	Recenc
5	10103	16	1642.25	16	54702	16	878	54702	2246	105.938	16	16	16	Bin 4	Bin 4	Bin 4	Bin 4	4	4	1
7	10105	15	1609.76	15	58871.1	15	939	58871.1	2233	107.267	15	15	15	Bin 4	Bin 4	Bin 4	Bin 4	4	4	1
8	10106	18	1531.04	18	56181.3	18	1361	56181.3	2227	84.7778	18	18	18	Bin 4	Bin 4	Bin 4	Bin 4	4	4	1
10	10108	16	1545.12	16	55245	16	971	55245	2213	99.4375	16	16	16	Bin 4	Bin 4	Bin 4	Bin 4	4	4	1
12	10110	16	1413.36	16	51017.9	16	1307	51017.9	2198	90.75	16	16	16	Bin 4	Bin 4	Bin 4	Bin 4	4	4	1
21	10120	15	1457.33	15	50397.7	15	815	50397.7	2156	94.8	15	15	15	Bin 4	Bin 4	Bin 4	Bin 4	4	4	1

- ➢ These are the top 5 best customers with highest monetary value, highest Frequency Value and least Recency.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
1	ORDER	QUANT	PRICEE	ORDER	SALES	ORDER	DAYS_S	MONET	RECEN(	MSRP	PRODU	DEALSI	CUSTO	QUANT	SALES [	MONET	RECEN(	Monet(	Frequer	Recenc
149	10264	7	543.67	7	19548.4	7	1492	19548.4	1729	83.1429	7	7	7	Bin 2	Bin 2	Bin 2	Bin 2	2	2	3
152	10267	6	566.53	6	23252.2	6	1683	23252.2	1722	90.5	6	6	6	Bin 2	Bin 2	Bin 2	Bin 2	2	2	3
156	10272	6	799.84	6	27149.3	6	504	27149.3	1709	132.333	6	6	6	Bin 2	Bin 2	Bin 2	Bin 2	2	2	3
163	10279	6	531.34	6	21986.3	6	1651	21986.3	1689	90.5	6	6	6	Bin 2	Bin 2	Bin 2	Bin 2	2	2	3
181	10297	7	651.26	7	18972	7	1021	18972	1651	89.4286	7	7	7	Bin 2	Bin 2	Bin 2	Bin 2	2	2	3
191	10307	9	871.71	9	27445.3	9	344	27445.3	1623	93.2222	9	9	9	Bin 2	Bin 2	Bin 2	Bin 2	2	2	3

- ➢ These are the top list of customers at the verge of churning.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
1	ORDER	QUANT	PRICEE	ORDER	SALES	ORDER	DAYS_S	MONET	RECEN(	MSRP	PRODU	DEALSI	CUSTO	QUANT	SALES [	MONET	RECEN(	Monet(	Frequer	Recenc
9	10107	8	925.49	8	25783.8	8	828	25783.8	2220	112.875	8	8	8	Bin 2	Bin 2	Bin 2	Bin 4	2	2	1
11	10109	6	750.92	6	27398.8	6	1241	27398.8	2206	127	6	6	6	Bin 2	Bin 2	Bin 2	Bin 4	2	2	1
30	10129	9	826.75	9	32376.3	9	820	32376.3	2112	91.1111	9	9	9	Bin 2	Bin 2	Bin 2	Bin 4	2	2	1
32	10131	8	664.19	8	20351	8	1244	20351	2108	83.75	8	8	8	Bin 2	Bin 2	Bin 2	Bin 4	2	2	1
33	10133	8	640.36	8	22167.7	8	1457	22167.7	2097	85.5	8	8	8	Bin 2	Bin 2	Bin 2	Bin 4	2	2	1
34	10134	7	817.74	7	25624.9	7	703	25624.9	2093	114.857	7	7	7	Bin 2	Bin 2	Bin 2	Bin 4	2	2	1

- ➢ These are the top 5 most loyal customers.

	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W
1	ORDER	SALES	ORDER	DAYS_S	MONET	RECEN(	MSRP	PRODU	DEALSI	CUSTO	QUANT	SALES [	MONET	RECEN(	Moneto	Freque	Recenc	Sales	Moneto	Frequen
160	14	52505.4	14	467	52505.4	1696	100.643	14	14	14	Bin 3	Bin 4	Bin 4	Bin 2	4	3	3	4	433	
166	13	52807.2	13	474	52807.2	1678	120.692	13	13	13	Bin 3	Bin 4	Bin 4	Bin 2	4	3	3	4	433	
169	13	48828.7	13	292	48828.7	1671	101	13	13	13	Bin 3	Bin 4	Bin 4	Bin 2	4	3	3	4	433	
175	14	53941.7	14	307	53941.7	1659	107.071	14	14	14	Bin 3	Bin 4	Bin 4	Bin 2	4	3	3	4	433	
189	14	54251.7	14	372	54251.7	1624	107.286	14	14	14	Bin 3	Bin 4	Bin 4	Bin 2	4	3	3	4	433	
206	14	55570.6	14	253	55570.6	1602	104.786	14	14	14	Bin 3	Bin 4	Bin 4	Bin 2	4	3	3	4	433	

- These are the top 5 most lost customers.

Part B:

Problem Statement:

A grocery store shared the transactional data with you. Your job is to conduct a thorough analysis of Point of Sale (POS) data, identify the most commonly occurring sets of items in the customer orders, and provide recommendations through which a grocery store can increase its revenue by popular combo offers & discounts for customers.

1) Exploratory Analysis --> Exploratory Analysis of data & an executive summary (in PPT) of your top findings, supported by graphs. --> Are there trends across months/years/quarters/days etc. that you are able to notice?

	Date	Order_id	Product
0	01-01-2018	1	yogurt
1	01-01-2018	1	pork
2	01-01-2018	1	sandwich bags
3	01-01-2018	1	lunch meat
4	01-01-2018	1	all- purpose

	Date	Order_id	Product
20636	25-02-2020	1138	soda
20637	25-02-2020	1138	paper towels
20638	26-02-2020	1139	soda
20639	26-02-2020	1139	laundry detergent
20640	26-02-2020	1139	shampoo

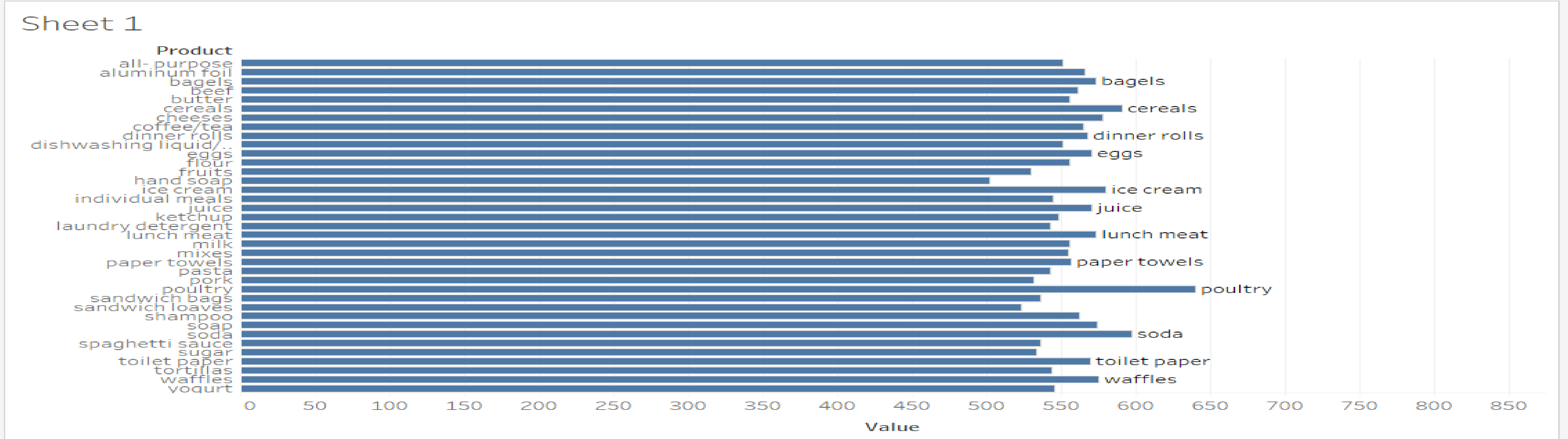
- The above pictures depicts the 5 first rows of the data and 5 last rows of the data.
- In this dataset we find 20641 entries and only 3 columns with Date, order_id and product.

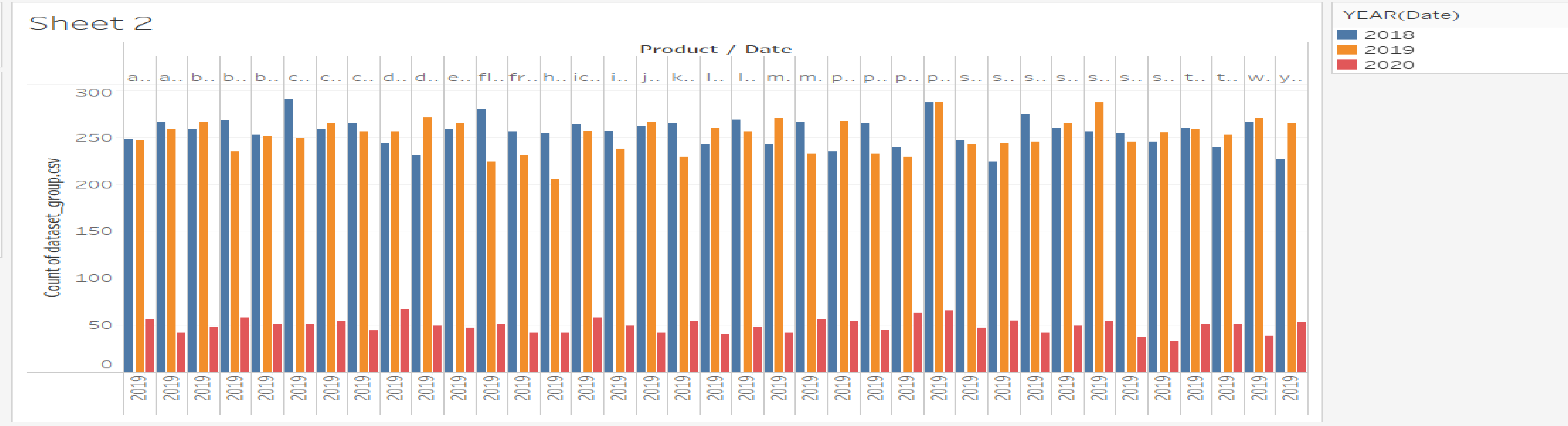
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 20641 entries, 0 to 20640
Data columns (total 3 columns):
 #   Column    Non-Null Count  Dtype 
---  ------    --------------  ----- 
 0   Date      20641 non-null  object
 1   Order_id  20641 non-null  int64 
 2   Product   20641 non-null  object
dtypes: int64(1), object(2)
memory usage: 483.9+ KB
```

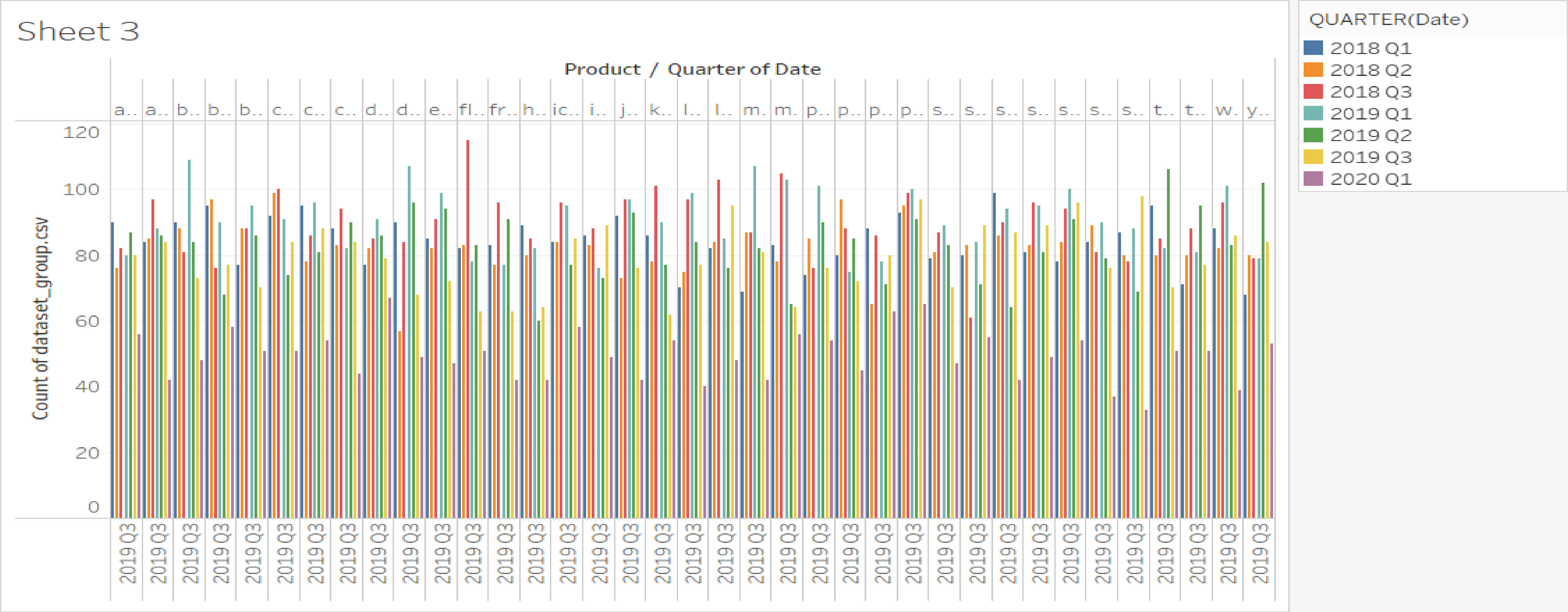
```
Number of duplicate rows = 4730
```

```
Date        0
Order_id    0
Product     0
dtype: int64
```

- If we see the above data images , we find that this dataset has 2 object type column and only 1 integer type column.
- We do not find any null values
- Duplicated rows however are 4730.
- Duplicate need not be treated since data has been duplicated as multiple orders placed by the same customer on one particular date.
- The below mentioned sheet1 image shows the product count along with product names.
- The below mentioned sheet 2 image shows the yearly basis analysis
- There is a slight variation in the yearly pattern.
- It has 2018, 2019 and 2020 year data.
- In comparison to the year 2020, 2018 and 2019 is high.

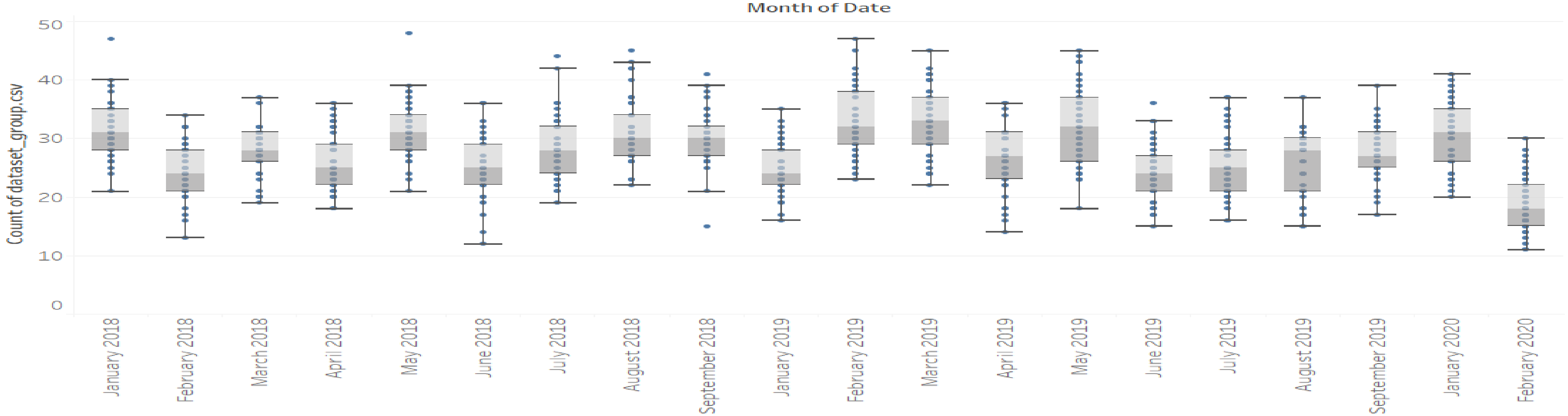

Sheet 1
Product
all- purpose
aluminum foil
bagels
beef
butter
cereals
cheeses
coffee/tea
dinner rolls
dishwashing liquid/..
eggs
flour
fruits
hand soap
ice cream
individual meals
juice
ketchup
laundry detergent
lunch meat
milk
mixes
paper towels
pasta
pork
poultry
sandwich bags
sandwich loaves
shampoo
soap
soda
spaghetti sauce
sugar
toilet paper
tortillas
waffles
yogurt
bagels
cereals
dinner rolls
eggs
ice cream
juice
lunch meat
paper towels
poultry
soda
toilet paper
waffles
0 50 100 150 200 250 300 350 400 450 500 550 600 650 700 750 800 850
Value


Sheet 2
Product / Date
a.. a.. b.. b.. b.. c.. c.. c.. d.. d.. e.. fl.. fr.. h.. ic.. i.. j.. k.. l.. l.. m.. m.. p.. p.. p.. p.. s.. s.. s.. s.. s.. s.. s.. t.. t.. w.. y..
Count of dataset_group.csv
300
250
200
150
100
50
0
2019
YEAR(Date)
2018
2019
2020

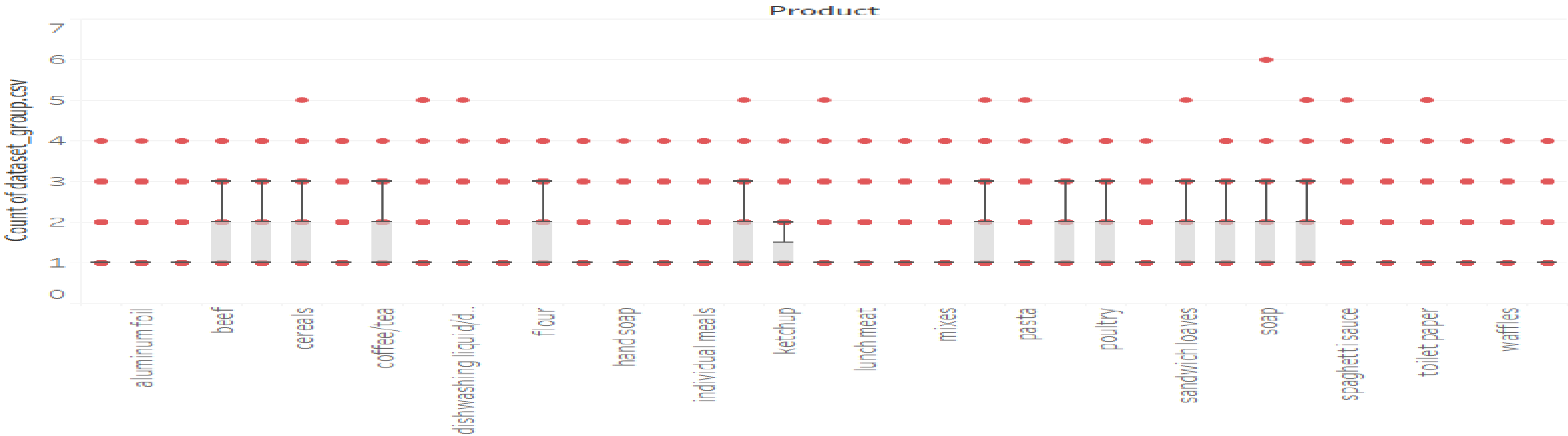


- ➢ The above plot has product calculated which is for quarterly year.
- ➢ This below mentioning sheet 4 plot is calculated on a monthly basis.
- ➢ For the year 2020, only 2 months data is recorded.
- ➢ A better understanding is given concerning the product and its count.
- ➢ The below image mentioning sheet 5 represents the product count and the order id in a box plot.

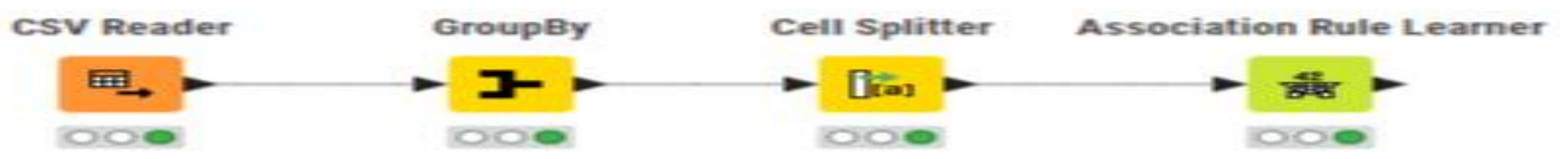
Sheet 4



Sheet 5



2) Use of Market Basket Analysis (Association Rules) -->Write Something about the association rules and its relevance in this case -->Add KNIME workflow image -->Write about threshold values of Support and Confidence.



The knime workflow :

- Market Basket Analysis (MBA) is a data mining technique used to discover associations between products or items frequently purchased together. Association rule mining is a key aspect of MBA, and it involves extracting patterns or rules from transactional data.
- Starting with the transactional data that records which products were purchased together in each transaction. Each transaction is typically represented as a set of items.
- After Identifying frequent itemsets, which are sets of items that frequently appear together in transactions. This is done using the Association Rule Learner.
- From the frequent itemsets, generated association rules that describe relationships between items. Association rules are in the form of "if {antecedent} then {consequent}" where antecedent and consequent are sets of items.

Frequent itemsets/Association rules (Table)

Rows: 18793 | Columns: 6

#	RowID	Support Number (double)	Confidence ↓ Number (double)	Lift Number (double)	Consequent String	implies String	Items Set
14336	rule1...	0.076	0.585	1.388	poultry	<---	[sandwich loaves,la...
17876	rule1...	0.099	0.579	1.49	dinner rolls	<---	[spaghetti sauce,po...
17878	rule1...	0.099	0.577	1.368	poultry	<---	[dinner rolls,spaghe...
16227	rule1...	0.079	0.573	1.36	poultry	<---	[mixes,sugar]
17680	rule1...	0.087	0.566	1.342	poultry	<---	[lunch meat,mixes]
17704	rule1...	0.087	0.566	1.342	poultry	<---	[dinner rolls,hand s...
17728	rule1...	0.088	0.565	1.341	poultry	<---	[dinner rolls,all- pur...
17739	rule1...	0.088	0.565	1.341	poultry	<---	[beef,sugar]
16862	rule1...	0.081	0.564	1.339	poultry	<---	[juice,sugar]
17781	rule1...	0.089	0.564	1.339	poultry	<---	[dinner rolls,juice]
17133	rule1...	0.083	0.563	1.498	individual meals	<---	[sandwich loaves,lu...
17702	rule1...	0.087	0.562	1.446	dinner rolls	<---	[poultry,hand soap]
17845	rule1...	0.091	0.562	1.334	poultry	<---	[dinner rolls,lunch ...
17746	rule1...	0.088	0.562	1.333	poultry	<---	[dinner rolls,milk]
17601	rule1...	0.085	0.561	1.33	poultry	<---	[toilet paper,sugar]
15814	rule1...	0.078	0.56	1.486	juice	<---	[shampoo,spaghetti...
17692	rule1...	0.087	0.559	1.435	eggs	<---	[beef,soda]
17784	rule1...	0.089	0.558	1.324	poultry	<---	[dinner rolls,coffee/...
17542	rule1...	0.085	0.557	1.323	poultry	<---	[yogurt,sandwich lo...
17839	rule1...	0.09	0.557	1.321	poultry	<---	[dinner rolls,mixes]
17874	rule1...	0.096	0.556	1.32	poultry	<---	[juice,aluminum foil]
17323	rule1...	0.083	0.556	1.432	dishwashing liquid/...	<---	[mixes,soda]
17724	rule1...	0.088	0.556	1.422	cheeses	<---	[cereals,sandwich b...
17802	rule1...	0.09	0.554	1.315	poultry	<---	[lunch meat,sugar]
17613	rule1...	0.085	0.554	1.315	poultry	<---	[dishwashing liquid...

- The above picture is the output after performing the Association Rule on the dataset .

- The threshold values of support and confidence are essential parameters in association rule mining, including Market Basket Analysis. These thresholds help determine which association rules are considered meaningful and actionable based on the characteristics of the dataset and the goals of the analysis.
- Support measures the frequency of occurrence of an itemset in the dataset. It indicates the proportion of transactions that contain the itemset. A high support value indicates that the itemset is frequently bought together.
- Setting a higher support threshold results in fewer but more reliable association rules.
- Rules with a support value below this threshold are considered insignificant and are often filtered out.
- Setting a lower support threshold results in more association rules but may include noise or spurious correlations.
- Confidence measures the reliability or strength of the association between items in a rule. It indicates the conditional probability of the consequent given the antecedent.
- Rules with a confidence value above this threshold are considered significant. Setting a higher confidence threshold ensures that only rules with a strong association between items are included.
- Rules with a confidence value below this threshold are considered weak and may not provide actionable insights. Setting a lower confidence threshold allows for the inclusion of more rules but may result in less reliable recommendations.

3) Suggestion of Possible Combos with Lucrative Offers --> Write recommendations --> Make discount offers or combos (or buy two get one free) based on the associations and your experience.

- The Grocery store can provide some combo offer for tea, coffee and sandwich loaves as they have good lift.
- The combination of hot dishwashing liquid / detergent and hand soap.
- A few discounts on combos can be offered for Tea, Coffee and Sandwich loaves with 10% discount on breakfast combo.
- Spaghetti sauce with tortillas with 5% discount on snack combo.