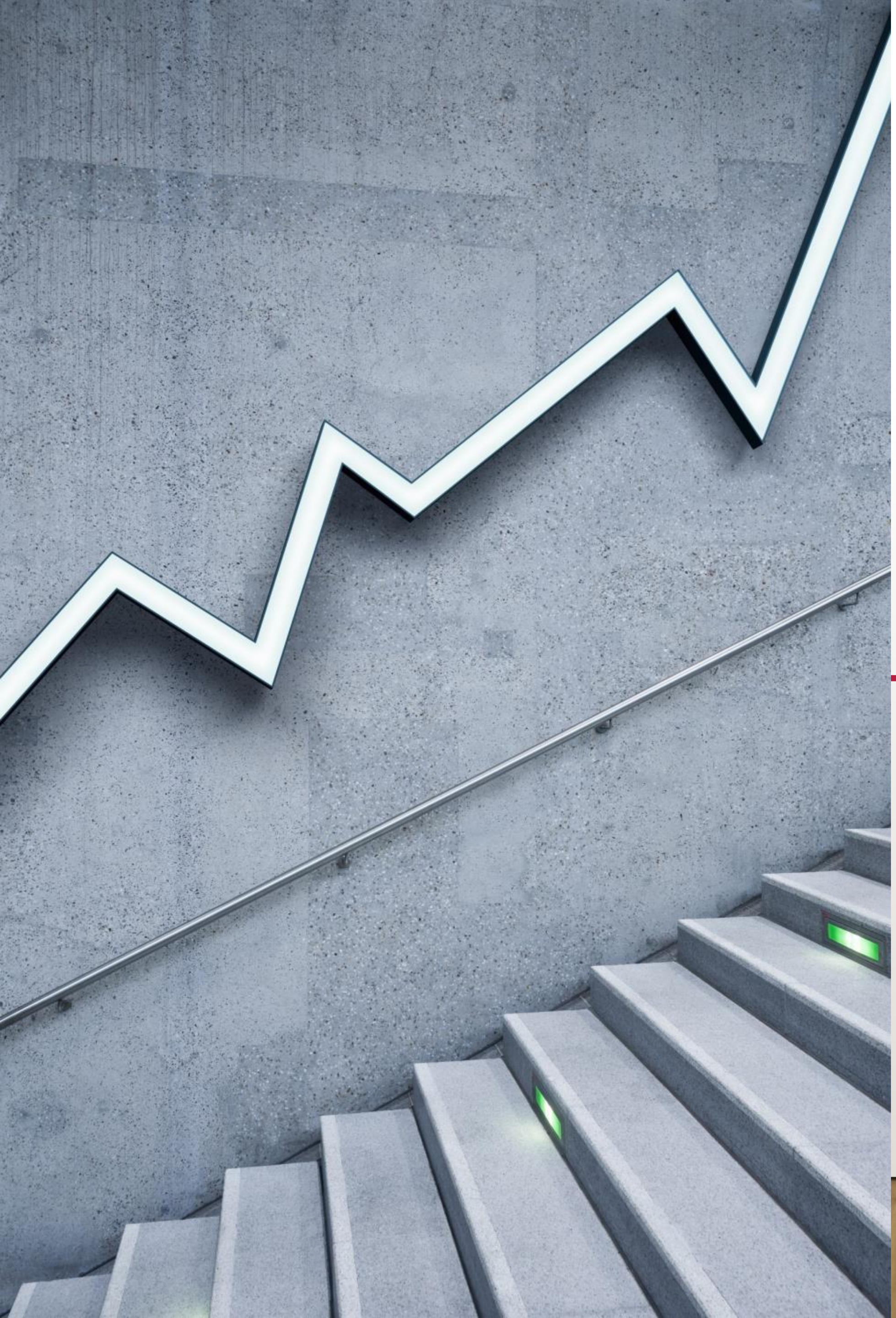


MARKETING AND RETAIL ANALYTICS PROJECT

- MILESTONE 1

- PRAKASH V MAHADOLE

Content

Content
Agenda & Executive Summary of the data
Exploratory Analysis and Inferences
Customer Segmentation using RFM analysis
Univariate, Bivariate, and multivariate analysis using data visualization
Sales Across different Categories of different features in the given data
Output table head
Inferences from RFM Analysis and identified segments

➢ PROBLEM STATEMENT -

An automobile parts manufacturing company has collected data of 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 magical data science skills to provide them with suitable insights about their data and their customers.

➢ FACT CHECK ABOUT GIVEN 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   datetime64[ns]
 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: datetime64[ns](1), float64(2), int64(5), object(12)
```

```
Out[8]: 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
```

➢ FACT CHECK ABOUT GIVEN DATA -

- Data contains 2747 rows and 20 columns.
- Data don't contain any null value.
- it contains datatype :

datetime64 (1)

Float64 (2)

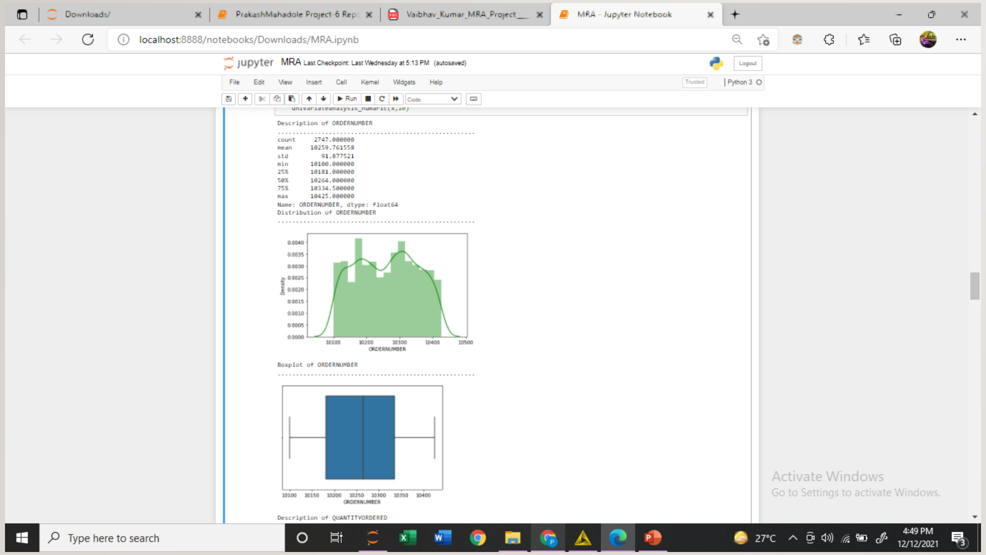
int64 (5)

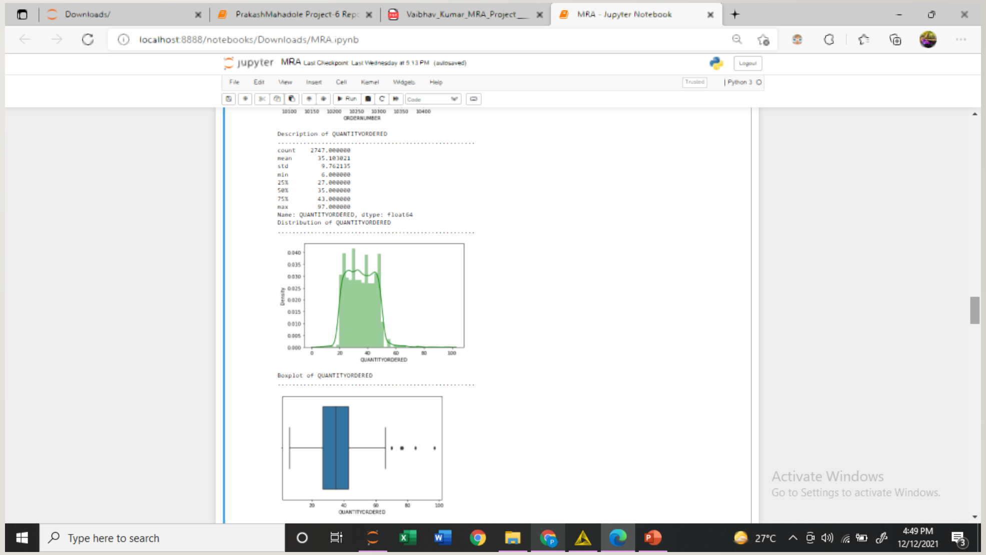
Object (12)

- Data description is given and data looks good

	ORDERNUMBER	QUANTITYORDERED	PRICEEACH	ORDERLINENUMBER	SALES	DAYS_SINCE_LASTORDER	MSRP
count	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	1757.085912	100.691664
std	91.877521	9.762135	42.042548	4.230544	1838.953901	819.280576	40.114802
min	10100.000000	6.000000	26.880000	1.000000	482.130000	42.000000	33.000000
25%	10181.000000	27.000000	68.745000	3.000000	2204.350000	1077.000000	68.000000
50%	10264.000000	35.000000	95.550000	6.000000	3184.800000	1761.000000	99.000000
75%	10334.500000	43.000000	127.100000	9.000000	4503.095000	2436.500000	124.000000
max	10425.000000	97.000000	252.870000	18.000000	14082.800000	3562.000000	214.000000

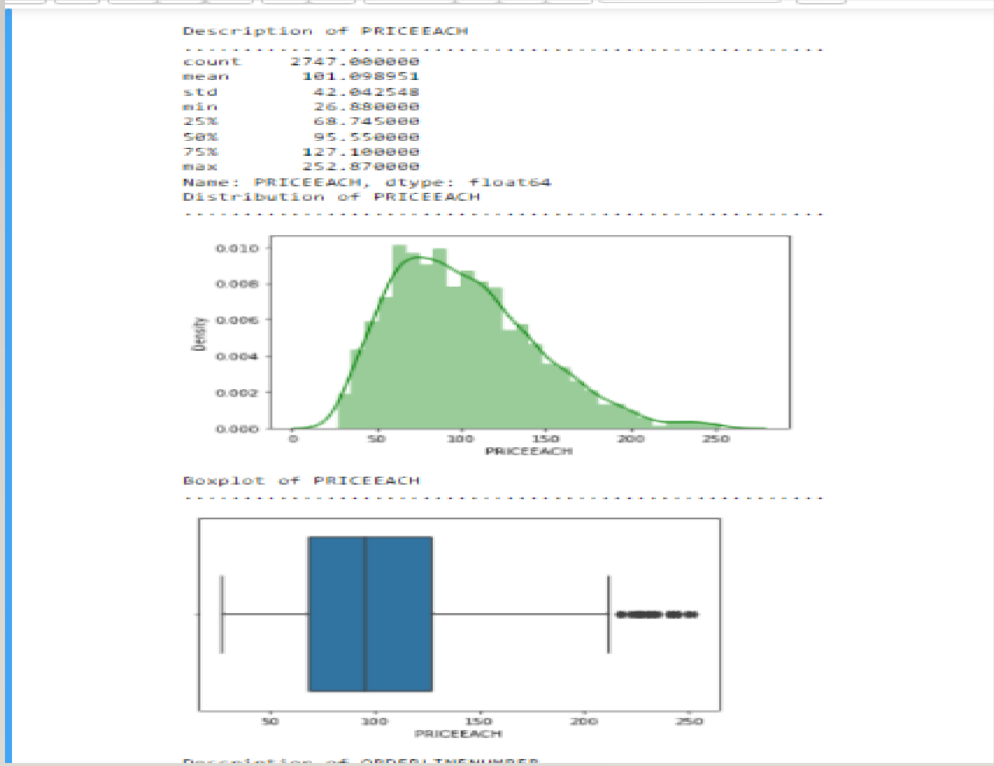
EXPLORATORY ANALYSIS –
FOR VARIABLE ORDERNUMBER AND QUANTITYORDERED

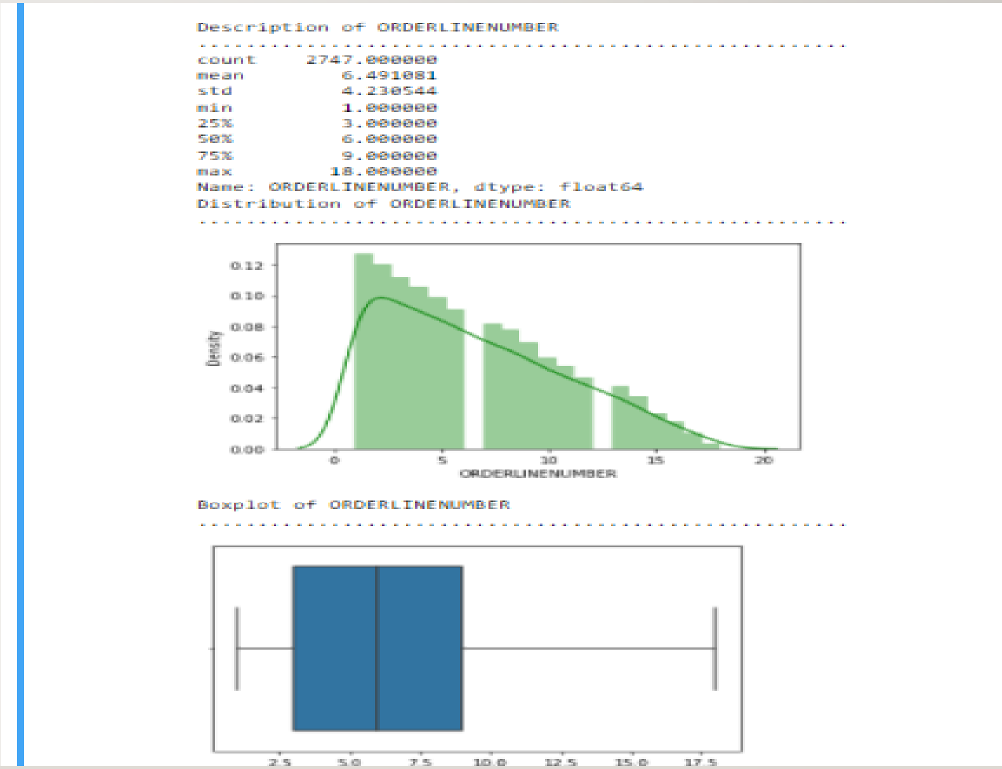




➢ EXPLORATORY ANALYSIS –

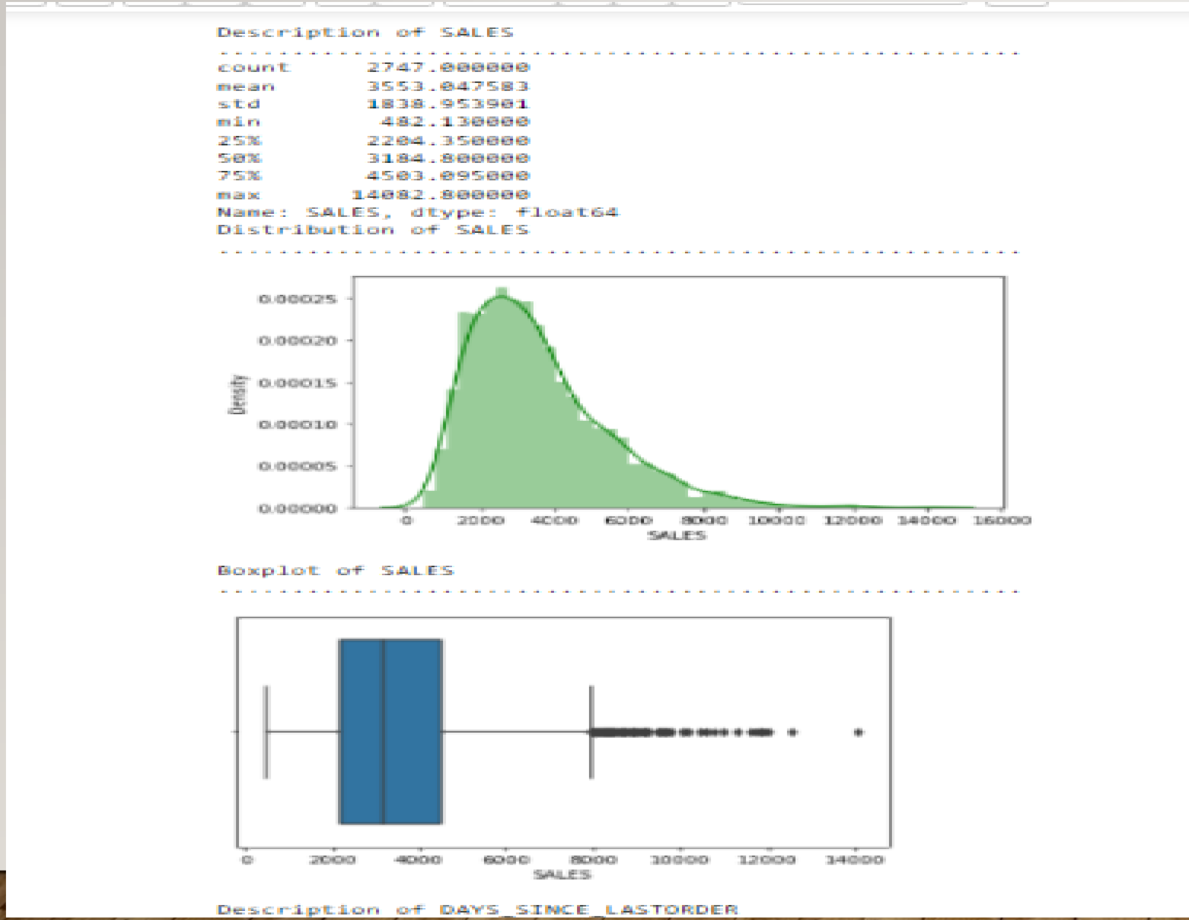
FOR VARIABLE PRICEEACH AND DERLINENUMBER

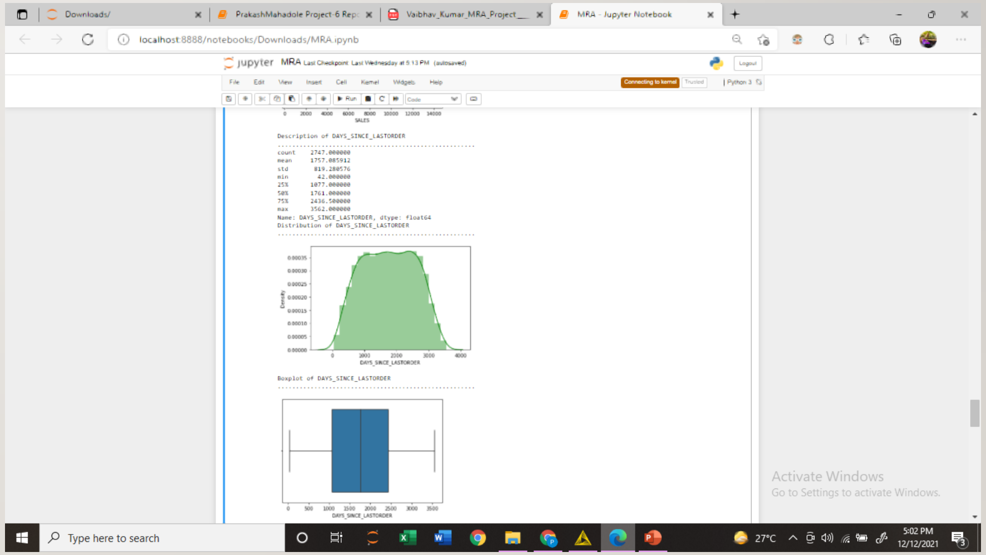


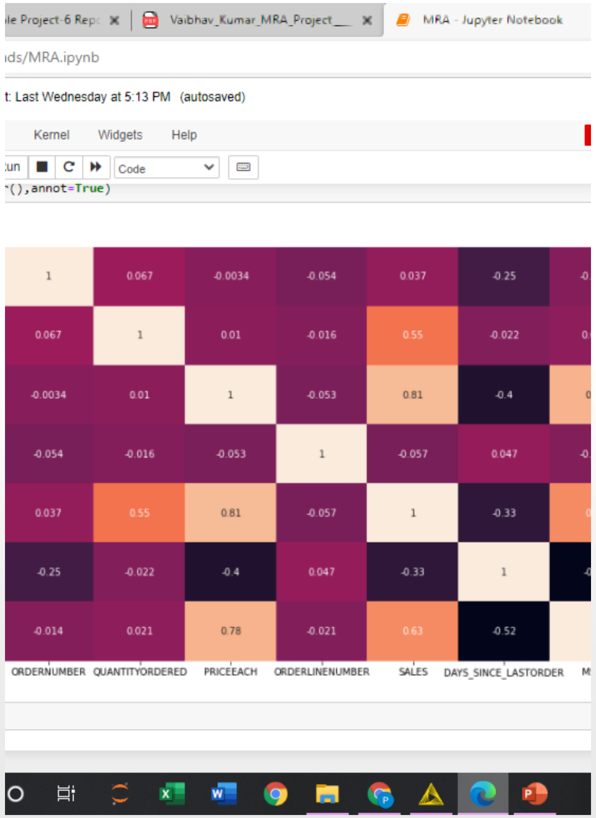


➢ EXPLORATORY ANALYSIS –

FOR VARIABLE SALES AND DAYS _SINCE_LAST_ORDERED







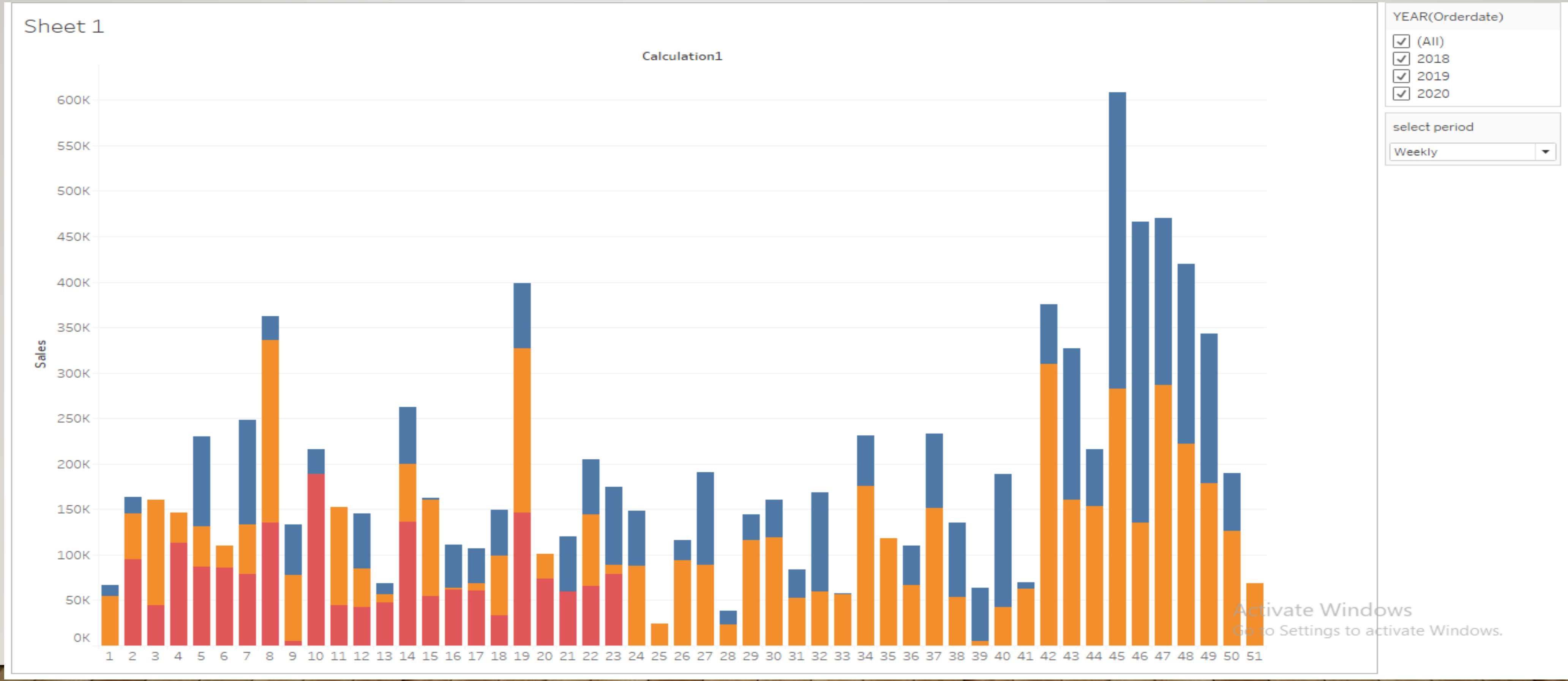
HEATMAP

The heatmap shows the correlation between two variables ,

SALES and PRICEEACH shows the high correlation .

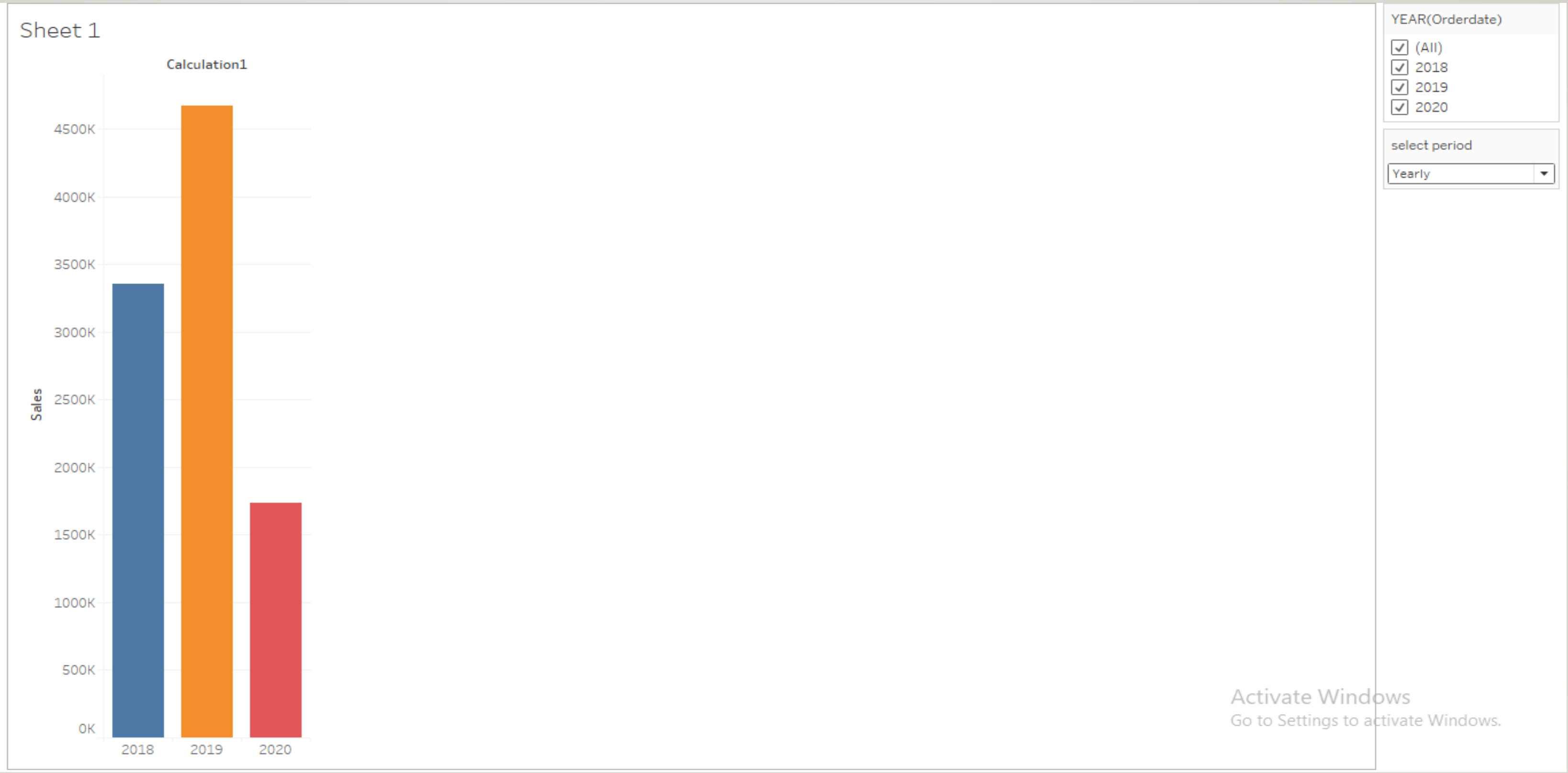
Also variable MSRP and PRICEEACH have a high correlation shown in above heatmap

➢ WEEKLY TREND IN SALES VARIABLE



➢ MONTHLY TREND IN SALES VARIABLE

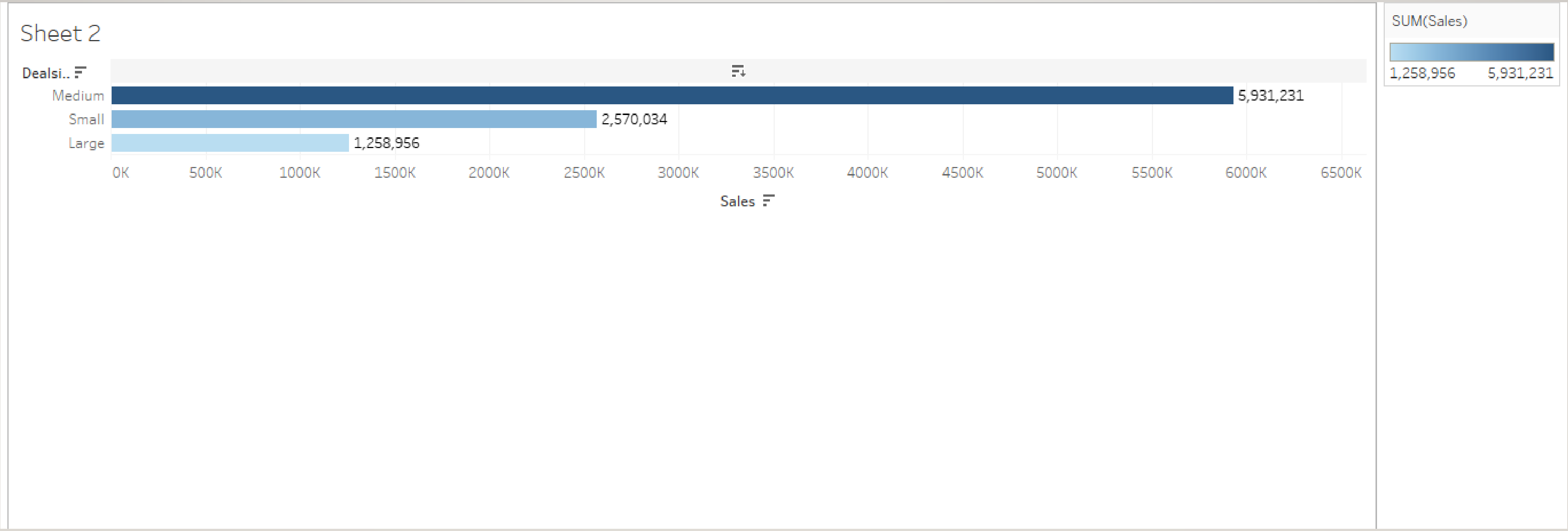
➢ YEARLY TREND IN SALES VARIABLE



➢ BIVARIATE ANALYSIS

SALES ACROSS DIFFERENT CATEGORIES OF DIFFERENT FEATURES IN THE GIVEN DATA

SALES VS DEALSIZE :

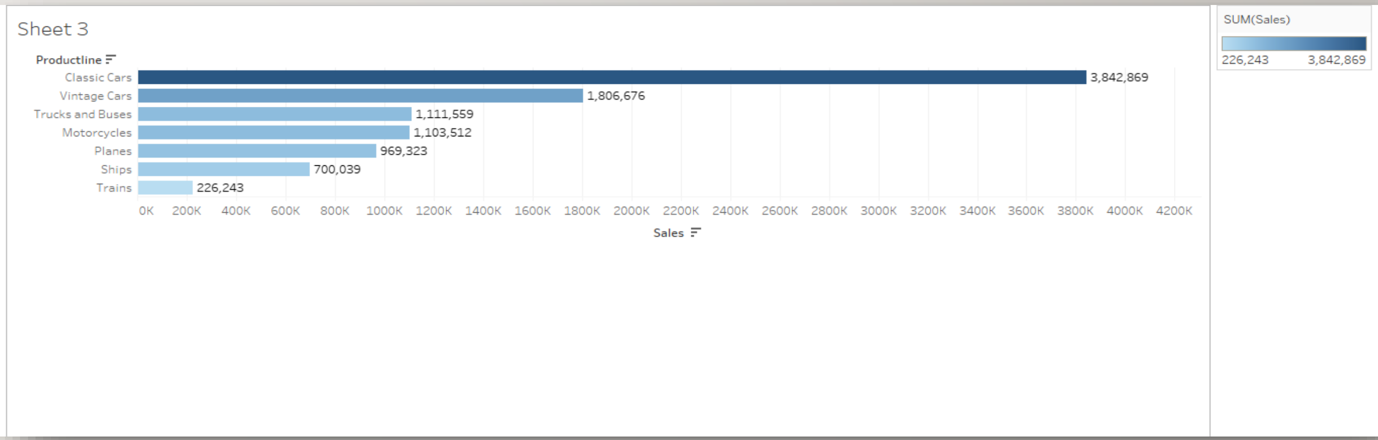


➢ BIVARIATE ANALYSIS

SALES ACROSS DIFFERENT CATEGORIES OF DIFFERENT FEATURES IN THE GIVEN DATA
SALES VS PRODUCTLINE :

➢ INFERENCE FROM THE ABOVE ANALYSIS:

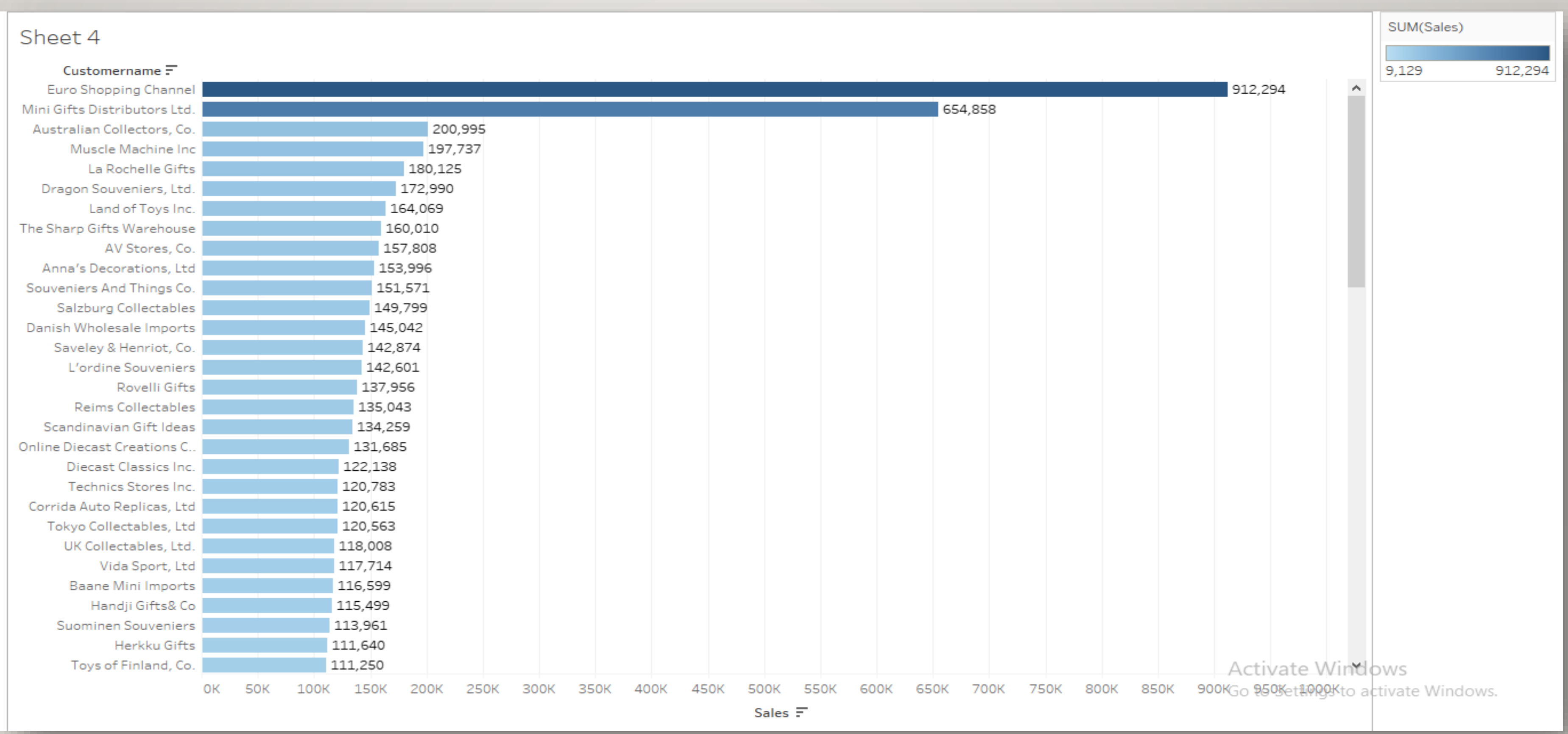
- The sales of large size deal is almost remain stagnant over the years and it can be presumed that company should focus on getting large size chunk projects .
- Maximum number of deals are Medium type of dealsize.
- Deals having larger dealsize across sales are minimum as compare to medium and smaller dealsize.
- In sales across productline 'Classic Cars' have higher number of sales as compare to other cars.
- The sales of product line like Trucks/Busses and motorcycles is in saturation stage and there is very less scope in them.



BIVARIATE ANALYSIS

SALES ACROSS DIFFERENT CATEGORIES OF DIFFERENT FEATURES IN THE GIVEN DATA

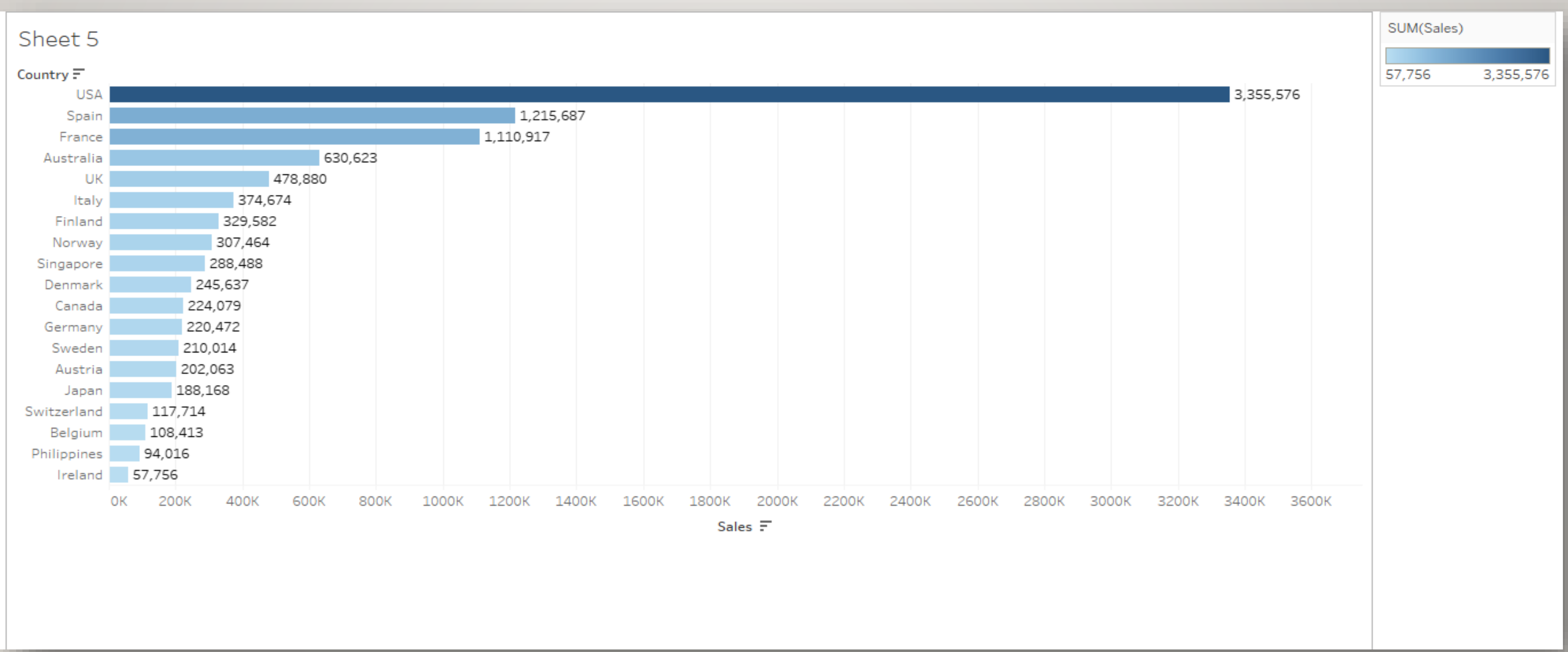
SALES VS CUSTOMERNAME :



➢ BIVARIATE ANALYSIS

SALES ACROSS DIFFERENT CATEGORIES OF DIFFERENT FEATURES IN THE GIVEN DATA

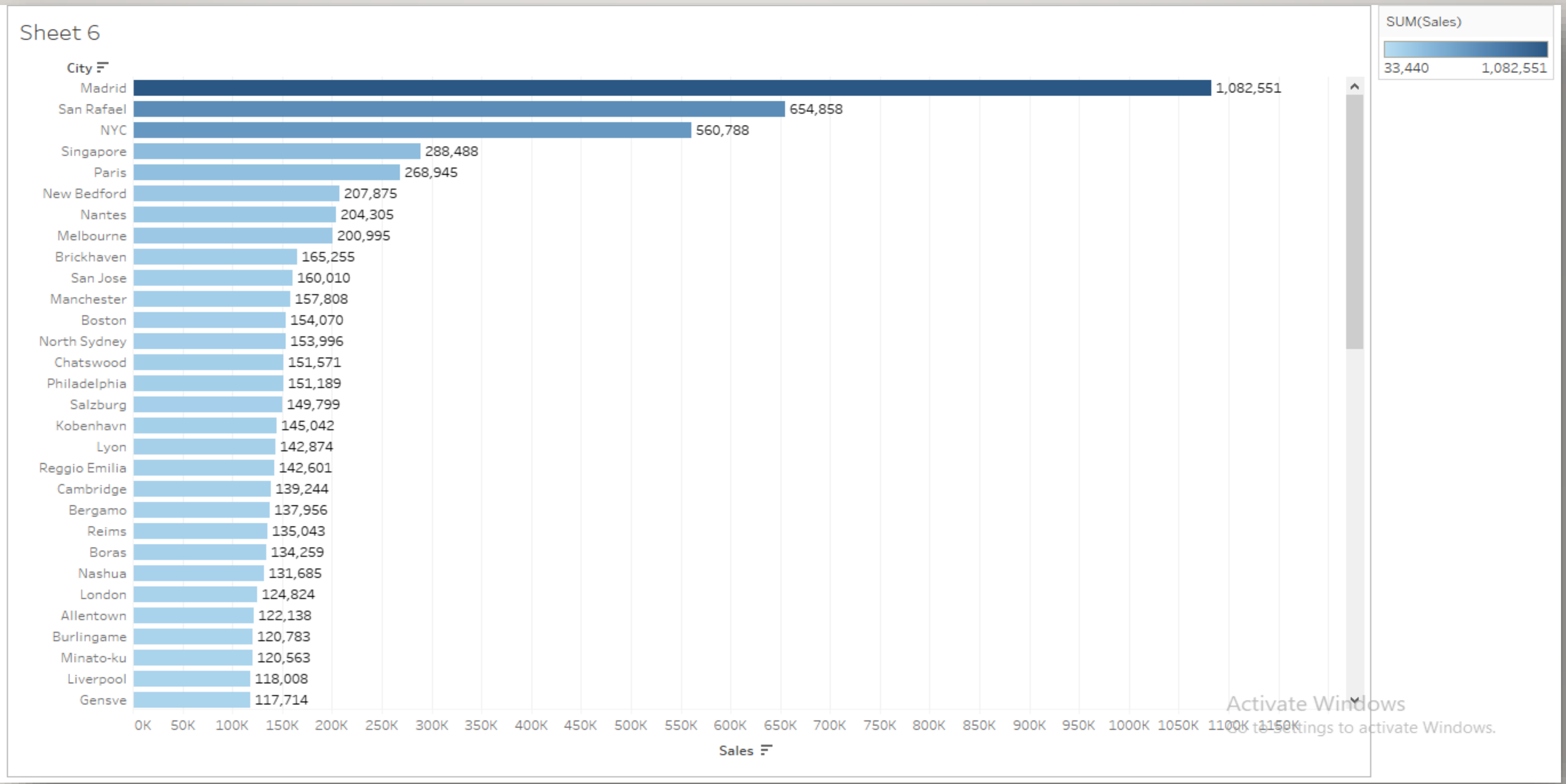
SALES VS COUNTRY :



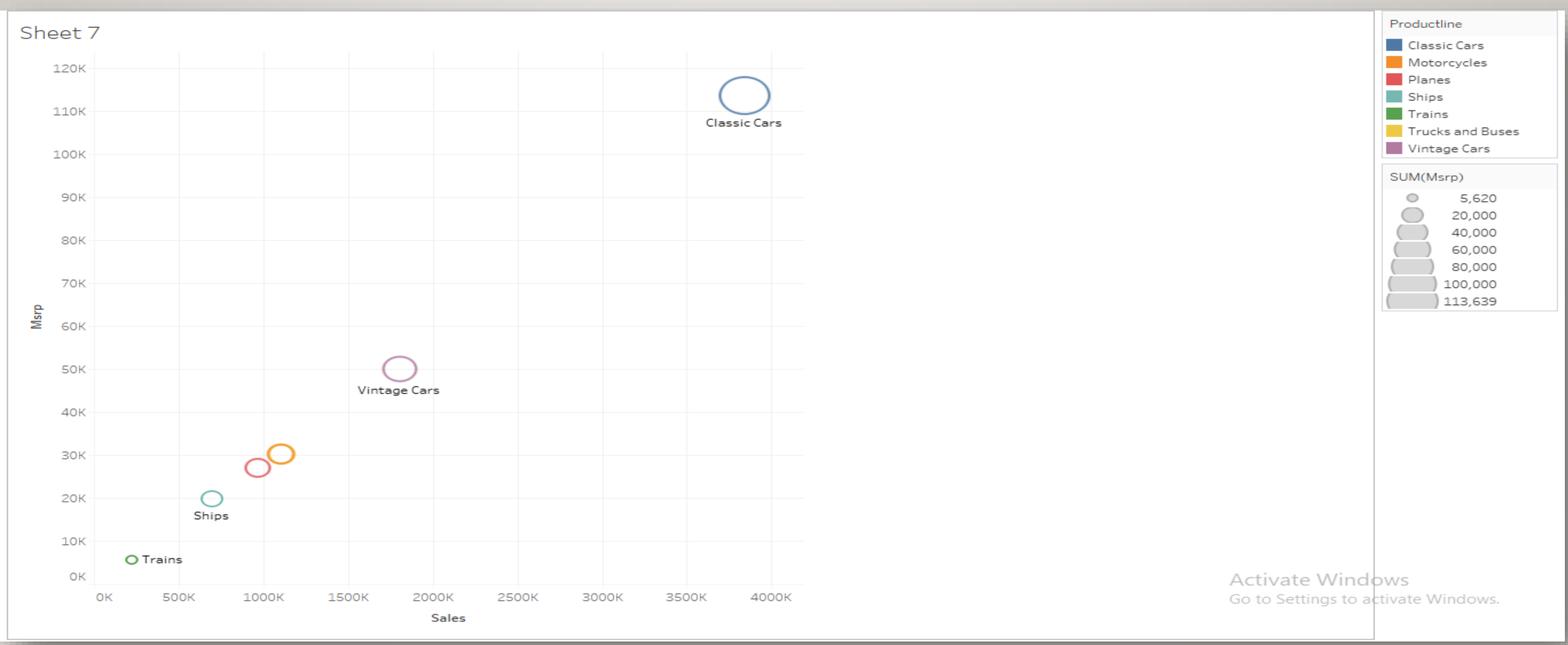
BIVARIATE ANALYSIS

SALES ACROSS DIFFERENT CATEGORIES OF DIFFERENT FEATURES IN THE GIVEN DATA

SALES VS CITY

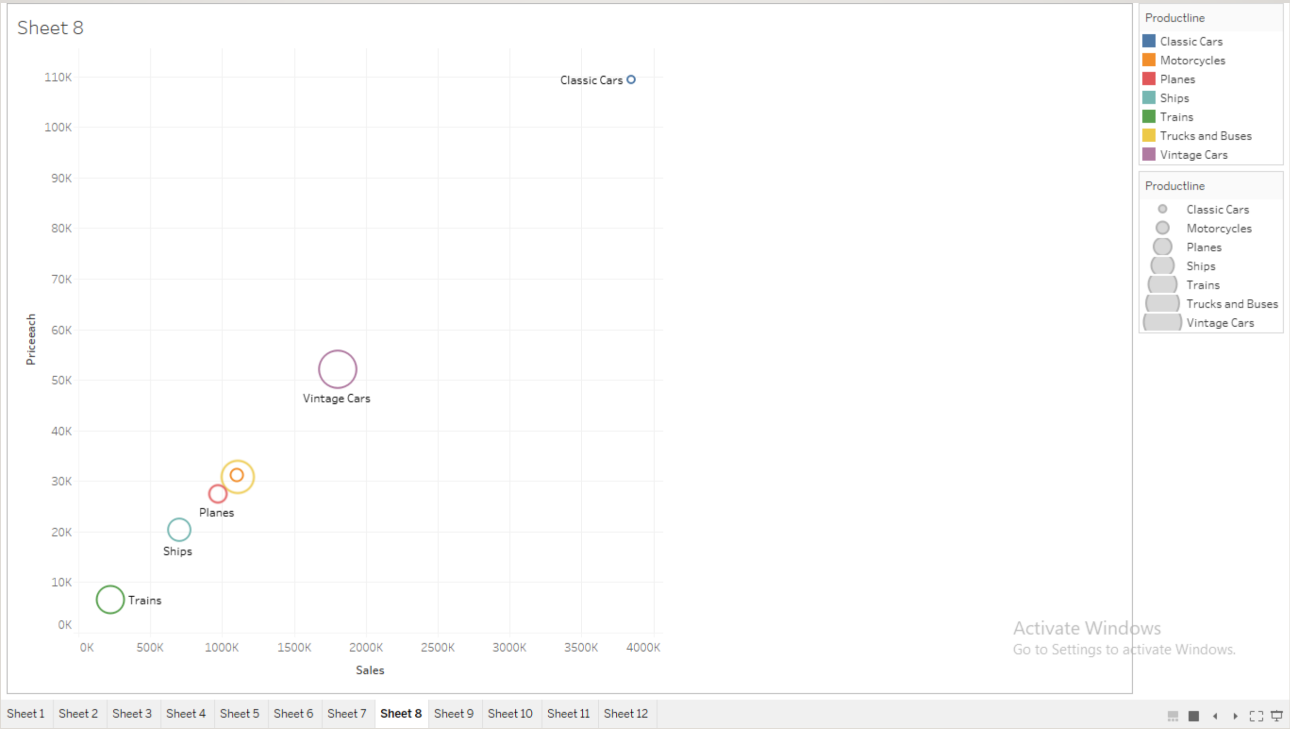


➢ MULTIVARIATE ANALYSIS
NUMARICAL VS NUMARICAL
SALES VS MSRP :

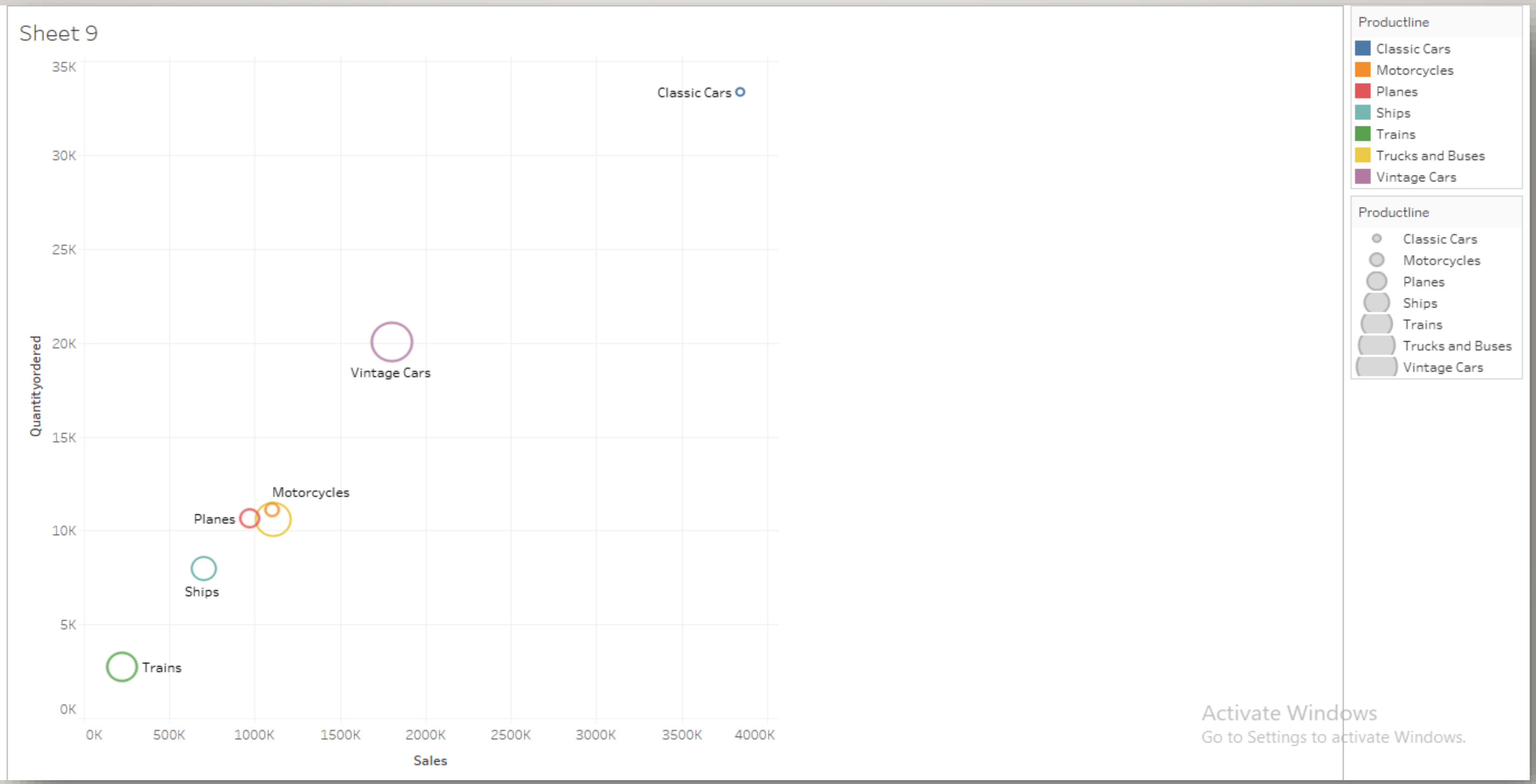


➢ MULTIVARIATE ANALYSIS

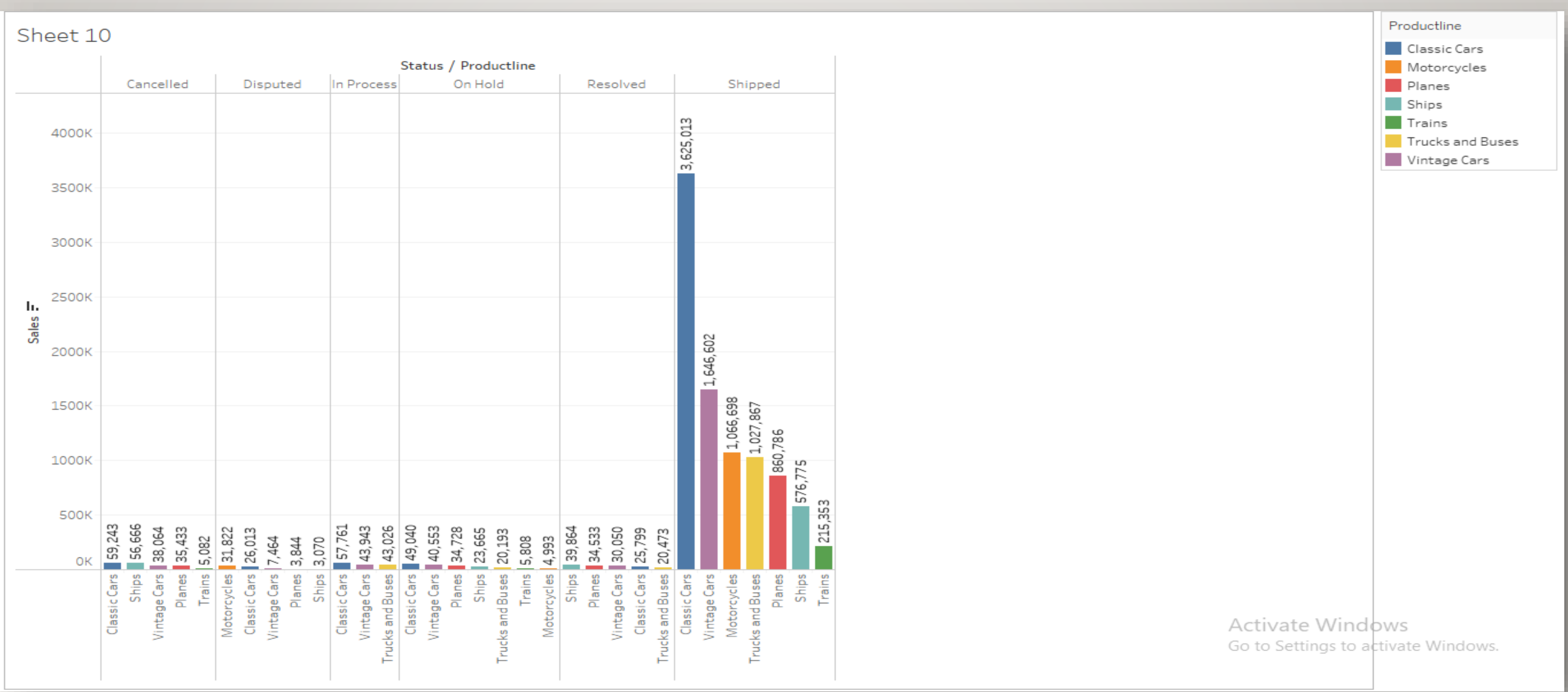
NUMARICAL VS NUMARICAL
SALES VS PRICEEACH :



➢ MULTIVARIATE ANALYSIS
NUMARICAL VS NUMARICAL
SALES VS QUANTITYORDERED :

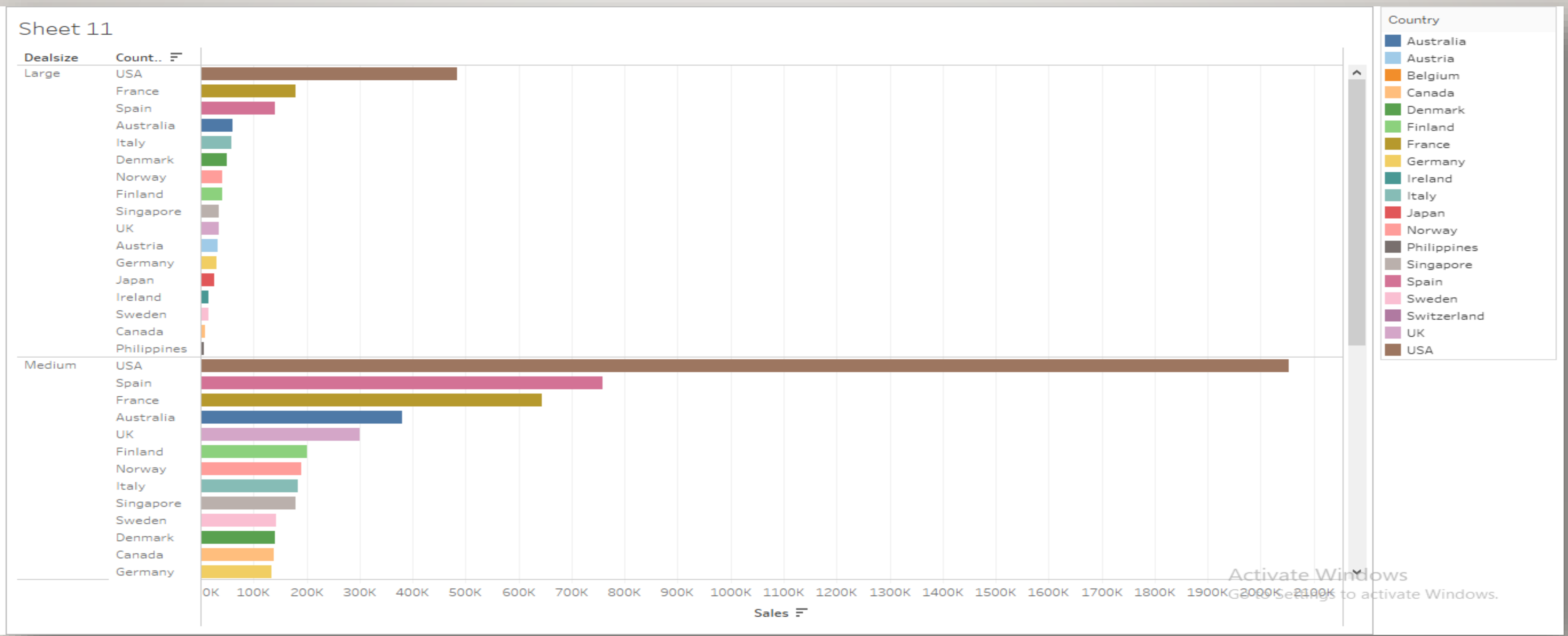


➤ MULTIVARIATE ANALYSIS
SALES ACROSS STATUS AND PRODUCTLINE :



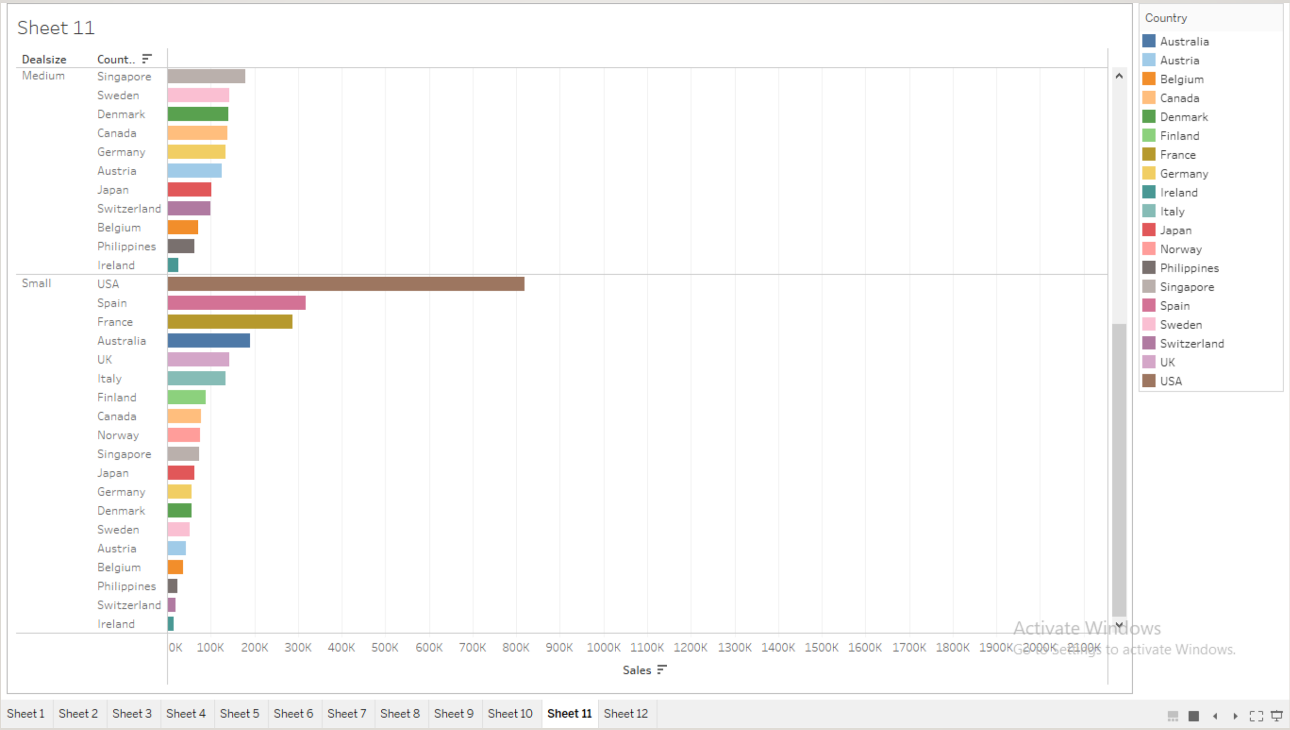
➢ MULTIVARIATE ANALYSIS

SALES ACROSS DEALSIZE ACCORDING TO COUNTRY :



MULTIVARIATE ANALYSIS

SALES ACROSS DEALSIZE ACCORDING TO COUNTRY :



➢ INFERENCE FROM THE ABOVE ANALYSIS:

- The sales of large size deal is almost remain stagnant over the years and it can be presumed that company should focus on getting large size chunk projects .
- Maximum number of deals are Medium type of dealsize.
- Deals having larger dealsize across sales are minimum as compare to medium and smaller dealsize.
- In sales across productline 'Classic Cars' have higher number of sales as compare to other cars.
- The sales of product line like Trucks/Busses and motorcycles is in saturation stage and there is very less scope in them.

➢ INFERENCE FROM THE ABOVE ANALYSIS:

- The company is customer driven because there major chunk of sales comes from 4 – 5 customer. So therefore company should focus more on customer scouting in a rationale way because in case there is client churn's it will impact the sales of the company grossly.
- Company have higher amount of Sales in country USA as compare to other country.
- Company drives export revenue mainly from large size deals followed by medium size deals and small size deals therefore it can be inferenced that large size deals is mainly from foreign customers and company should focus on domestic sales as well

➢ CUSTOMER SEGMENTATION USING RFM ANALYSIS

- The tool used for Customer segmentation using RFM analysis is knime .
- We have used country column for recency analysis and chosen country as USA
- We have created 4 segments

1. Low
2. Medium
3. High
4. Extremely high

➢ CUSTOMER SEGMENTATION USING RFM ANALYSIS

➢ RFM Analysis using knime.

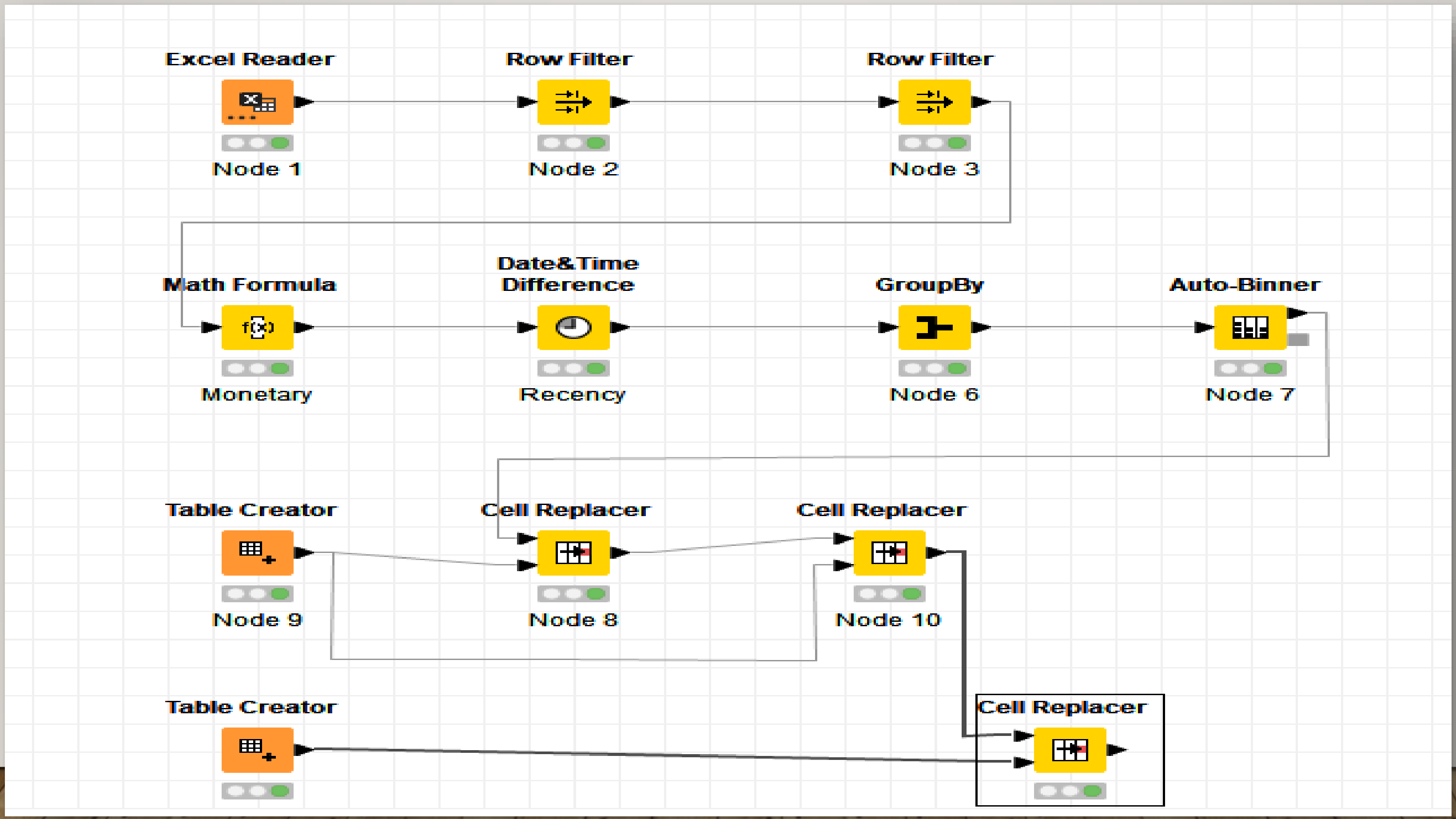
- We have divided the customers into four bins i.e. Bin-1, Bin-2, Bin-3, Bin-4 and mapped them to recency, frequency and monetary columns.
- After performing RFM analysis, it was found that Recency consists of top 25% customer which can be termed as active customers.
- Bin-3 consist of customers which are at risk.
- After that the bottom consists of inactive customers which are in Bin-4 and in lowest quartile.

➢ RFM ANALYSIS:

OUTPUT TABLE HEAD :

Row ID	PRODU...	PHONE	ADDRESSLI...	CITY	POSTA...	COUNTRY	CONTA...	CONTA...	DEALSIZE	Monetary	Recency	ORDER...	Moneta...	Recenc...	Moneta...	Freque...	Recen
Row0	3029	6175558428	16780 Pompton St.	Brickhaven	58339	USA	Taylor	Leslie	Medium	26,479.26	739	Bin 1	Bin 1	Bin 1	L	L	X
Row1	3380	3105552373	4097 Douglas Av.	Glendale	92561	USA	Young	Leslie	Medium	9,129.35	672	Bin 1	Bin 1	Bin 1	L	L	X
Row2	1949	6175555555	4658 Baden Av.	Cambridge	51247	USA	Tseng	Kyung	Medium	36,163.62	948	Bin 1	Bin 1	Bin 2	L	L	H
Row3	1949	2155554695	782 First Street	Philadelphia	71270	USA	Cervantes	Francisca	Medium	67,506.97	789	Bin 1	Bin 1	Bin 2	L	L	H
Row4	1949	2125558493	5905 Pompton St.	NYC	10022	USA	Hernandez	Maria	Medium	77,795.2	751	Bin 1	Bin 1	Bin 1	L	L	X
Row5	4757	7605558146	361 Furth Circle	San Diego	91217	USA	Thompson	Valarie	Medium	87,489.23	1019	Bin 1	Bin 1	Bin 3	L	L	M
Row6	1949	6175558555	7825 Douglas Av.	Brickhaven	58339	USA	Nelson	Allen	Medium	81,577.98	691	Bin 1	Bin 1	Bin 1	L	L	X
Row7	1678	2155551555	7586 Pompton St.	Allentown	70267	USA	Yu	Kyung	Medium	122,138.14	560	Bin 1	Bin 1	Bin 1	L	L	X
Row8	4962	6175552555	6251 Ingle Ln.	Boston	51003	USA	Franco	Valarie	Medium	70,859.78	960	Bin 1	Bin 1	Bin 2	L	L	H
Row9	1678	5085552555	1785 First Street	New Bedford	50553	USA	Benitez	Violeta	Medium	98,923.73	648	Bin 1	Bin 1	Bin 1	L	L	X
Row10	1678	2035552570	25593 South Bay...	Bridgewater	97562	USA	King	Julie	Medium	101,894.79	585	Bin 1	Bin 1	Bin 1	L	L	X
Row11	1662	2035554407	2440 Pompton St.	Glendale	97561	USA	Lewis	Dan	Small	57,294.42	738	Bin 1	Bin 1	Bin 1	L	L	X
Row12	4757	6175559555	8616 Spinnaker Dr.	Boston	51003	USA	Yoshido	Juri	Medium	83,209.88	584	Bin 1	Bin 1	Bin 1	L	L	X
Row13	1678	2125557818	897 Long Airport ...	NYC	10022	USA	Yu	Kwai	Small	164,069.44	757	Bin 2	Bin 1	Bin 1	L	M	X
Row14	1678	6175558555	39323 Spinnaker ...	Cambridge	51247	USA	Hernandez	Marta	Medium	103,080.38	790	Bin 1	Bin 1	Bin 2	L	L	H
Row15	2823	2125551957	5290 North Pend...	NYC	10022	USA	Kuo	Kee	Medium	33,144.93	769	Bin 1	Bin 1	Bin 1	L	L	X
Row16	2016	9145554562	3758 North Pend...	White Plains	24067	USA	Frick	Steve	Medium	85,555.99	788	Bin 1	Bin 1	Bin 2	L	L	H
Row17	4757	5085559555	4575 Hillside Dr.	New Bedford	50553	USA	Tam	Wing C	Medium	108,951.13	704	Bin 1	Bin 1	Bin 1	L	L	X
Row18	1949	4155551450	5677 Strong St.	San Rafael	97562	USA	Nelson	Valarie	Large	654,858.06	561	Bin 4	Bin 4	Bin 1	X	X	X
Row19	2016	2155559857	11328 Douglas Av.	Philadelphia	71270	USA	Hernandez	Rosa	Small	83,682.16	755	Bin 1	Bin 1	Bin 1	L	L	X
Row20	1108	2125557413	4092 Furth Circle	NYC	10022	USA	Young	Jeff	Large	197,736.94	741	Bin 2	Bin 2	Bin 1	M	M	X
Row21	1949	6035558647	2304 Long Airpor...	Nashua	62005	USA	Young	Valarie	Medium	131,685.3	768	Bin 1	Bin 1	Bin 1	L	L	X
Row22	1099	6175557555	7635 Spinnaker Dr.	Brickhaven	58339	USA	Barajas	Miguel	Large	57,197.96	823	Bin 1	Bin 1	Bin 2	L	L	H
Row23	4473	4155554312	2793 Furth Circle	Brisbane	94217	USA	Taylor	Sue	Medium	50,218.51	1035	Bin 1	Bin 1	Bin 3	L	L	M
Row24	1129	7025551838	8489 Strong St.	Las Vegas	83030	USA	King	Sue	Medium	82,751.08	743	Bin 1	Bin 1	Bin 1	L	L	X
Row25	1949	2035559545	567 North Pendal...	New Haven	97823	USA	Murphy	Leslie	Medium	79,472.07	952	Bin 1	Bin 1	Bin 2	L	L	H
Row26	1678	6505556809	9408 Furth Circle	Burlingame	94217	USA	Hirano	Juri	Medium	120,783.07	706	Bin 1	Bin 1	Bin 1	L	L	X
Row27	1678	2015559350	7476 Moss Rd.	Newark	94019	USA	Brown	William	Medium	83,228.19	617	Bin 1	Bin 1	Bin 1	L	L	X
Row28	4757	4085553659	3086 Ingle Ln.	San Jose	94217	USA	Frick	Sue	Large	160,010.27	598	Bin 1	Bin 1	Bin 1	L	L	X
Row29	1678	6265557265	78934 Hillside Dr.	Pasadena	90003	USA	Young	Julie	Medium	104,561.96	699	Bin 1	Bin 1	Bin 1	L	L	X
Row30	1678	2125551500	2678 Kingston Rd.	NYC	10022	USA	Frick	Michael	Small	88,041.26	767	Bin 1	Bin 1	Bin 1	L	L	X
Row31	1949	3105553722	3675 Furth Circle	Burbank	94019	USA	Thompson	Steve	Medium	46,084.64	1047	Bin 1	Bin 1	Bin 3	L	L	M

➢ WORKFLOW IMAGE OF KNIME



➢ INFERENCES FROM RFM ANALYSIS AND IDENTIFIED SEGMENTS

- Top Customers :

1. Customer "Mini Gifts Distributors" having large type dealsize with a high monetary value , high frequency and high recency .
2. We have our second best customer "Muscle Machine Inc." with medium monetary, medium frequency and high recency.
3. The third one is "Land of Toys Inc." with high recency, medium frequency .

We have found best three customers.

➢ INFERENCES FROM RFM ANALYSIS AND IDENTIFIED SEGMENTS

Lost Customers :

1. Customer "Collectable Mini Design" have a low recency, low frequency and low monetary.
2. Customer "Signal Collectibles Ltd." have a low recency, low frequency and low monetary.
3. Customer "West Coast Collectable" have a low recency, low frequency and low monetary.

- These are the lost customers having low recency , low frequency and low monetary.

➢ INFERENCES FROM RFM ANALYSIS AND IDENTIFIED SEGMENTS

- Loyal Customers:

1. Customer “Mini Gifts Distributors” having high monetary, high frequency and high recency .
2. Customer “Muscle Machine Inc.” with medium monetary, medium frequency and high recency.
3. Customer “Land of Toys Inc.” with high recency, medium frequency.

➢ INFERENCES FROM RFM ANALYSIS:

- “Champions” are those best customers, who bought most recently, most often, and are heavy spenders. They can become early adopters for new products and will help promote the brand.
- “Potential Loyalists” are most recent customers with average frequency and who spent a good amount.
- “New Customers” are those customers who have a high overall RFM score but are not frequent shoppers. Start building relationships with these customers by providing onboarding support and special offers to increase their visits.
- “At Risk” Customers are those customers who purchased often and spent big amounts, but haven’t purchased recently.
- “Can’t Lose Them” are customers who used to visit and purchase quite often, but haven’t been visiting recently.
- Conclusions - We found that most of the customers belong to Bin-z which is categorized as medium meaning that customers in this category are at risk and they might switch to other supplier as well in future if there will be price issue