Marketing & Retail Analytics

Part A - RFM

Agenda & Executive Summary of the Data

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.

Info

- The data has 20 variables
- All variables does not any null values
- 1 datetime, 5 int, 2 float and 12 object variables
- All the variables have their respective datatypes

```
<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)
memory usage: 429.3+ KB
```

Shape

```
The numner of rows in the data set =  2747
The numner of columns in the data set =  20
```

Summary - Numerical

- Quantity Ordered is normally distributed
- Price goes from 101 to 252. Right Skewed
- Sales is right skewed and have outliers

	count	mean	std	min	25%	50%	75%	max
ORDERNUMBER	2747.0	10259.76	91.88	10100.00	10181.00	10264.00	10334.50	10425.00
QUANTITYORDERED	2747.0	35.10	9.76	6.00	27.00	35.00	43.00	97.00
PRICEEACH	2747.0	101.10	42.04	26.88	68.74	95.55	127.10	252.87
ORDERLINENUMBER	2747.0	6.49	4.23	1.00	3.00	6.00	9.00	18.00
SALES	2747.0	3553.05	1838.95	482.13	2204.35	3184.80	4503.09	14082.80
DAYS_SINCE_LASTORDER	2747.0	1757.09	819.28	42.00	1077.00	1761.00	2436.50	3562.00
MSRP	2747.0	100.69	40.11	33.00	68.00	99.00	124.00	214.00

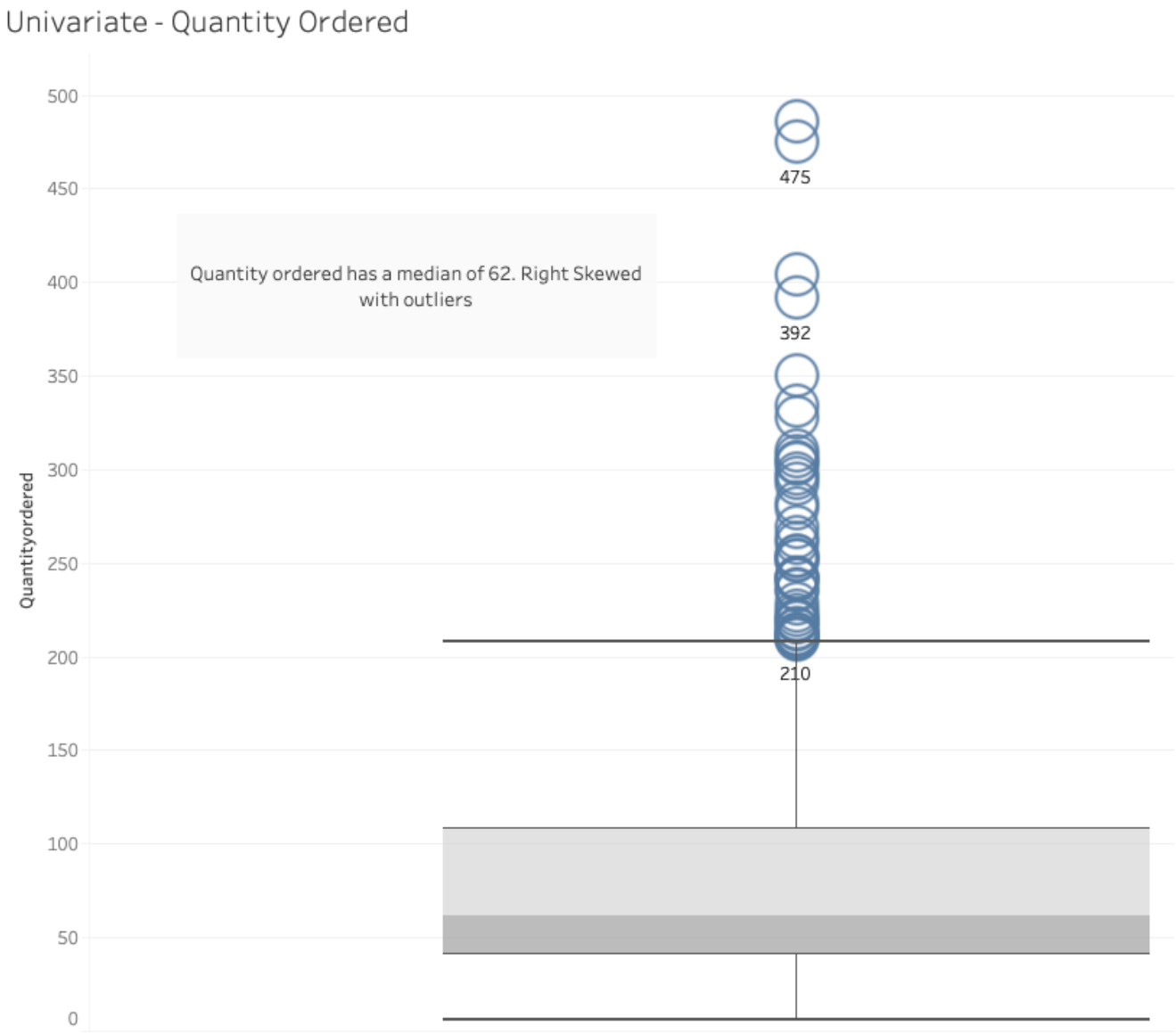
Summary - Object

- Product Code had 109 unique SKU's
- There are 89 unique customers
- 71 unique cities and 19 unique countries
- Medium has the higher frequency in the deal size with 1349 out of 2747
- Classic cars is the highest ordered productline

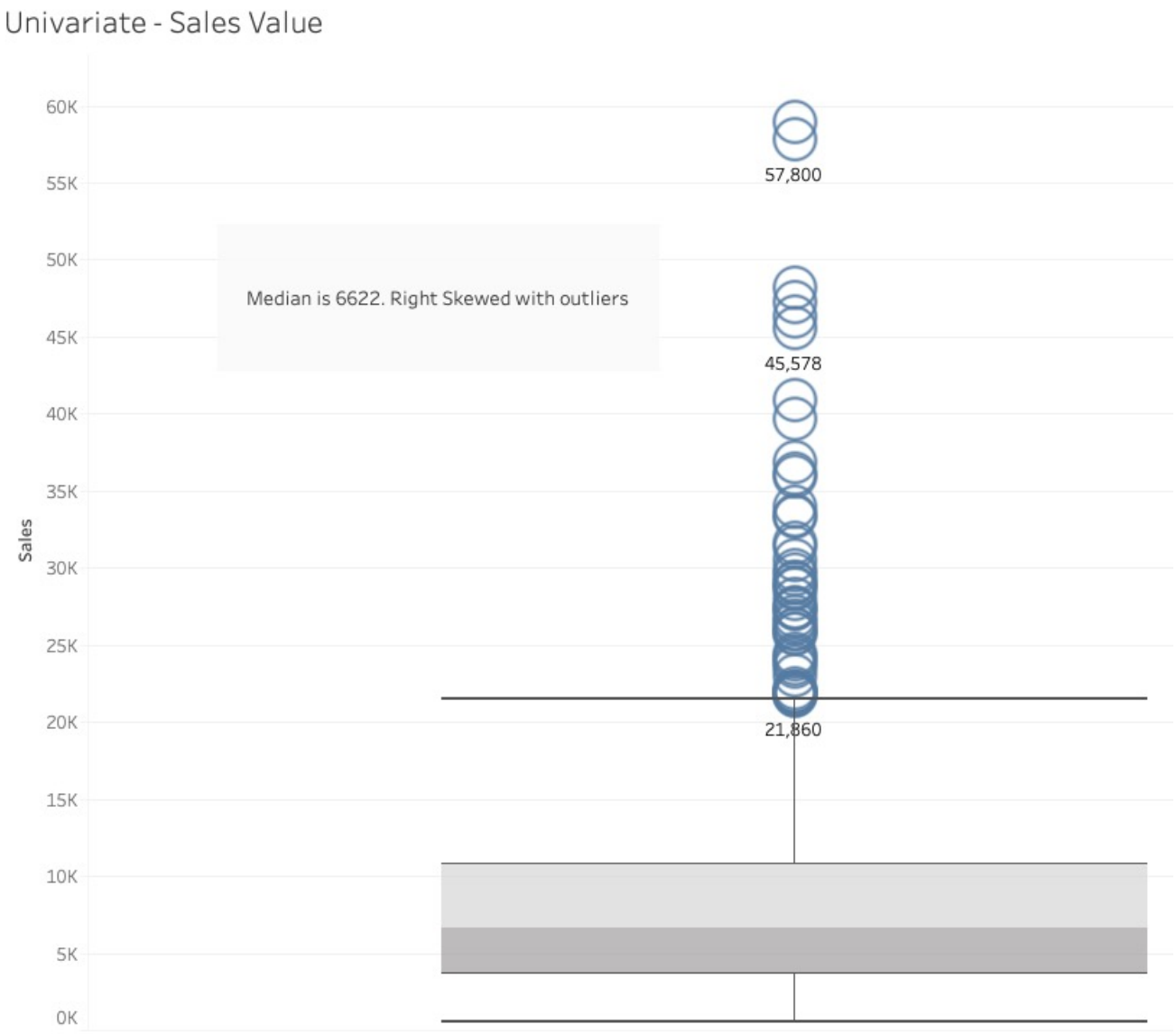
	count	unique	top	freq
STATUS	2747	6	Shipped	2541
PRODUCTLINE	2747	7	Classic Cars	949
PRODUCTCODE	2747	109	S18_3232	51
CUSTOMERNAME	2747	89	Euro Shopping Channel	259
PHONE	2747	88	(91) 555 94 44	259
ADDRESSLINE1	2747	89	C/ Moralzarzal, 86	259
CITY	2747	71	Madrid	304
POSTALCODE	2747	73	28034	259
COUNTRY	2747	19	USA	928
CONTACTLASTNAME	2747	76	Freyre	259
CONTACTFIRSTNAME	2747	72	Diego	259
DEALSIZE	2747	3	Medium	1349

Exploratory Analysis and Inferences

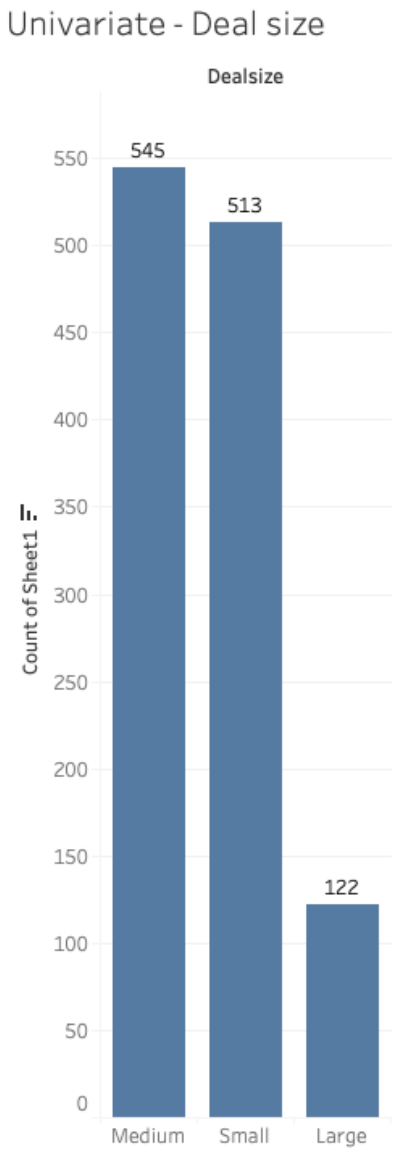
Univariate Analysis



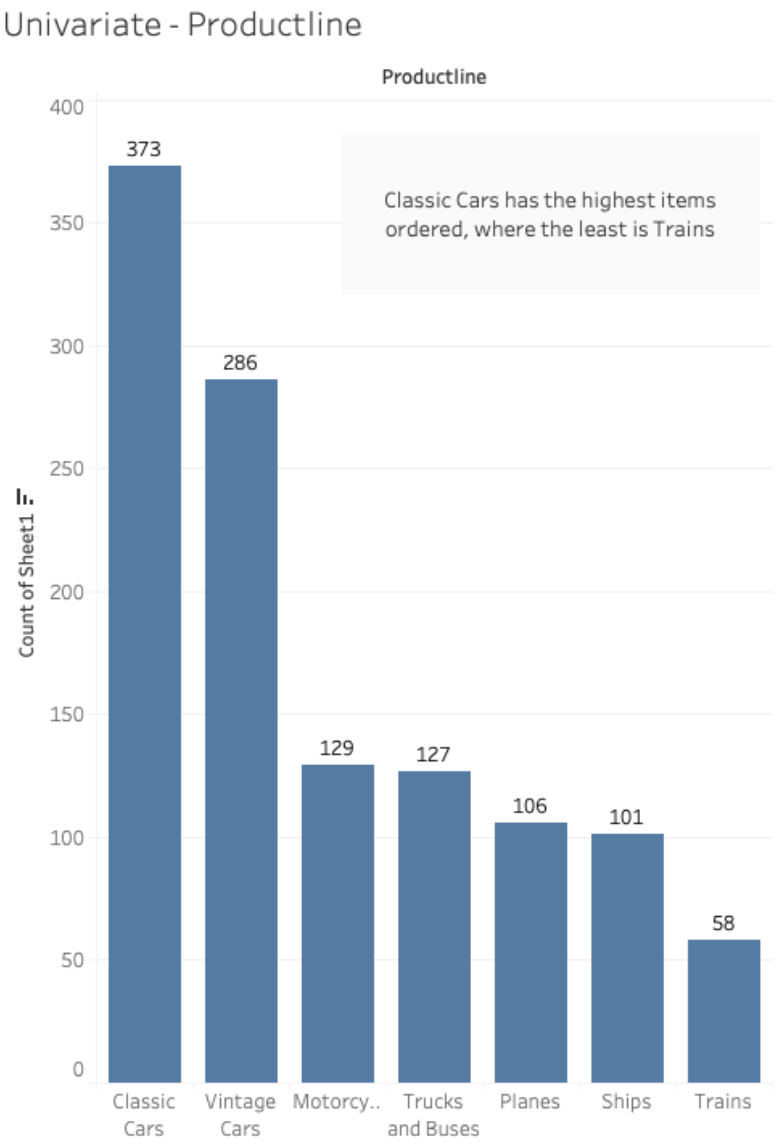
Univariate Analysis



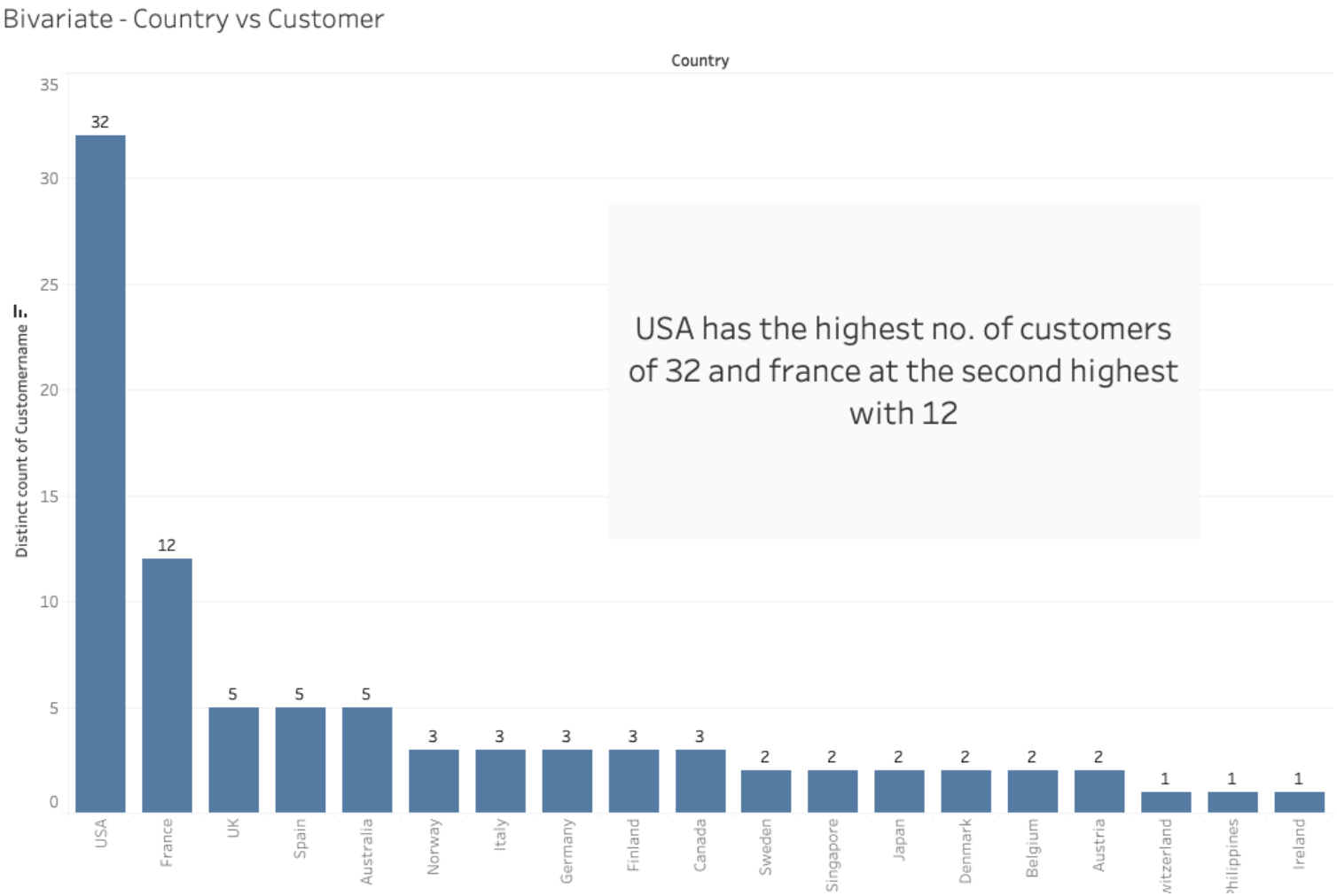
Univariate Analysis



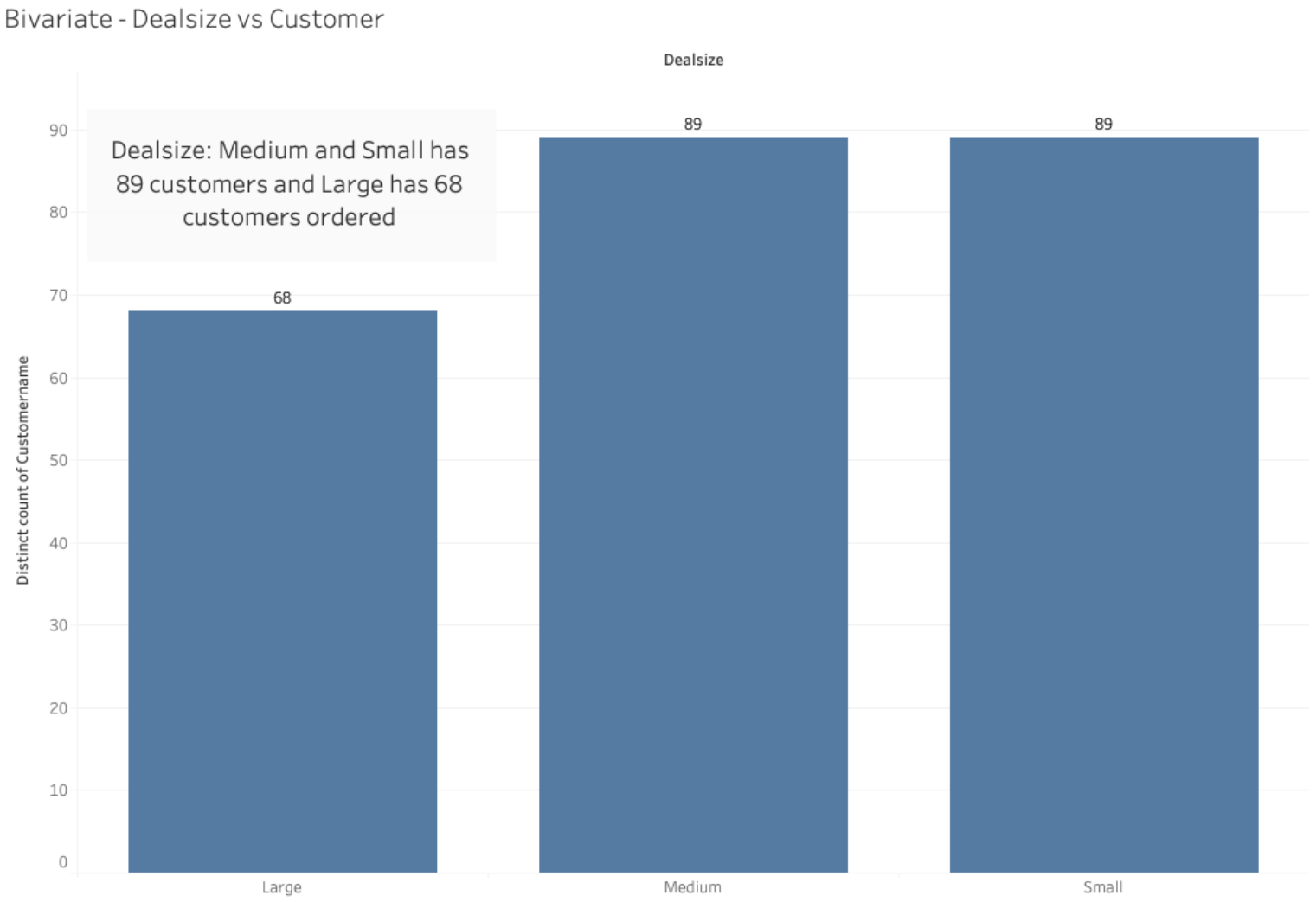
Univariate Analysis



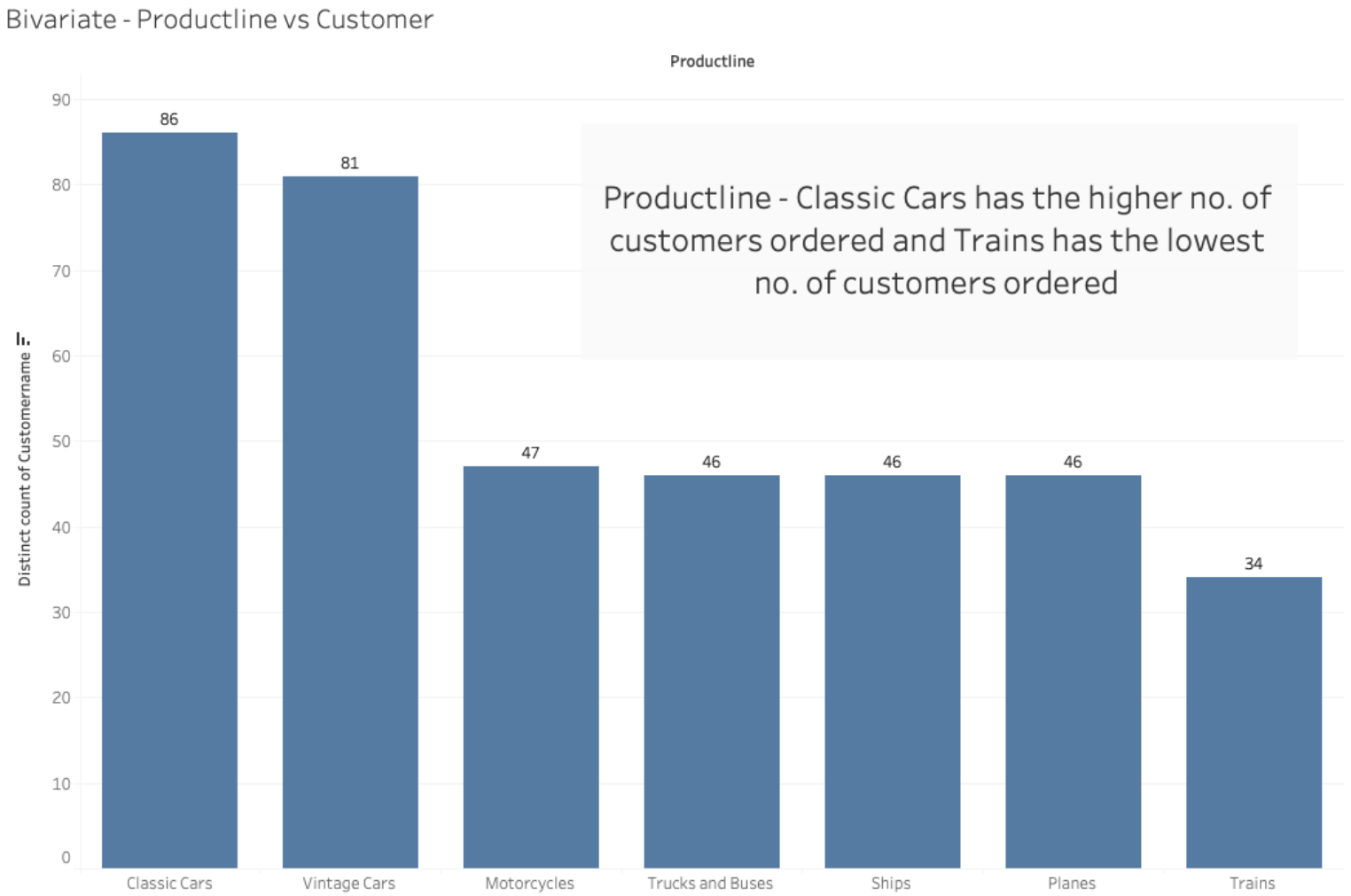
Bivariate Analysis



Bivariate Analysis

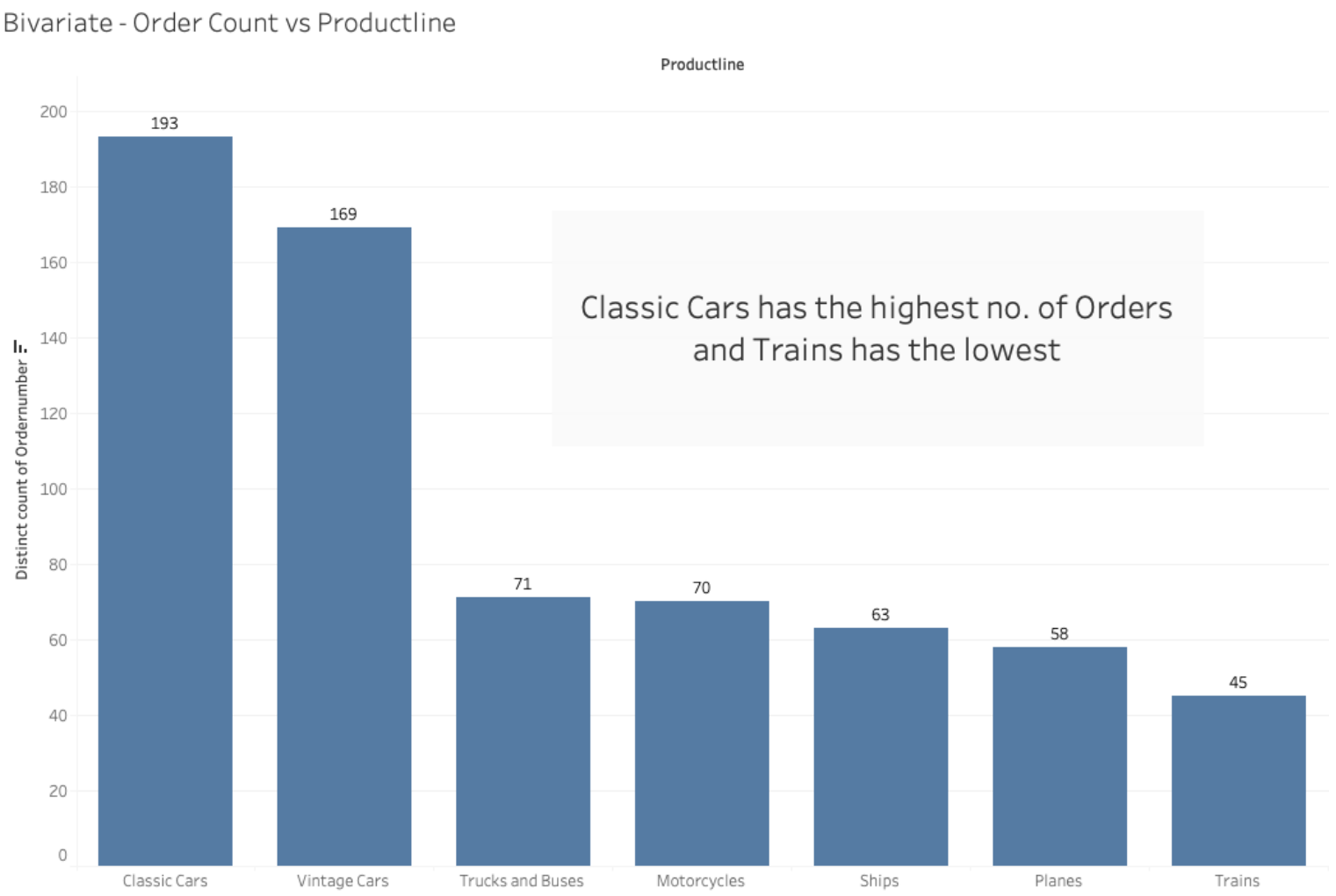


Bivariate Analysis



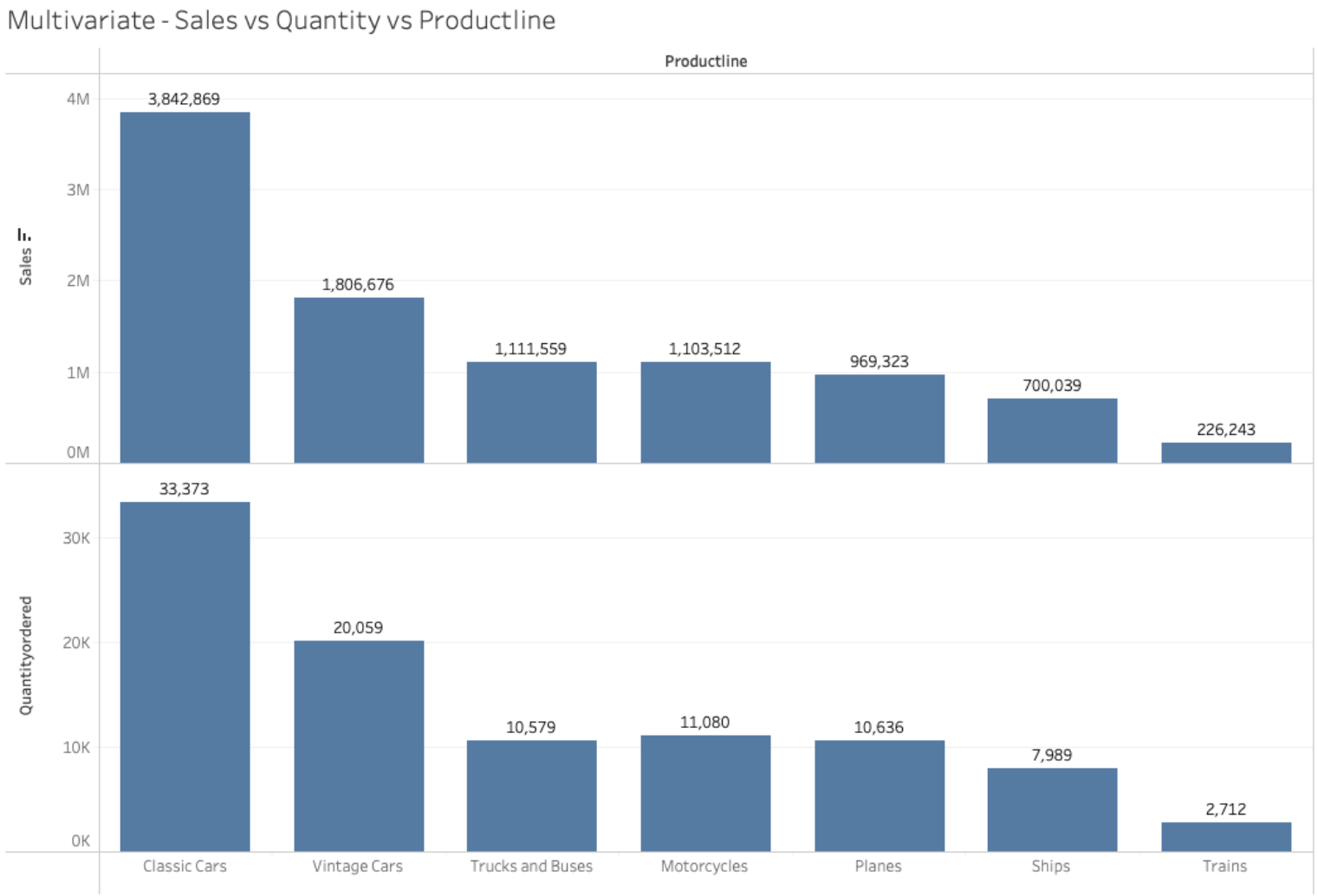
Bivariate Analysis

Bivariate Analysis

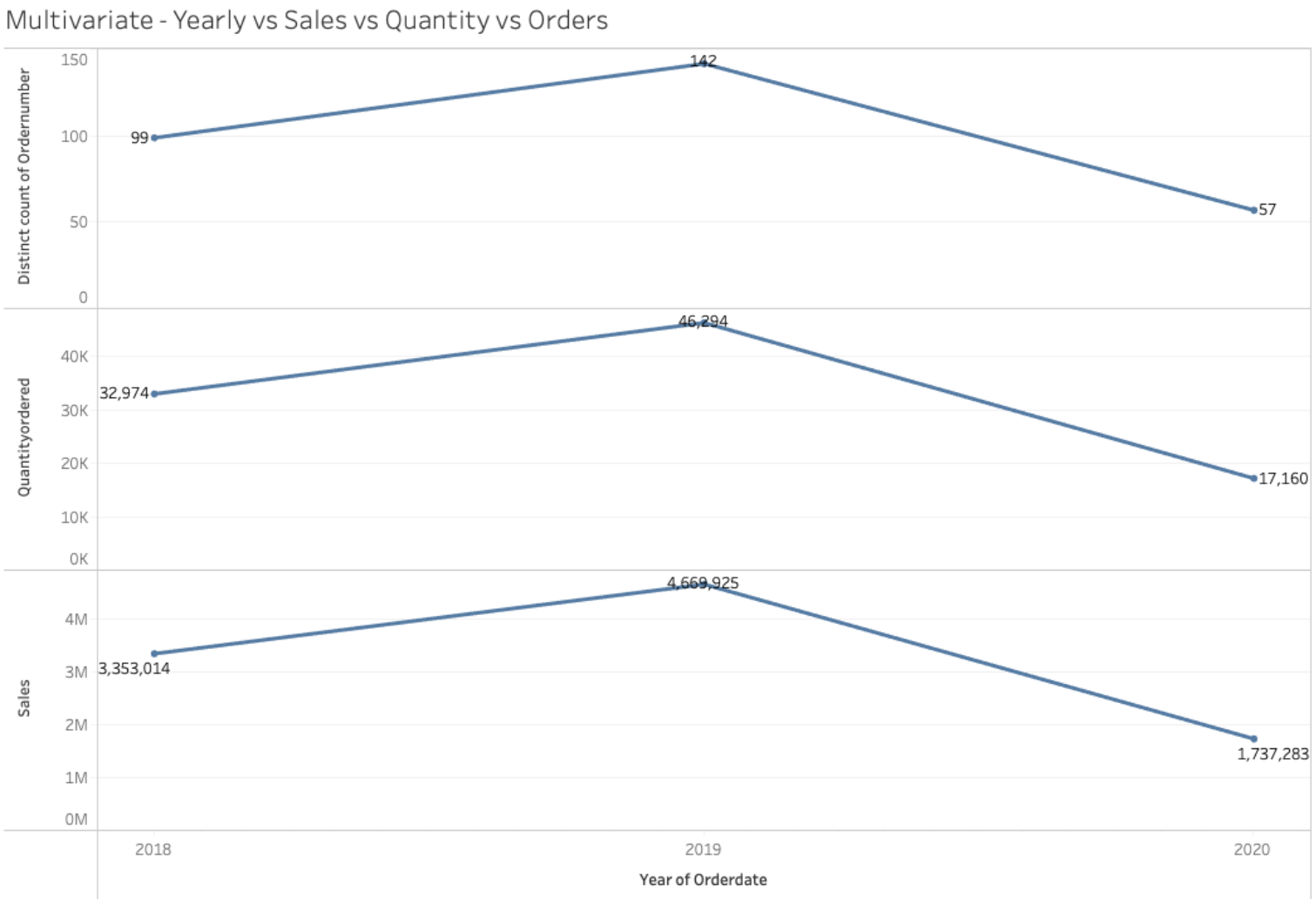


Multivariate Analysis

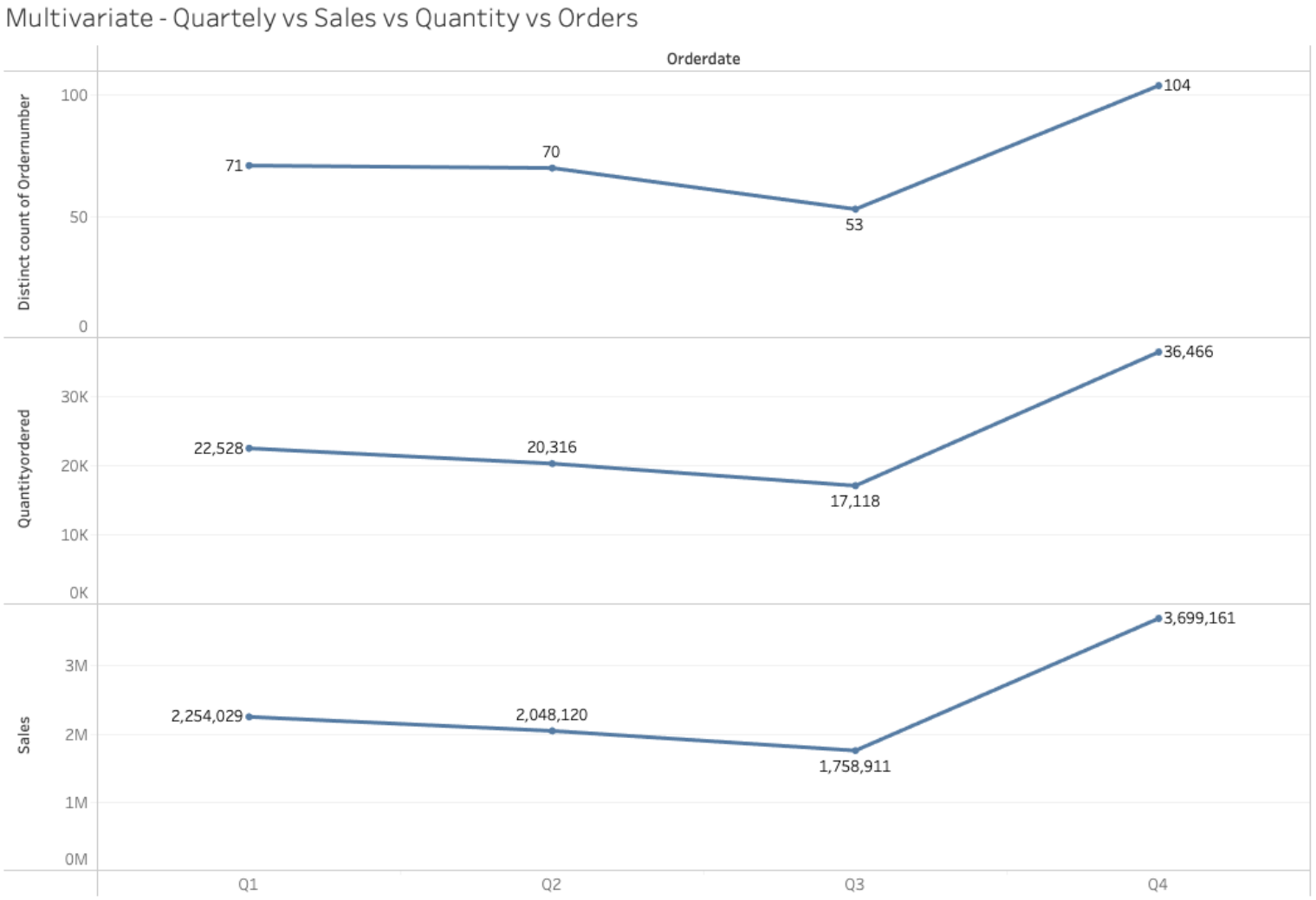
Multivariate Analysis



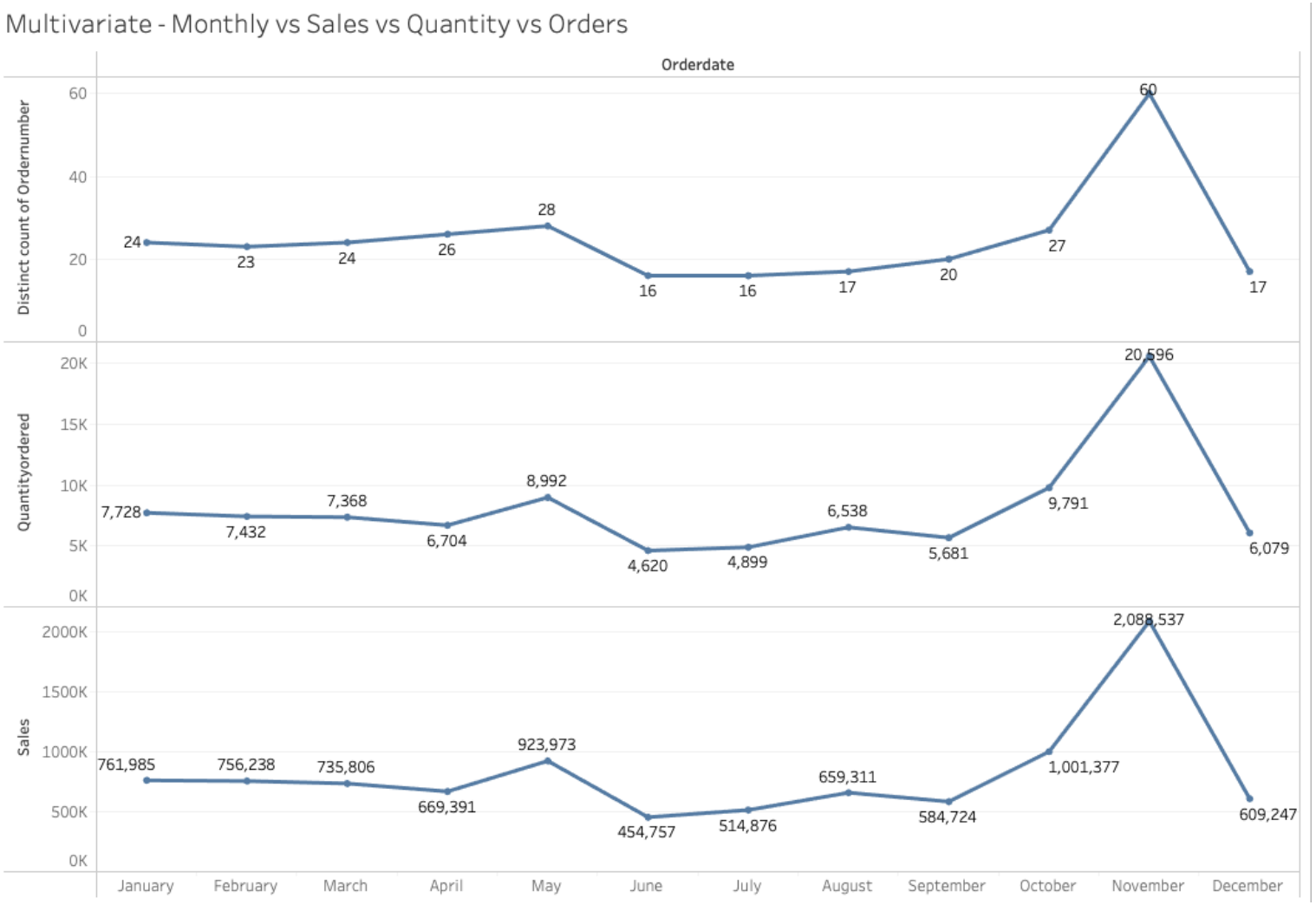
Yearly Trend



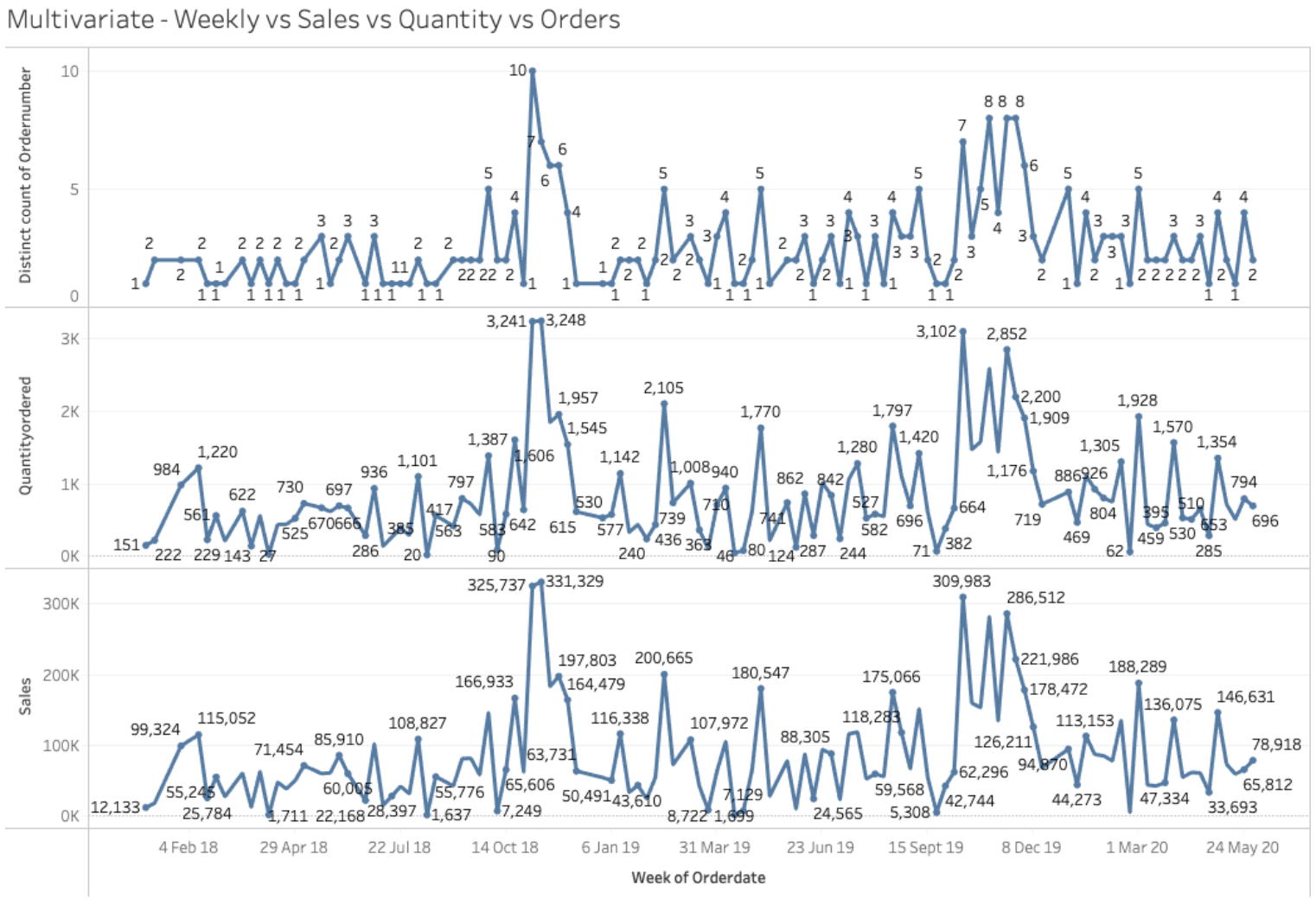
Quarterly Trend



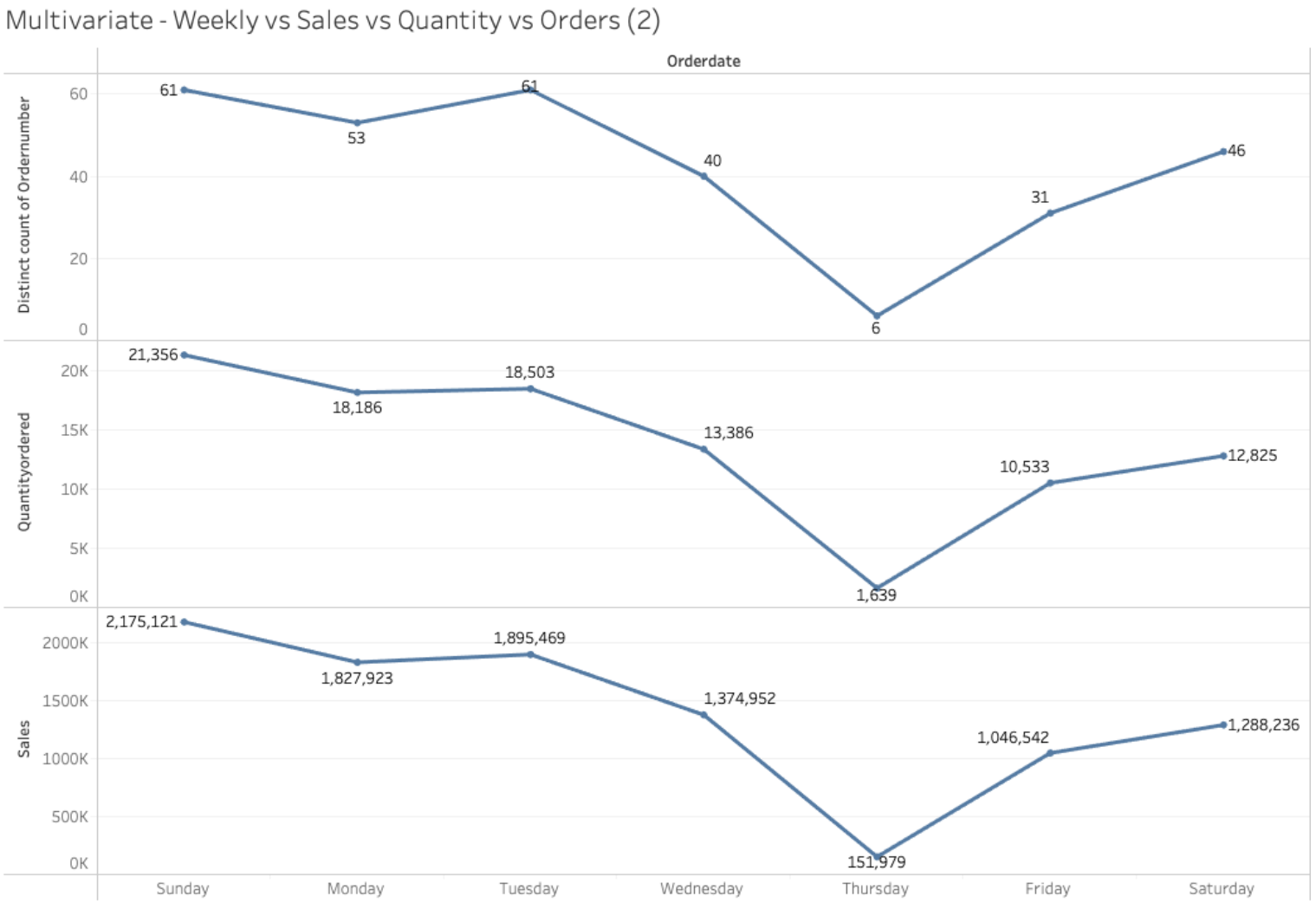
Monthly Trend



Weekly Trend



Weekday Trend



EDA - Inferences

- Quantity ordered has a median of 62. Right Skewed with outliers
- Deal size of medium is higher
- Classic Cars has the highest items ordered, where the least is Trains
- USA has the highest no. of customers of 32 and France at the second highest with 12
- New York city with highest no. of customers
- Deal size: Medium and Small has 89 customers and Large has 68 customers ordered
- Product line - Classic Cars has the higher no. of customers ordered and Trains has the lowest no. of customers ordered

EDA - Inferences

- 89 customers have orders that are shipped
- USA has the highest Orders with 103 and Switzerland & Ireland with 2 orders each
- Madrid city has the highest Orders with 31 and Munich with 1 order
- Large deal size has the lowest order count of 114
- Yearly sales, quantity ordered, and order count shows an increasing trend from 2018-19 and decreasing from 2019-20
- There was a sharp increase in the month of November across years
- There was a dip in sales during Thursday

Customer Segmentation using RFM analysis (4 segments)

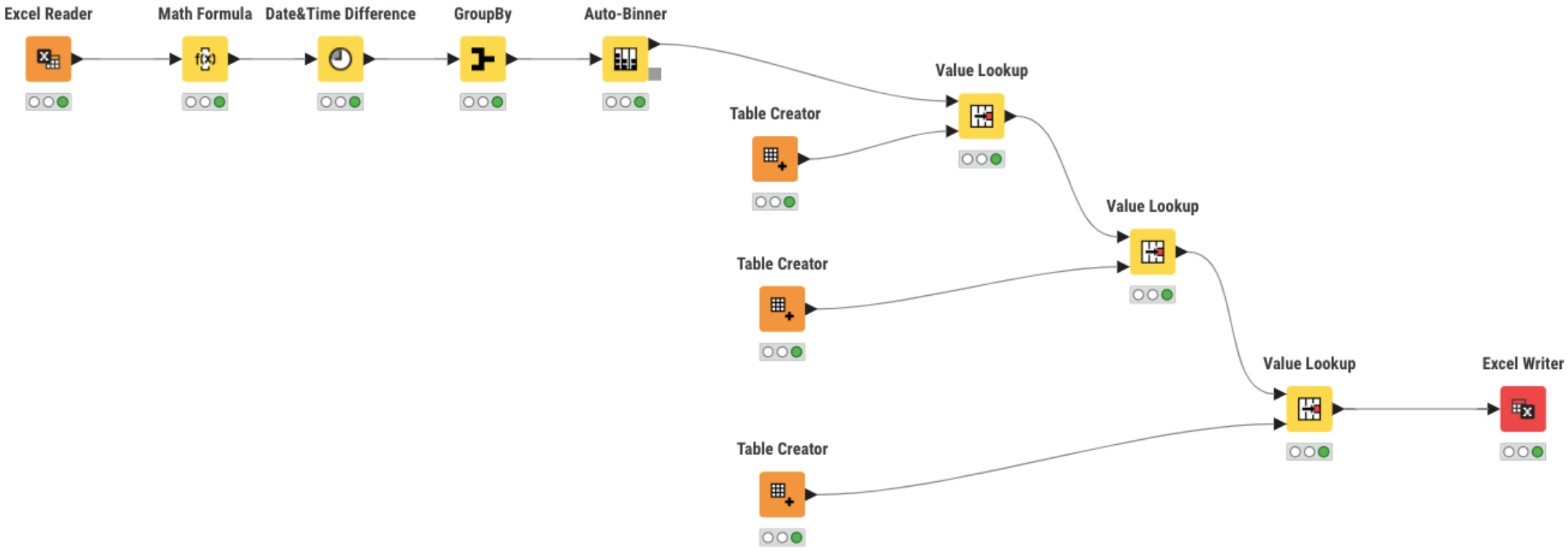
What is RFM?

RFM is a technique used in Sales & Marketing domain for customer segmentation based on their purchase pattern in terms of Recency, Frequency and Monetary

What all parameters used, and assumptions made?

- **RECENCY:** Days since the last order. Calculated by subtracting the order date from the max. date in the data set given (01-AUG-2020)
- **FREQUENCY:** The Order number count. The unique orders are considered customer wise
- **MONETARY:** The sum of total sales value (Quantity Ordered x Price of Each Product) customer wise

Showcase the KNIME workflow image



What results are there in the output table head?

Unique count*(ORDERNUMBER) [Binned]	Min*(RECENCY) [Binned]	Sum(MONETARY) [Binned]

Recency_Bin	RECENCY	Match Found (#2)

Freq_Bin	FREQUENCY	Match Found

Monetary_Bin	MONETARY	Match Found (#1)

What results are there in the output table head?

Unique count*(ORDERNUMBER) [Binned]	Min*(RECENCY) [Binned]	Sum(MONETARY) [Binned]	Freq_Bin	FREQUENCY	Match Found	Monetary_Bin	MONETARY	Match Found (#1)	Recency_Bin	RECENCY	Match Found (#2)
Bin 2	Bin 3	Bin 4	Bin 2	LM	TRUE	Bin 4	H	TRUE	Bin 3	LM	TRUE
Bin 2	Bin 1	Bin 1	Bin 2	LM	TRUE	Bin 1	L	TRUE	Bin 1	H	TRUE
Bin 1	Bin 4	Bin 3	Bin 1	L	TRUE	Bin 3	HM	TRUE	Bin 4	L	TRUE
Bin 4	Bin 2	Bin 4	Bin 4	H	TRUE	Bin 4	H	TRUE	Bin 2	HM	TRUE
Bin 2	Bin 3	Bin 1	Bin 2	LM	TRUE	Bin 1	L	TRUE	Bin 3	LM	TRUE
Bin 2	Bin 1	Bin 1	Bin 2	LM	TRUE	Bin 1	L	TRUE	Bin 1	H	TRUE
Bin 4	Bin 2	Bin 4	Bin 4	H	TRUE	Bin 4	H	TRUE	Bin 2	HM	TRUE
Bin 2	Bin 2	Bin 1	Bin 2	LM	TRUE	Bin 1	L	TRUE	Bin 2	HM	TRUE
Bin 1	Bin 4	Bin 1	Bin 1	L	TRUE	Bin 1	L	TRUE	Bin 4	L	TRUE
Bin 2	Bin 1	Bin 3	Bin 2	LM	TRUE	Bin 3	HM	TRUE	Bin 1	H	TRUE
Bin 2	Bin 2	Bin 1	Bin 2	LM	TRUE	Bin 1	L	TRUE	Bin 2	HM	TRUE
Bin 4	Bin 3	Bin 3	Bin 4	H	TRUE	Bin 3	HM	TRUE	Bin 3	LM	TRUE
Bin 1	Bin 4	Bin 1	Bin 1	L	TRUE	Bin 1	L	TRUE	Bin 4	L	TRUE

Inferences from RFM Analysis and identified segments

RFM Matrix

RECENCY	FREQUENCY	H	HM	LM	L	Grand Total
H	H	9	2	0	0	11
	HM	0	0	0	0	0
	LM	1	3	6	2	12
	L	0	0	0	0	0
H Total		**10**	**5**	**6**	**2**	**23**
HM	H	5	0	0	1	6
	HM	0	0	0	0	0
	LM	1	7	3	4	15
	L	0	0	0	1	1
HM Total		**6**	**7**	**3**	**6**	**22**
LM	H	1	1	2	0	4
	HM	0	0	0	0	0
	LM	4	3	3	1	11
	L	0	2	3	2	7
LM Total		**5**	**6**	**8**	**3**	**22**
L	H	0	0	0	0	0
	HM	0	0	0	0	0
	LM	1	1	0	0	2
	L	0	3	5	12	20
L Total		**1**	**4**	**5**	**12**	**22**
Grand Total		**22**	**22**	**22**	**23**	**89**

RFM Matrix

- 9 customers with HHH are active customers. They are the golden customers of the organization
- 44 customers on the segment of LM & HM are at risk. They need more attention where to be focused to get them in the high recency & frequency range
- 12 customers with LLL are almost inactive customers. Reasons need to be analyzed to get then back in the business

Who are your best customers? (give at least 5)

CUSTOMERNAME	Unique count*(ORDERNUMBER)	Min*(RECENCY)	Sum(MONETARY)
Euro Shopping Channel	26	1	912294
Mini Gifts Distributors Ltd.	17	3	654858
La Rochelle Gifts	4	1	180125
Souveniers And Things Co.	4	3	151571
Salzburg Collectables	4	15	149799

Recency: H, Frequency: H, Monetary: H

Which customers are on the verge of churning? (give at least 5)

CUSTOMERNAME	Unique count*(ORDERNUMBER)	Min*(RECENCY)	Sum(MONETARY)
Australian Collectables, Ltd	3	23	64591
Alpha Cognac	3	65	70488
Quebec Home Shopping Network	3	31	74205
Petit Auto	3	2	74973
Lyon Souveniers	3	76	78570

Recency: H, Frequency: L, Monetary: L

Who are your lost customers? (give at least 5)

CUSTOMERNAME	Unique count*(ORDERNUMBER)	Min*(RECENCY)	Sum(MONETARY)
Anna's Decorations, Ltd	4	84	153996
Australian Collectors, Co.	5	185	200995
Dragon Souveniers, Ltd.	5	91	172990
Land of Toys Inc.	4	199	164069
Muscle Machine Inc	4	183	197737

Recency: L, Frequency: H, Monetary: H

Who are your loyal customers? (give at least 5)

CUSTOMERNAME	Unique count*(ORDERNUMBER)	Min*(RECENCY)	Sum(MONETARY)
Euro Shopping Channel	26	1	912294
Mini Gifts Distributors Ltd.	17	3	654858
Danish Wholesale Imports	5	47	145042
Reims Collectables	5	63	135043
La Rochelle Gifts	4	1	180125

Recency: H, Frequency: H, Monetary: H/M/L

Thank You