

MRA Project ML 1

Priyadharshini K

PGP – DSBA Nov-C Batch, 2021

kpriyadharshinik@gmail.com

Tableau Public Link

https://public.tableau.com/views/MRA_16325905246800/Sheet1?:language=en-US&:display_count=n&:origin=viz_share_link



Agenda [Table OF CONTENT]

- Problem Statement
- Executive Summary
- EDA
- EDA Summary
- RFM
- RFM Summary

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.

This project aims to find the underlying buying patterns of the customers of an automobile part manufacturer based on the past 3 years of the Company's transaction data and hence recommend customized marketing strategies for different segments of customers.

ABOUT DATA

- ▶ The dataset has a total of datetime64(1), float64(2), int64(5), object(12) datatypes. The dataset has categorical, continuous and discrete data.
- ▶ Shape of the dataset: (2747,20)
- ▶ No duplicate values are present in the dataset.
- ▶ Null values/missing values are not present in the dataset.
- ▶ Object Datatype variables – STATUS, PRODUCTLINE, PRODUCTCODE, CUSTOMERNAME, PHONE, ADDRESSLINE1, CITY, POSTALCODE, COUNTRY, CONTACTLASTNAME, CONTACTFIRSTNAME, DEALSIZE
- ▶ Int Datatype variables – ORDERNUMBER, QUANTITYORDERED, ORDERLINENUMBER, DAYS_SINCE_LASTORDER, MSRP
- ▶ Float Datatype variables Discrete variable – PRICEEACH, SALES

EDA - Univariate Analysis

Tool Used: Python

➤ Skewness

```
Skewness of ORDERNUMBER : -0.006991678308042113
Skewness of QUANTITYORDERED : 0.36908467165651987
Skewness of PRICEEACH : 0.6968408548715753
Skewness of ORDERLINENUMBER : 0.5750125802758338
Skewness of SALES : 1.1553084908947713
Skewness of DAYS_SINCE_LASTORDER : -0.00298177880902193
Skewness of MSRP : 0.5753319981585184
```

- Skewness is a measure of symmetry, or more precisely, the lack of symmetry.
- For skewness, if the value is greater than + 1.0, the distribution is right skewed.
- If the value is less than -1.0, the distribution is left skewed. Skewness of the variables present in the dataset is between -0.5 to 0.5, indicates that the data is symmetrical.
- Sales - Right skewed
- MSRP, Quantity ordered, order number, days since last order - Symmetrical

➤ Kurtosis

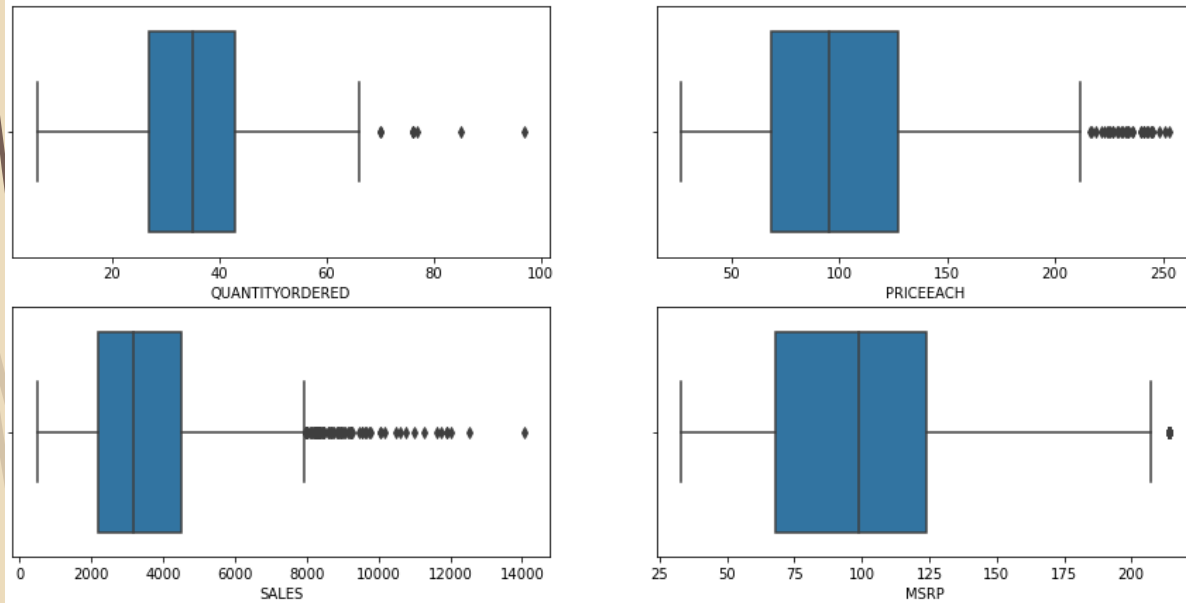
```
Kurtosis of ORDERNUMBER : -1.1544900051275273
Kurtosis of QUANTITYORDERED : 0.43987587769877345
Kurtosis of PRICEEACH : 0.22591902460639401
Kurtosis of ORDERLINENUMBER : -0.5921445636443714
Kurtosis of SALES : 1.767690983161713
Kurtosis of DAYS_SINCE_LASTORDER : -1.0247855171043092
Kurtosis of MSRP : -0.14141981873118636
```

- Kurtosis is a measure of whether the data are heavy-tailed or light-tailed relative to a normal distribution. Data sets with high kurtosis tend to have heavy tails, or outliers. Data sets with low kurtosis tend to have light tails, or lack of outliers.
- For kurtosis, the general guideline is that if the number is greater than +1, the distribution is too peaked. Likewise, a kurtosis of less than -1 indicates a distribution that is too flat.
- Quantity ordered, Price each , MSRP - follows relative normal distribution
- Sales - Heavy tailed towards left

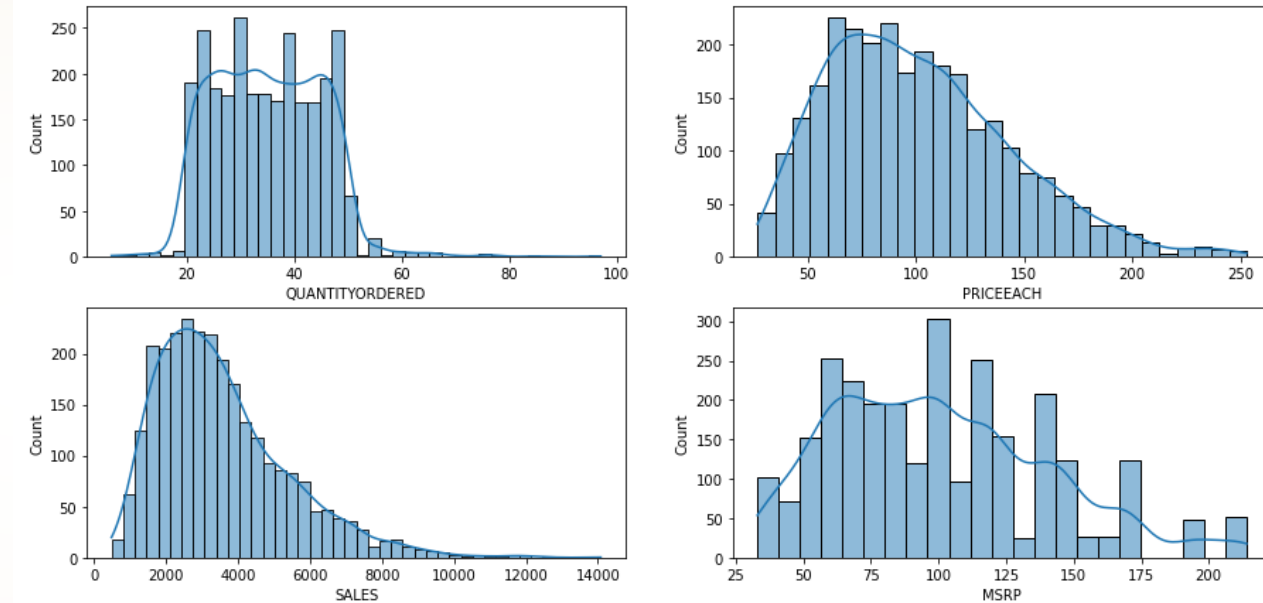
EDA - Univariate Analysis

Tool Used: Python

Box Plot



Histogram

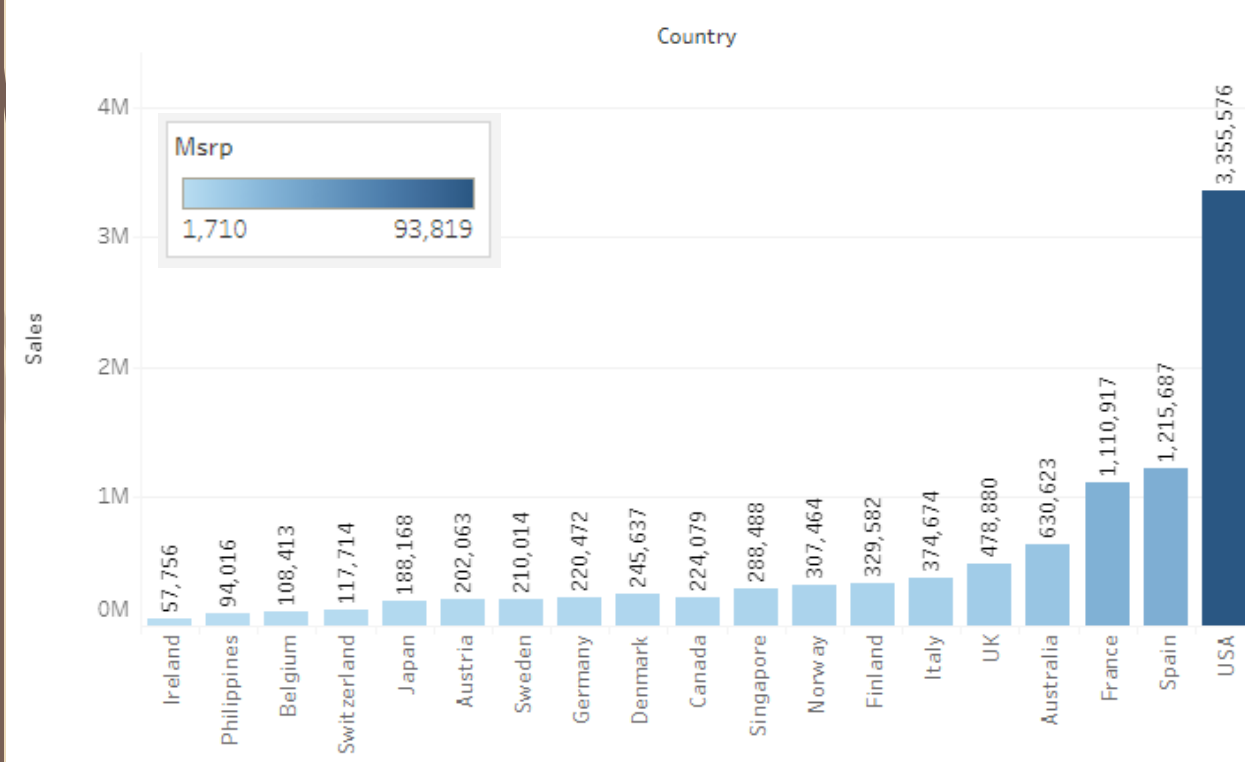


- Outliers are present in the dataset.
- Sales, Price each, Quantity ordered, MSRP variable is normally distributed.

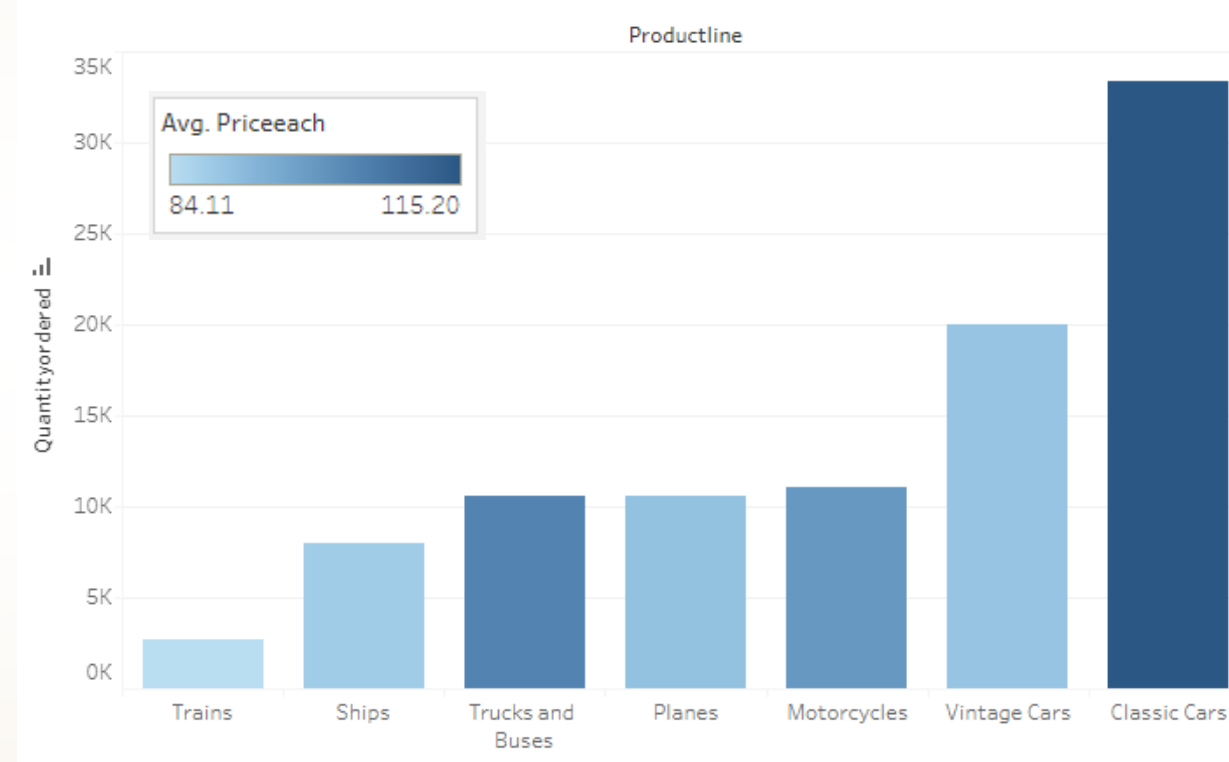
EDA - Multivariate Analysis

Tool Used: Tableau

Country vs Sales

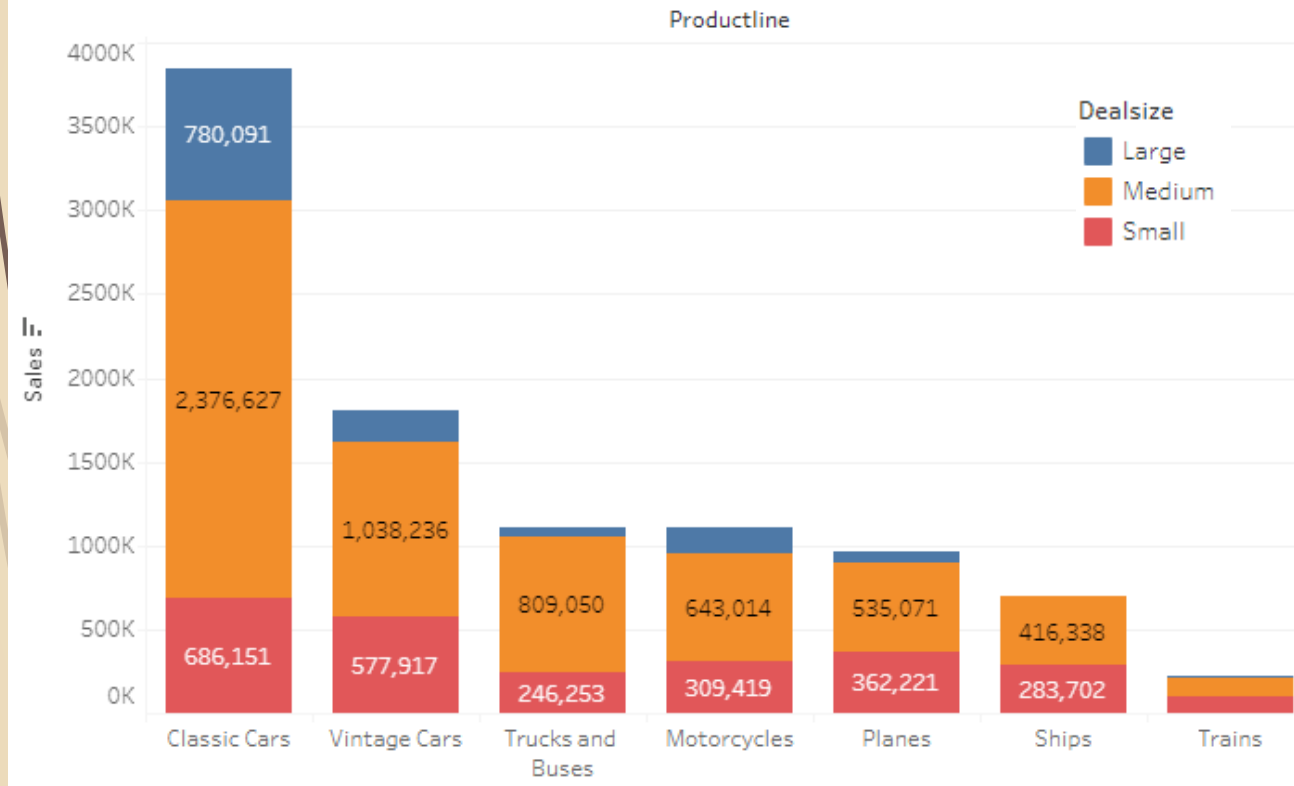


Product line vs Quantity Ordered

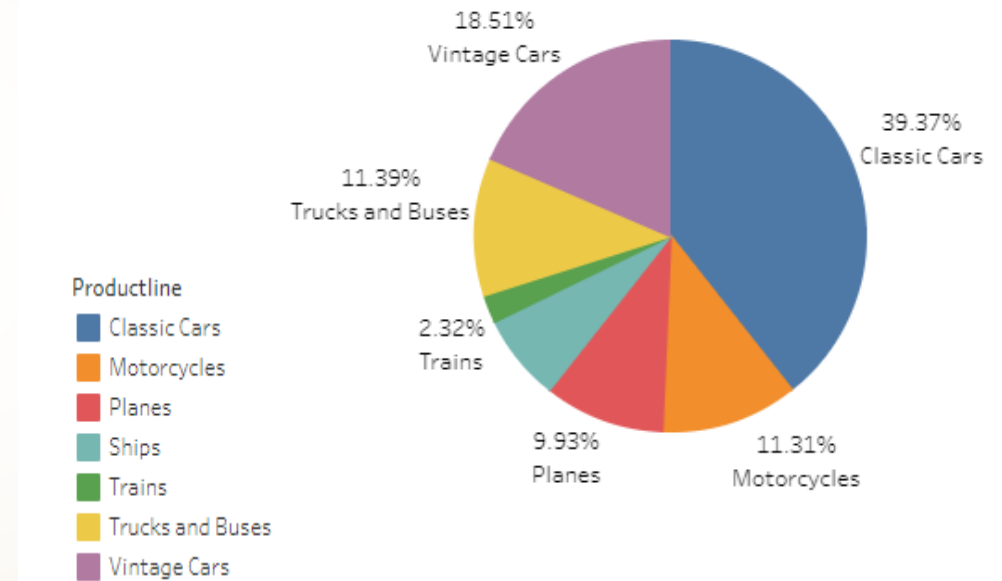


- USA ranks first in terms of Sales and MSRP followed by Spain, France, Australia and UK in top 5 places.
- Classic Cars tops most in terms Quantity ordered.

Product Line vs Sales (Deal size)

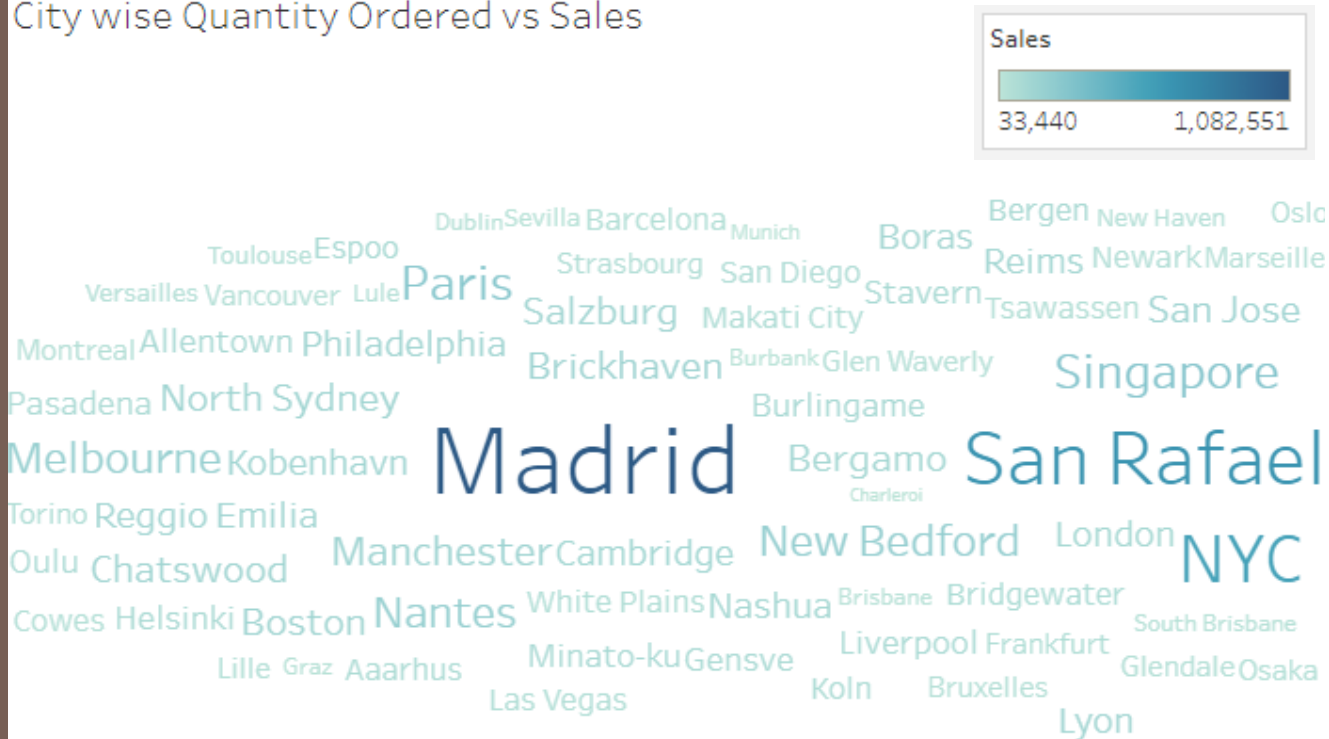


Product Line vs Sales

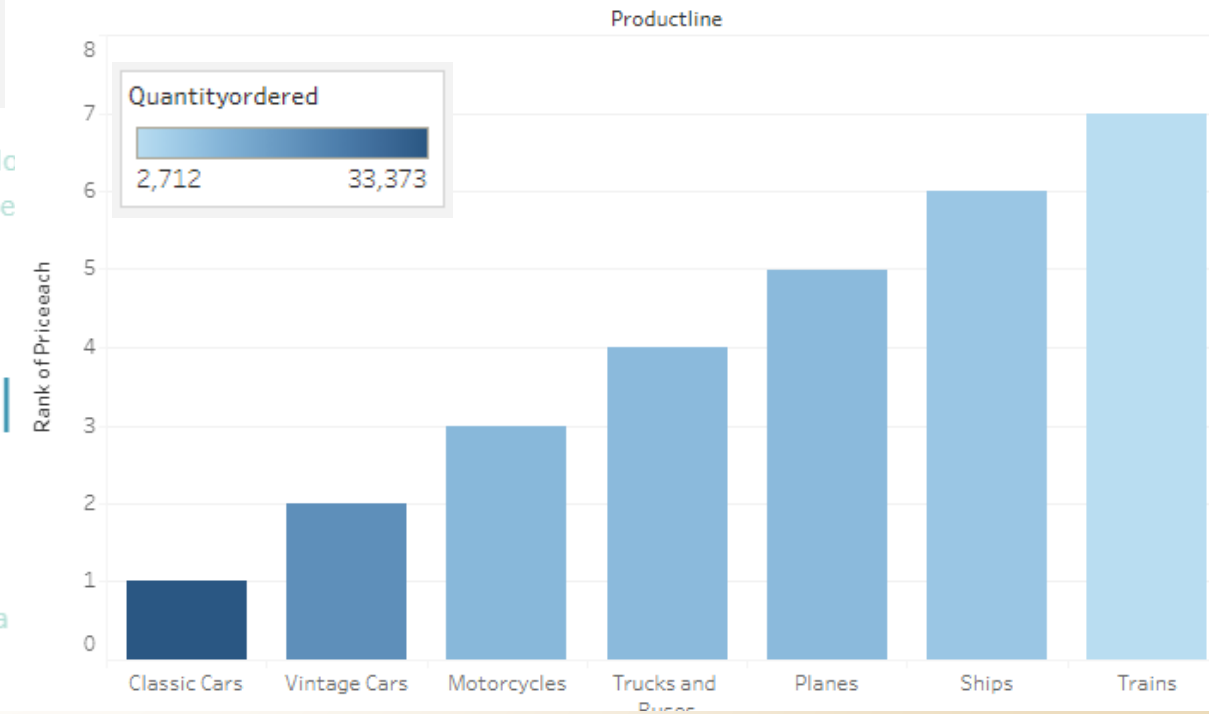


- Classic Cars tops first in terms of Sales, followed by Vintage cars and Trucks/Buses with predominant Medium size deals.
- Classic Cars contributes to about 40% of the total sales with trains being minimum with 2%

City wise Quantity Ordered vs Sales



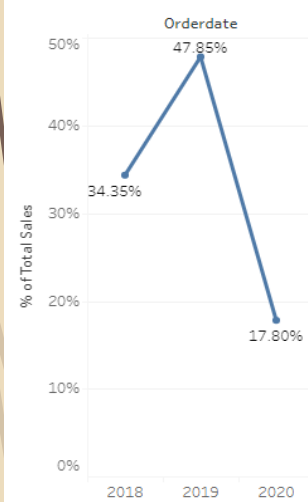
Rank wise price of Product Line



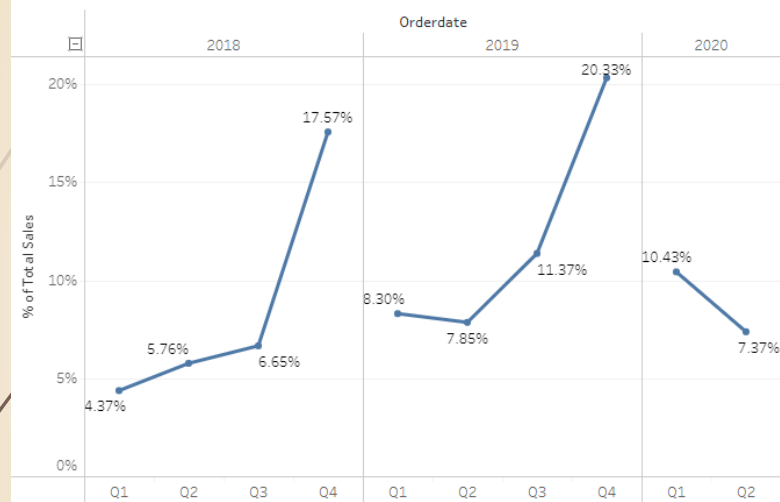
- Madrid tops Quantity Ordered, and Sales followed by San Rafael and NYC.
- Price of Classic cars is the minimum, but Quantity ordered for Classic cars is the maximum with Trains being costliest with minimum order quantity.

Weekly, Monthly, Quarterly, Yearly Trends in Sales

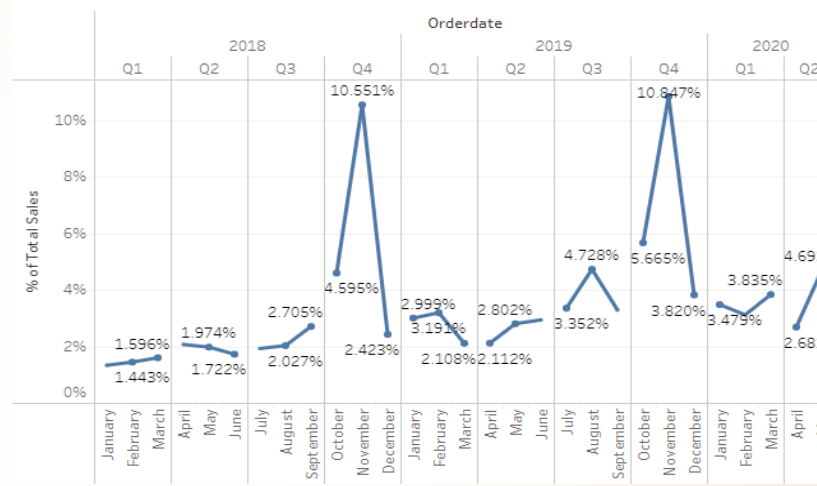
Yearly Sales trend



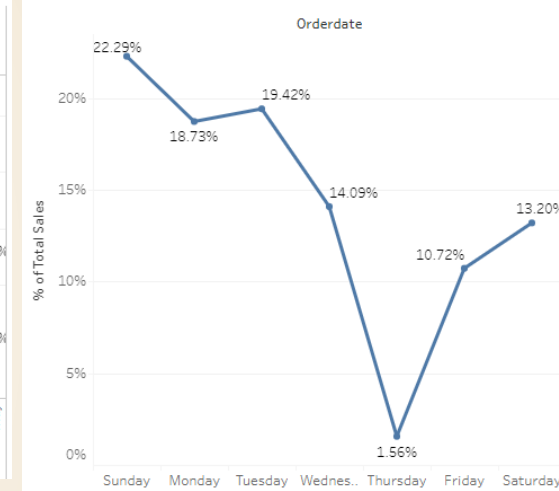
Quarterly Sales trend



Monthly Sales trend

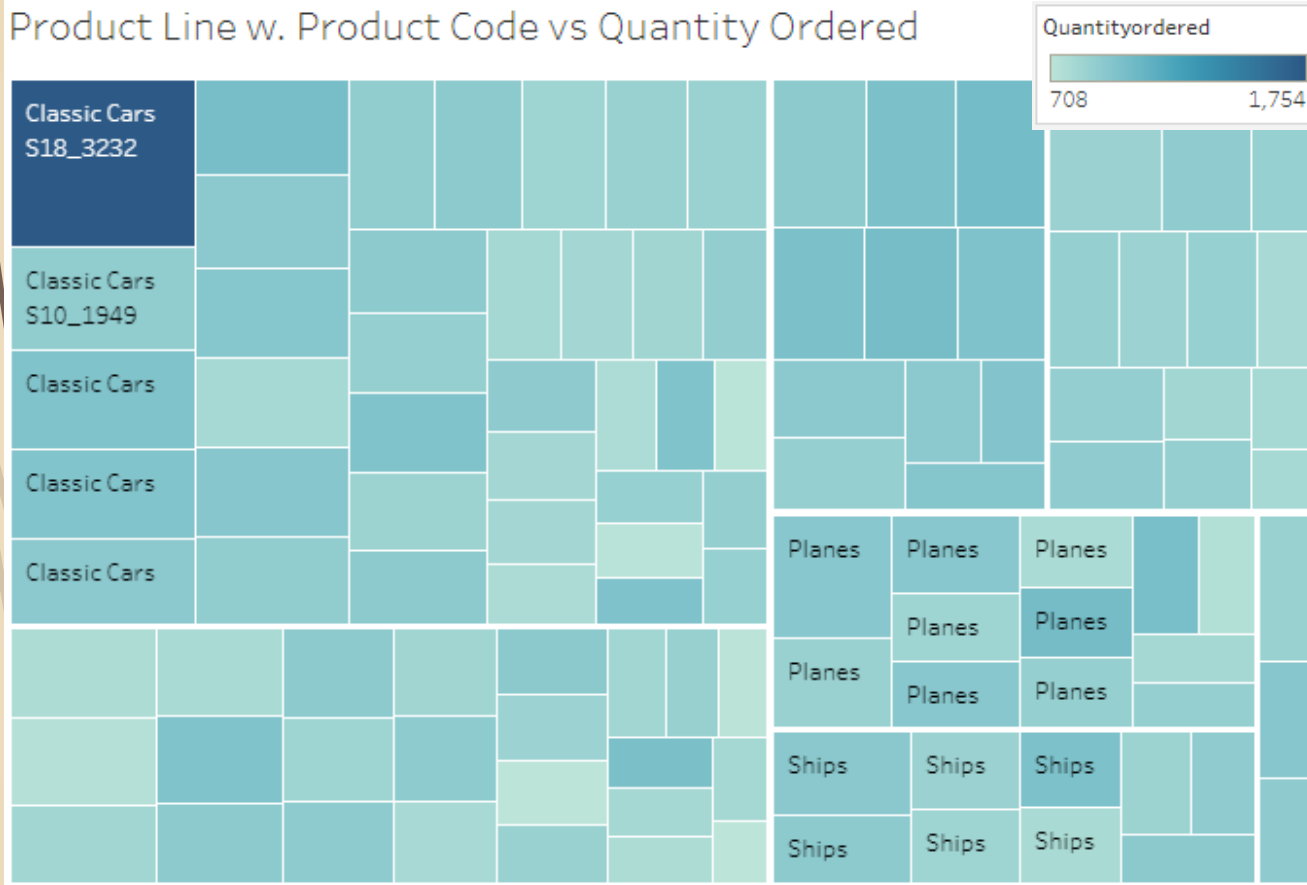


Weekly Sales Trend

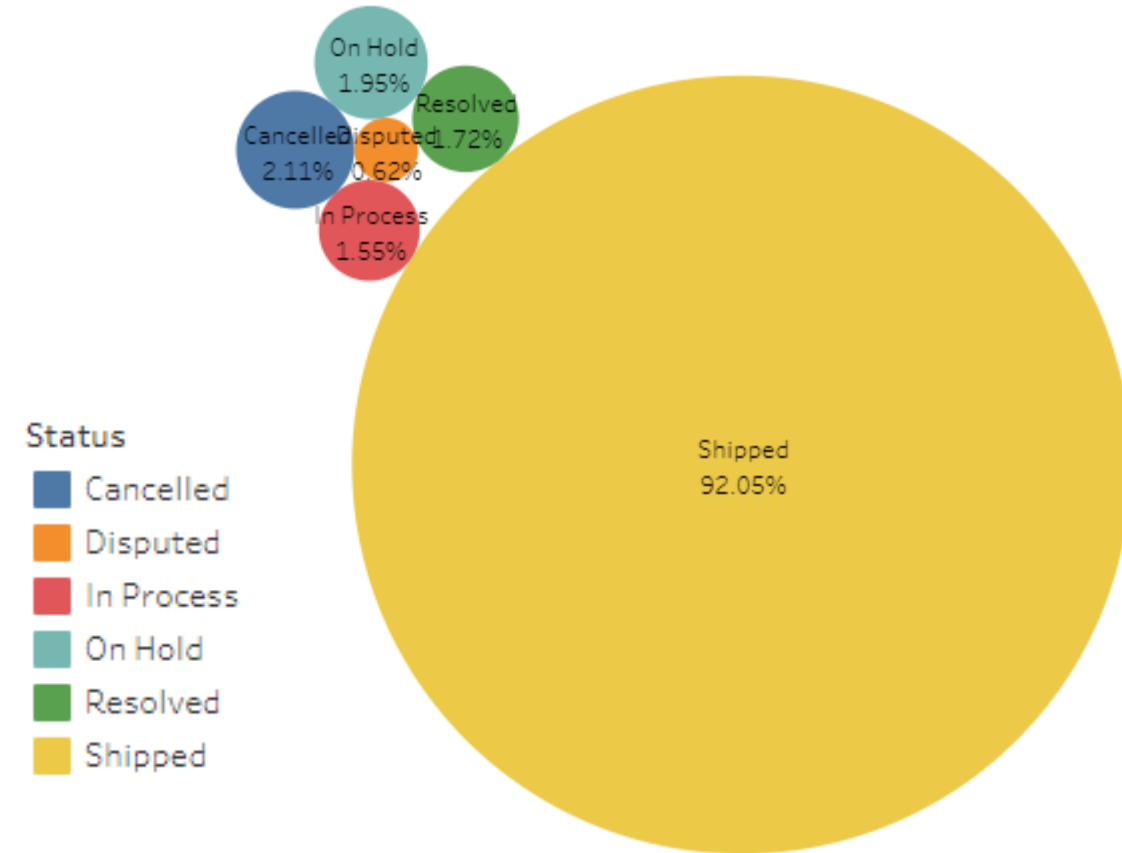


- Yearly : 2019 has recorded the max. sales (47% of Total Sales.)
- Quarterly : Q4 has the maximum sales with Q1 minimum.
- Monthly : November month has the peak Sales
- Weekly : Sunday records the max. sales (22%) while Thursday records minimum sales (1%)

Product Line w. Product Code vs Quantity Ordered

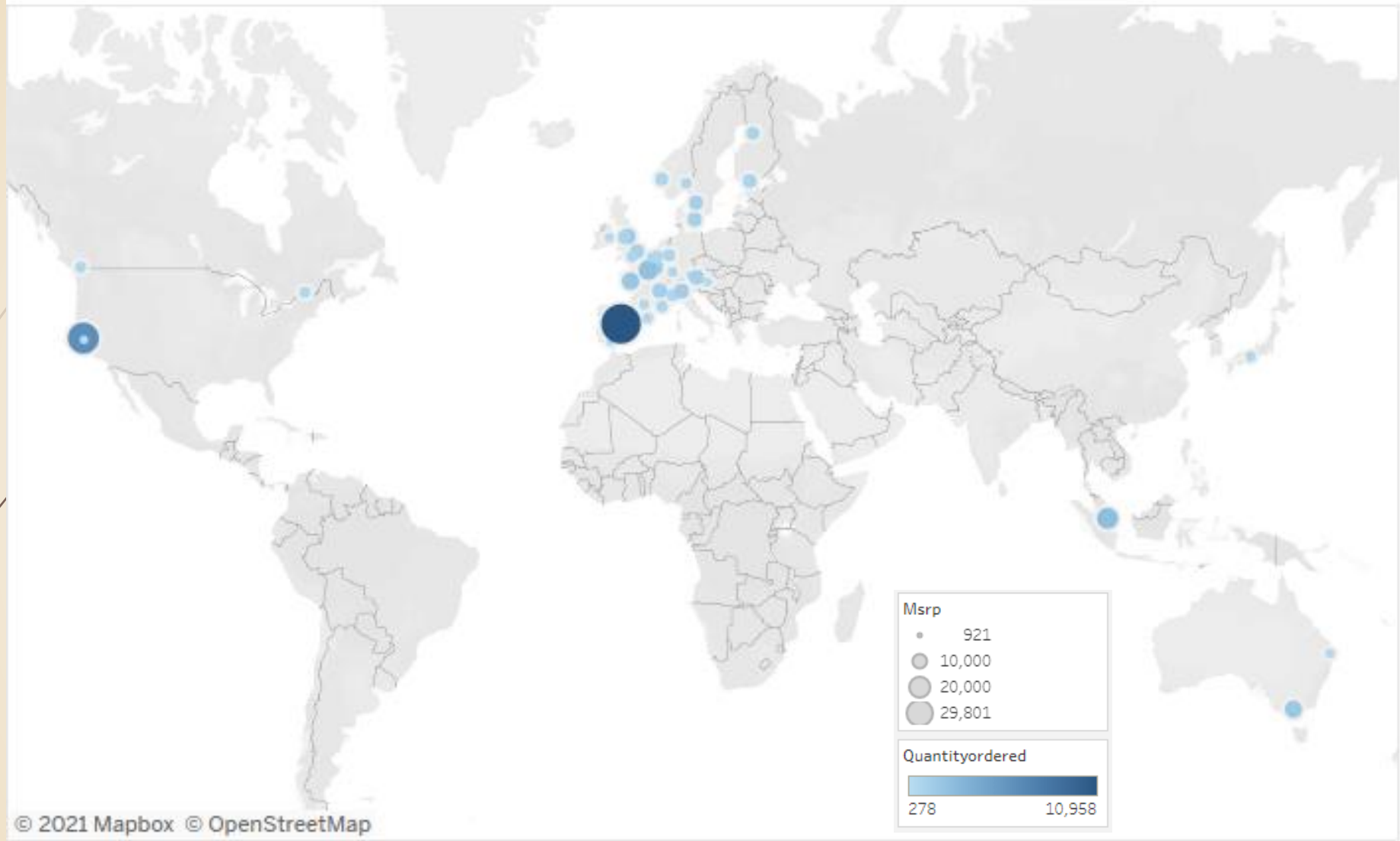


Quantity Ordered vs Status



- Classic Cars with Product code S18_3232 has the maximum Quantity ordered
- Only 92% of the total orders get shipped, while 2 % of orders get cancelled.

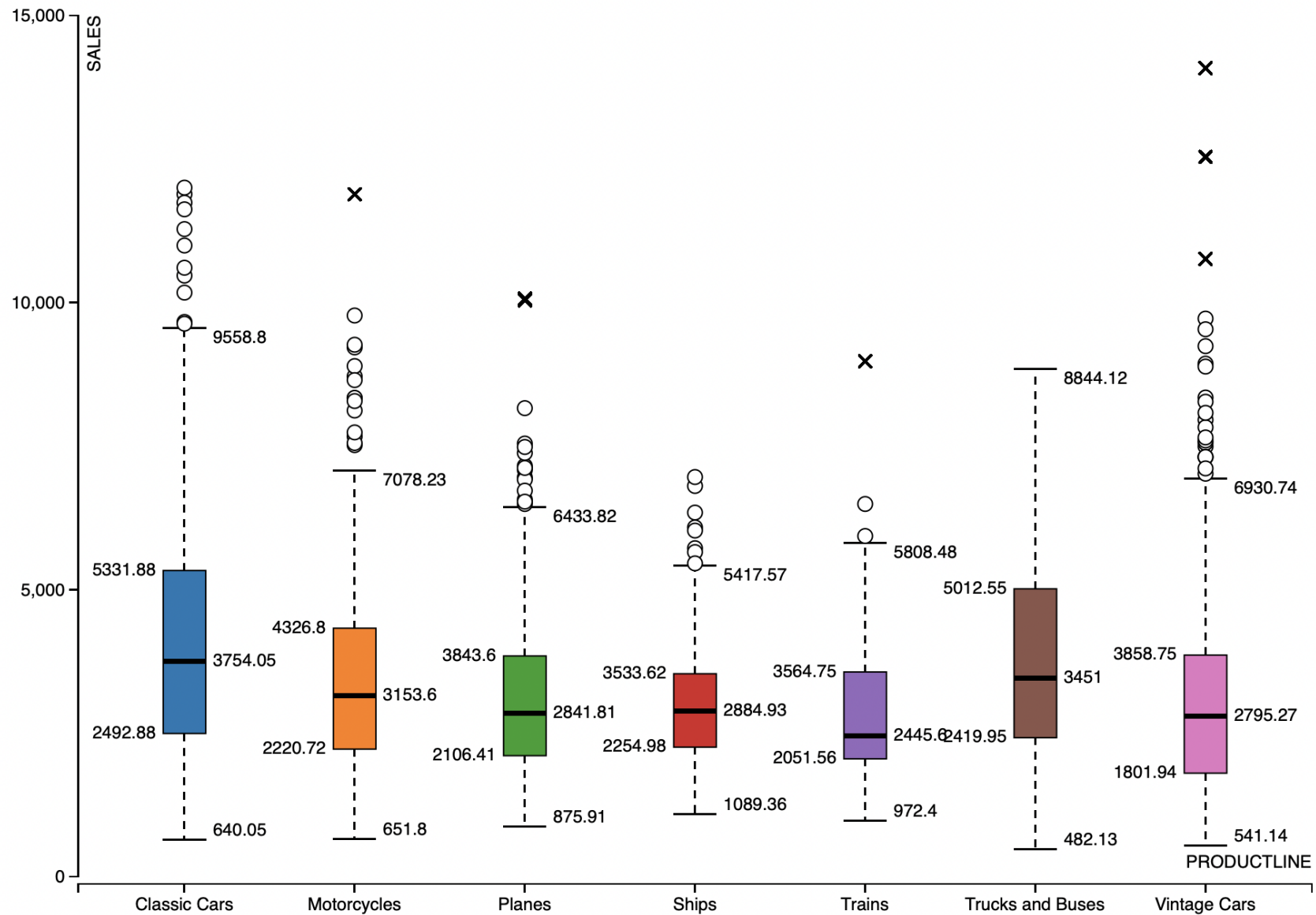
Country wise MSRP vs Quantity Ordered



Spain has the maximum of Quantity ordered, maximum MSRP

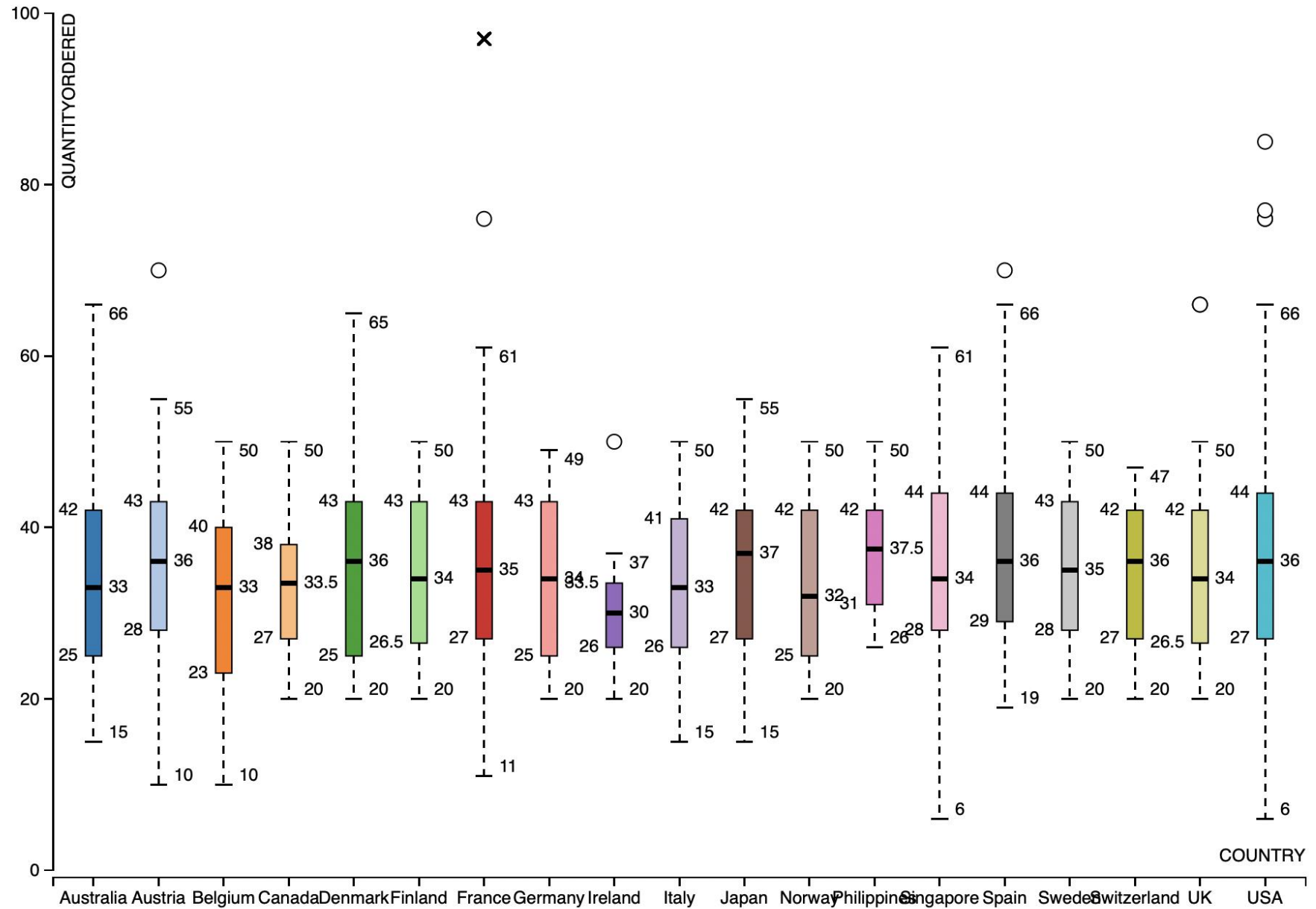
EDA IN KNIME

Conditional Box Plot



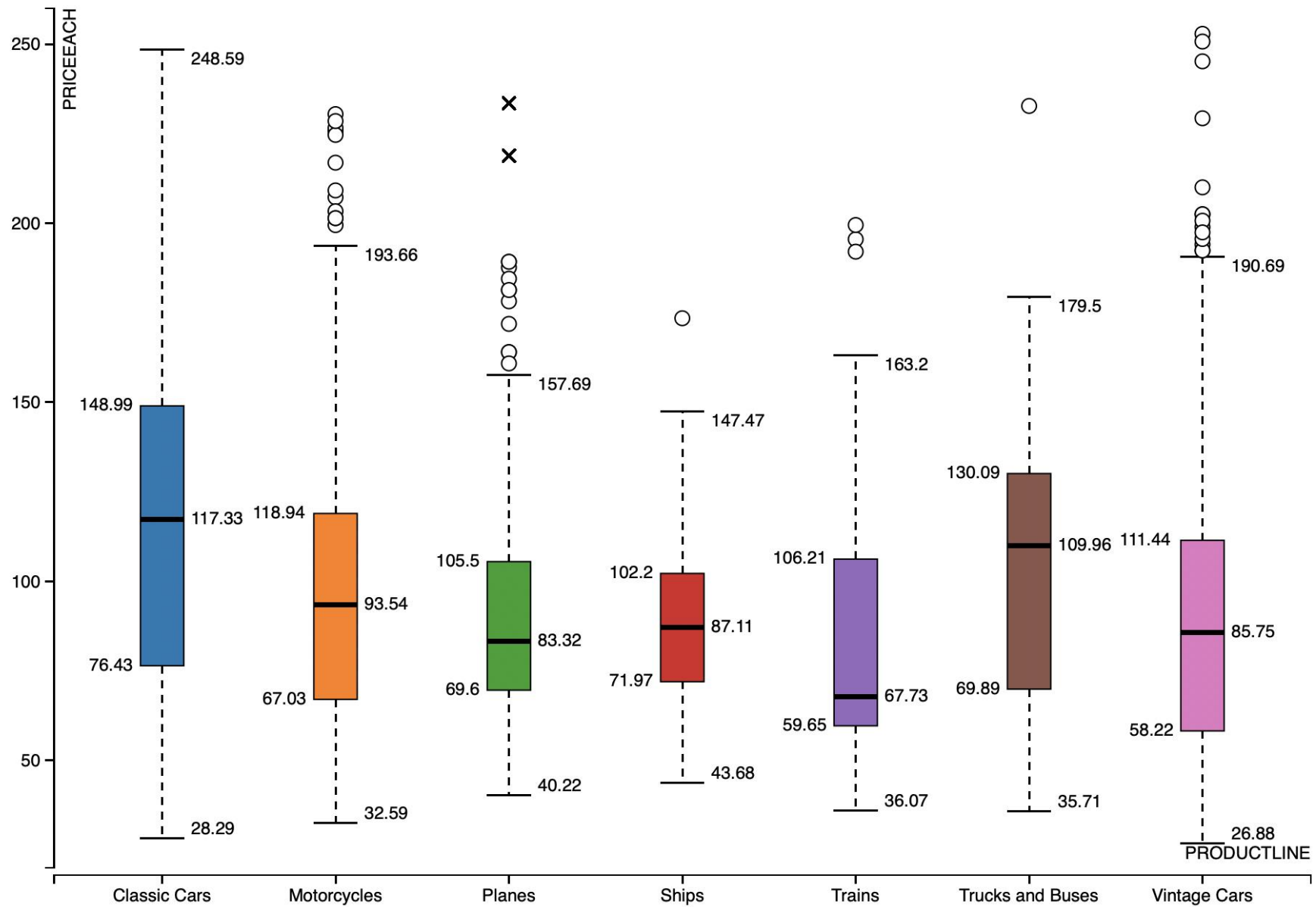
EDA IN KNIME

Conditional Box Plot



EDA IN KNIME

Conditional Box Plot



EDA IN KNIME














Numeric

Nominal

Data Preview

Clear Sorting

Search:

Column 	Exclude Column	Minimum 	Maximum 	Mean 	Standard Deviation 	Variance 
 ORDERNUMBER	<input type="checkbox"/>	10100	10425	10259.762	91.878	8441.479
 QUANTITYORDERED	<input type="checkbox"/>	6	97	35.103	9.762	95.299
 PRICEEACH	<input type="checkbox"/>	26.880	252.870	101.099	42.043	1767.576
 ORDERLINENUMBER	<input type="checkbox"/>	1	18	6.491	4.231	17.897
 SALES	<input type="checkbox"/>	482.130	14082.800	3553.048	1838.954	3381751.448
 DAYS_SINCE_LASTORDER	<input type="checkbox"/>	42	3562	1757.086	819.281	671220.663
 MSRP	<input type="checkbox"/>	33	214	100.692	40.115	1609.197

	QUANTITYORDERED	<input type="checkbox"/>	6	97	35.103	9.762	95.299	0.369
--	-----------------	--------------------------	---	----	--------	-------	--------	-------

Kurtosis 0.443

Overall Sum 96428

No. zeros 0

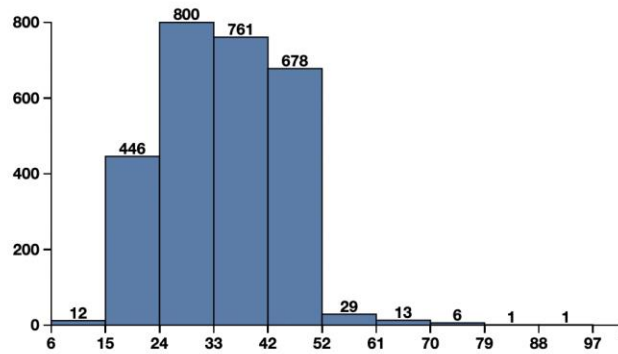
No. missings 0

No. NaN 0

No. +∞ 0

No. -∞ 0

Histogram



	PRICEEACH	<input type="checkbox"/>	26.880	252.870	101.099	42.043	1767.576	0.697
--	-----------	--------------------------	--------	---------	---------	--------	----------	-------

Kurtosis 0.229

Overall Sum 277718.819

No. zeros 0

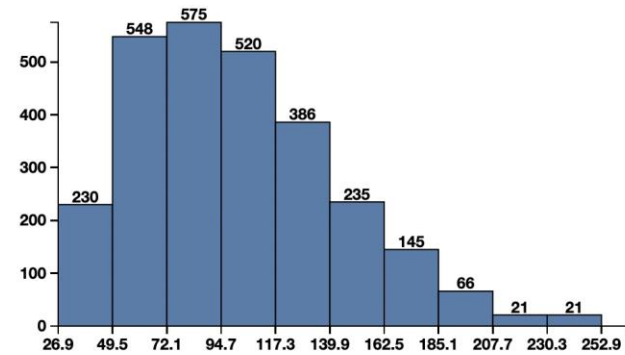
No. missings 0

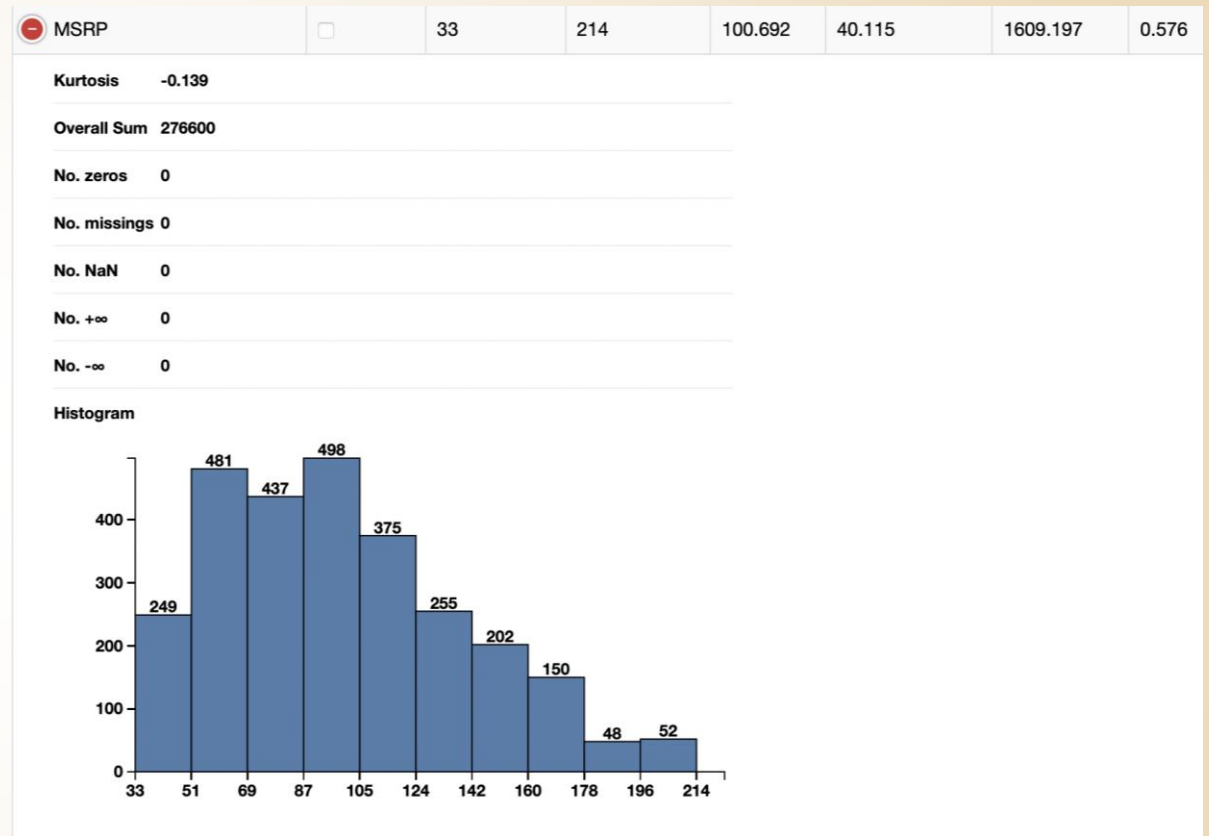
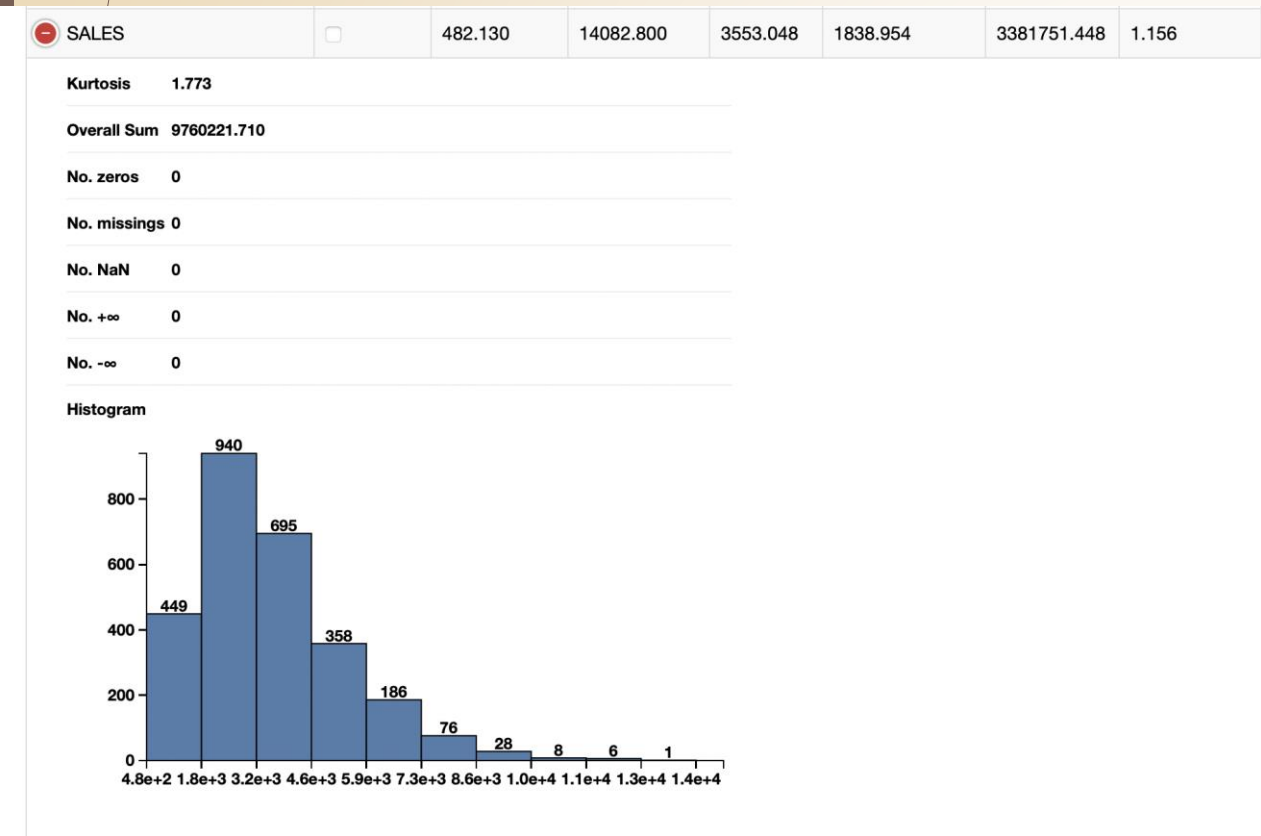
No. NaN 0









No. +∞ 0


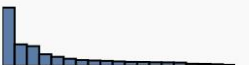

No. -∞ 0

Histogram





Column 	Exclude Column	No. missings 	Unique values 	All nominal values 	Frequency Bar Chart
STATUS	<input type="checkbox"/>	0	6	Shipped, Cancelled, Resolved, On Hold, In Process, Disputed	
PRODUCTLINE	<input type="checkbox"/>	0	7	Classic Cars, Vintage Cars, Motorcycles, Planes, Trucks and Buses, Ships, Trains	
PRODUCTCODE	<input type="checkbox"/>	0	109	S18_3232, S24_1444, S32_2509, S24_2840, S50_1392, [...], S18_4933, S18_4409, S18_1749, S24_3969, S18_2248	
CUSTOMERNAME	<input type="checkbox"/>	0	89	Euro Shopping Channel, Mini Gifts Distributors Ltd., Australian Collectors, Co., La Rochelle Gifts, AV Stores, Co., [...], Microscale Inc., Auto-Moto Classics Inc., Royale Belge, Atelier graphique, Boards & Toys Co.	

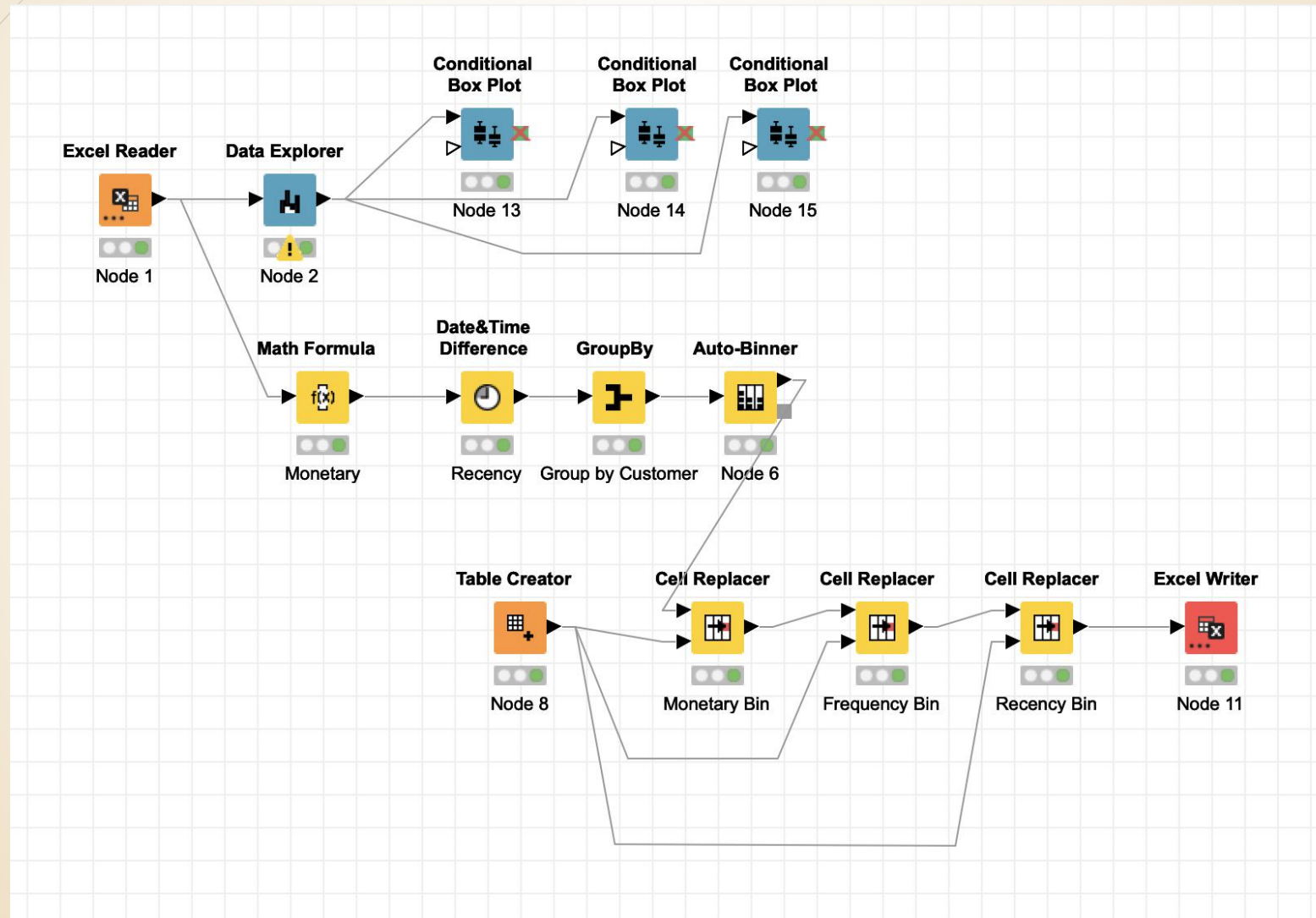
CITY	<input type="checkbox"/>	0	71	Madrid, San Rafael, NYC, Singapore, Paris, [...], Sevilla, South Brisbane, Munich, Burbank, Charleroi	
COUNTRY	<input type="checkbox"/>	0	19	USA, Spain, France, Australia, UK, [...], Japan, Belgium, Switzerland, Philippines, Ireland	
DEALSIZE	<input type="checkbox"/>	0	3	Medium, Small, Large	

RFM

Customer Segmentation using RFM is made using KNIME.

Segmented as VH (Best Customers), H (Loyal Customers), M (Good Customers), L (Lost/Inactive Customers)

KNIME Workflow:



ASSUMPTIONS MADE

- CUSTOMERNAME, ORDERLINENUMBER, SALES, STATUS, PRODUCTLINE, MSRP, PRODUCTCODE, ADDRESSLINE1, PHONE, CITY, POSTALCODE, COUNTRY, CONTACTFIRSTNAME, CONTACTLASTNAME, DEALSIZE do not significantly contribute to the RFM Analysis output
- Recency = Current Date – Order Date
- Frequency = Count of Order number
- Monetary = Quantity ordered * Price each

PARAMETERS USED

- ORDER NUMBER
- QUANTITY ORDERED
- PRICE EACH
- SALES
- ORDER DATE

HEAD OF OUTPUT TABLE

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	CUSTOMERNAME	ORDERNUMBER	QUANTITYORDERED	PRICEEACH	SALES	ORDERDATE	PRODUCTLINE	MSRP	Monetary	Recency	RNUMBER [Bin]	Monetary [Bin]	Recency [Bin]	Monetary_Bin	frequency_Bin	Recency_Bin
2	AV Stores, Co.	51	34.86	91.08	157807.81	43752.00	Vintage Cars	92.84	157807.81	679.00	Bin 4	Bin 4	Bin 3	VH	VH	M
3	Alpha Cognac	20	34.35	101.16	70488.44	43412.00	Ships	97.15	70488.44	547.00	Bin 1	Bin 1	Bin 1	L	L	VH
4	Amica Models &	26	32.42	110.85	94117.26	43694.00	Vintage Cars	107.65	94117.26	748.00	Bin 2	Bin 3	Bin 4	H	M	L
5	Anna's Decorati	46	31.93	106.42	153996.13	43899.00	Classic Cars	104.72	153996.13	566.00	Bin 4	Bin 4	Bin 2	VH	VH	H
6	Atelier graphique	7	38.57	92.24	24179.96	43735.00	Classic Cars	95.57	24179.96	671.00	Bin 1	Bin 1	Bin 3	L	L	M
7	Australian Colle	23	30.65	90.04	64591.46	43425.00	Vintage Cars	88.13	64591.46	505.00	Bin 2	Bin 1	Bin 1	L	M	VH
8	Australian Colle	55	35.02	104.59	200995.41	43516.00	Motorcycles	103.53	200995.41	667.00	Bin 4	Bin 4	Bin 2	VH	VH	H
9	Australian Gift M	15	36.33	110.55	59469.12	43410.00	Classic Cars	111.53	59469.12	602.00	Bin 1	Bin 1	Bin 2	L	L	H
10	Auto Assoc. & Ci	18	35.39	99.49	64834.32	43749.00	Vintage Cars	100.39	64834.32	716.00	Bin 1	Bin 1	Bin 4	L	L	L

Customers based on segments identified from RFM Analysis

Best customers (VH segment)

- Danish Wholesale Imports
- Euro Shopping Channel
- La Rochelle Gifts
- L'ordine Souveniers
- Mini Gifts Distributors Ltd



Loyal customers (H segment)

- Auto Canal Petit
- Tokyo Collectables, Ltd
- UK Collectables, Ltd.



Good customers (M segment)

- Gifts4AllAges.com
- Petit Auto
- Quebec Home Shopping Network
- Tekni Collectables Inc.



Lost customers (L segment)

- Alpha Cognac

Insights from RFM Analysis

- Coupons points to reward Best customers
- Exclusive Offers to retain Loyal customers
- Discounts to retain Good customers
- Targeted Marketing strategy to retain Lost customers

Recency	Frequency	Monetary				Grand Total
		VH	H	M	L	
VH	VH	9	1			10
	H	1	3			4
	M		1	4	1	6
	L			2	1	3
H	VH	6	1			7
	H		4	1		5
	M		2	2		4
	L				6	6
M	VH	4				4
	H	1	2			3
	M		4	6	1	11
	L			2	2	4
L	VH	1				1
	H		3			3
	M		1	3		4
	L			2	12	14
Grand Total		22	22	22	23	89