PGPDSBA Online FEB_A 2021 MRA Project Milestone 1

By.

Pavan Kumar R Naik

PGPDSBA Online

FEB_A 2021

Contents of the presentation:

- Problem Statement

- Data analysis –
  - Info, Shape, and Summary

- EDA – Inferential analysis
  - Outliers and Distribution for Variables
  - Correlation analysis
  - Univariate analysis'
  - Trend
  - Multivariate analysis
  - Country wise and city wise

- Customer Segmentation – RFM analysis
  - Output head in Knime tool
  - Knime workflow
  - Segmentation and inference & recommendations

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.

Data Dictionary:

ORDERNUMBER :	Order Number	CUSTOMERNAME :	customer
QUANTITYORDERED :	Quantity ordered	PHONE :	Phone of the customer
PRICEEACH :	Price of Each item	ADDRESSLINE1 :	Address of customer
ORDERLINENUMBER :	order line	CITY :	City of customer
SALES :	Sales amount	POSTALCODE :	Postal Code of customer
ORDERDATE :	Order Date	COUNTRY :	Country customer
DAYS_SINCE_LASTORDER :	Days_ Since_Lastorder	CONTACTLASTNAME :	Contact person customer
STATUS :	Status of order like Shipped or not	CONTACTFIRSTNAME :	Contact person customer
PRODUCTLINE :	Product line – CATEGORY	DEALSIZE :	Size of the deal based on Quantity and Item Price
MSRP :	Manufacturer's Suggested Retail Price	PRODUCTCODE :	Code of Product

Data Analysis:

The initial analysis is done using the Python in Jupiter notebook using NumPy, Pandas, Matplotlib, and Seaborn packages.

- Info:

  Dataset has 20 variables where 12 are categorical 7 Numerical and 1 date field and there is no missing values are found with a total of 2747 records.

```
<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
```

- **Data Shape:**

  Dataset has 2747 rows and 20 columns (2747, 20)

- **Data Summary:**

	count	mean	std	min	25%	50%	75%	max
ORDERNUMBER	2747.0	10259.761558	91.877521	10100.00	10181.000	10264.00	10334.500	10425.00
QUANTITYORDERED	2747.0	35.103021	9.762135	6.00	27.000	35.00	43.000	97.00
PRICEEACH	2747.0	101.098951	42.042548	26.88	68.745	95.55	127.100	252.87
ORDERLINENUMBER	2747.0	6.491081	4.230544	1.00	3.000	6.00	9.000	18.00
SALES	2747.0	3553.047583	1838.953901	482.13	2204.350	3184.80	4503.095	14082.80
DAYS_SINCE_LASTORDER	2747.0	1757.085912	819.280576	42.00	1077.000	1761.00	2436.500	3562.00
MSRP	2747.0	100.691664	40.114802	33.00	68.000	99.00	124.000	214.00

- Looking at the summary sheet we can assume the sales amount ranges from 482.13$ to 14082.8$ with mean and median tending to equal which meant to be normally distributed
- And even the Quantity ordered data suggests the distribution is normal

- EDA - Distribution:
  Quantity ordered and each price data tell its slightly right-skewed and has few outliers
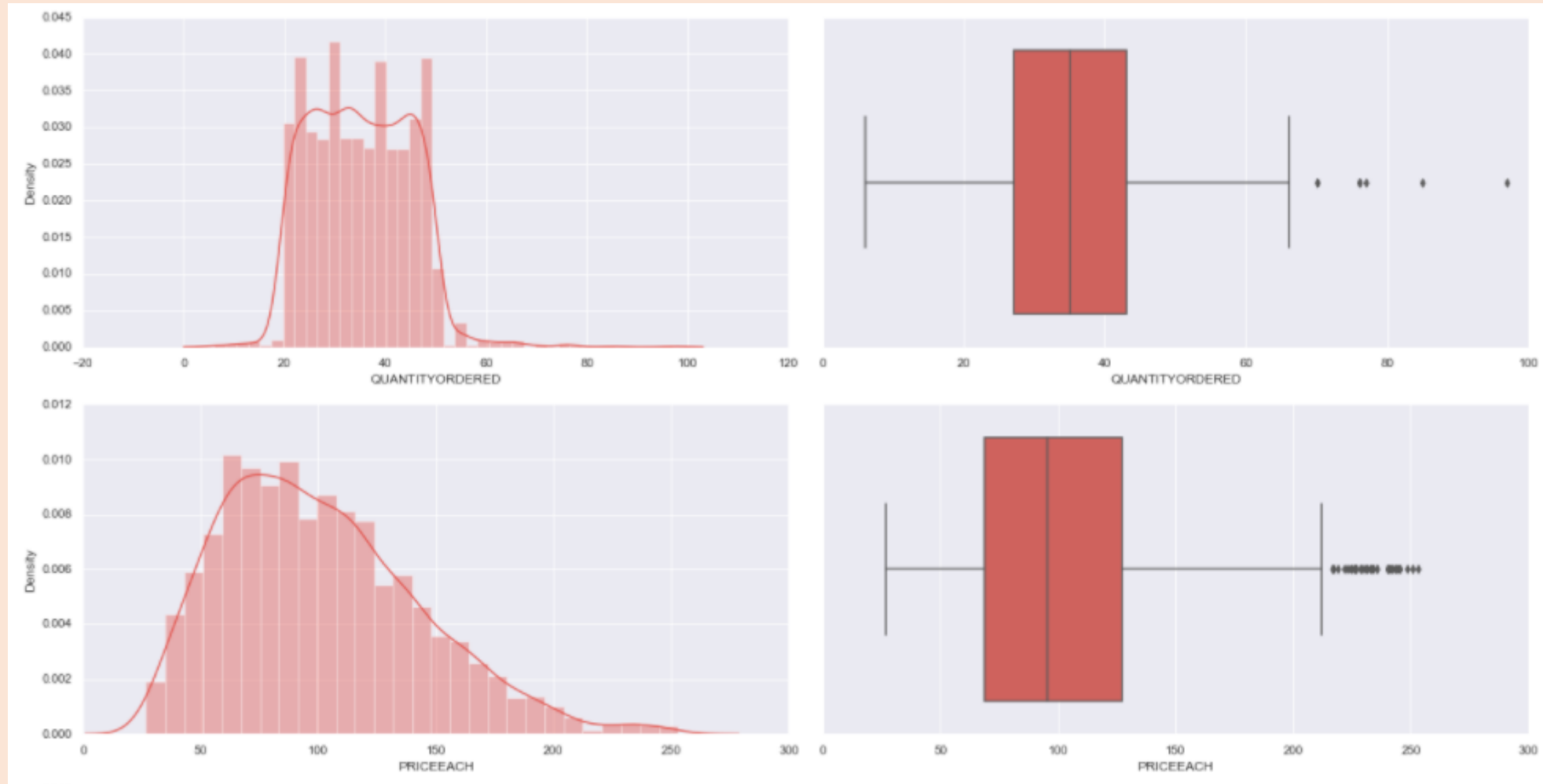
- EDA - Distribution:
  MSRP data suggest normally distributed and Sales amount is right-skewed with multiple outliers
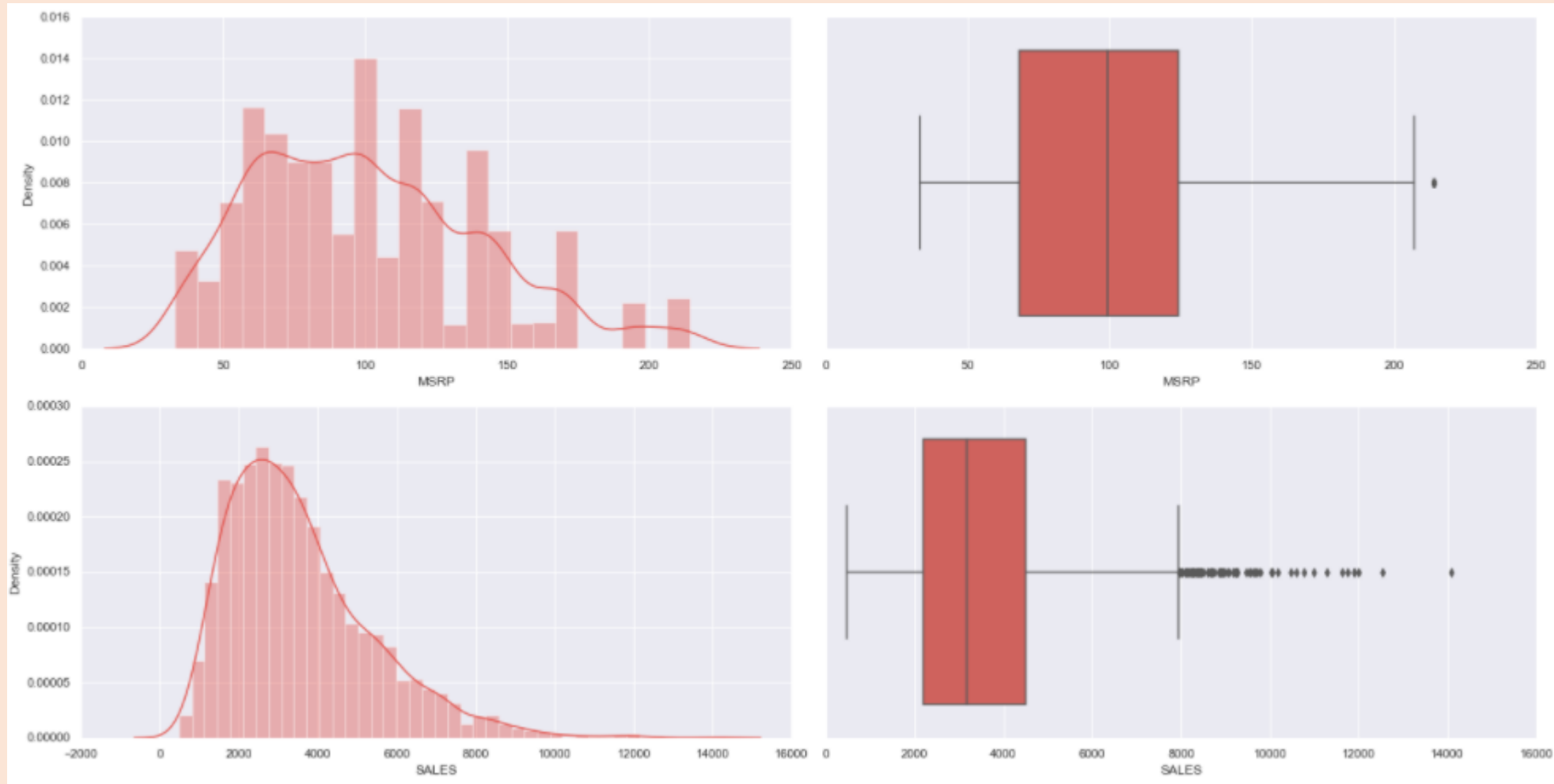
- Correlation:

  Each price and Sales, MSRP and each price, quantity ordered and sales are highly correlated variables
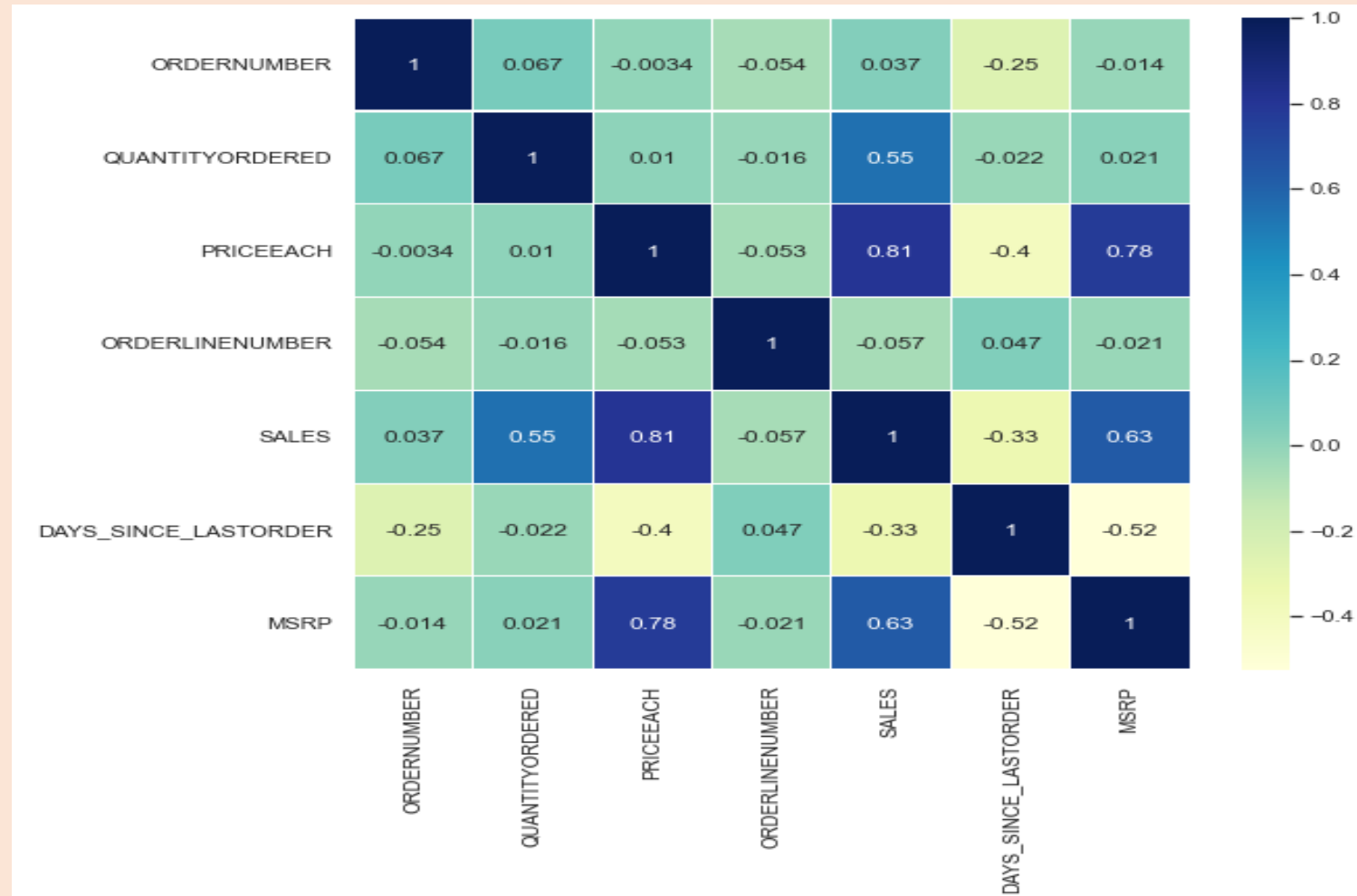
Univariate Analysis:
Highest sales and products ordered have recorded in USA, Spain, and France whereas the least with Ireland
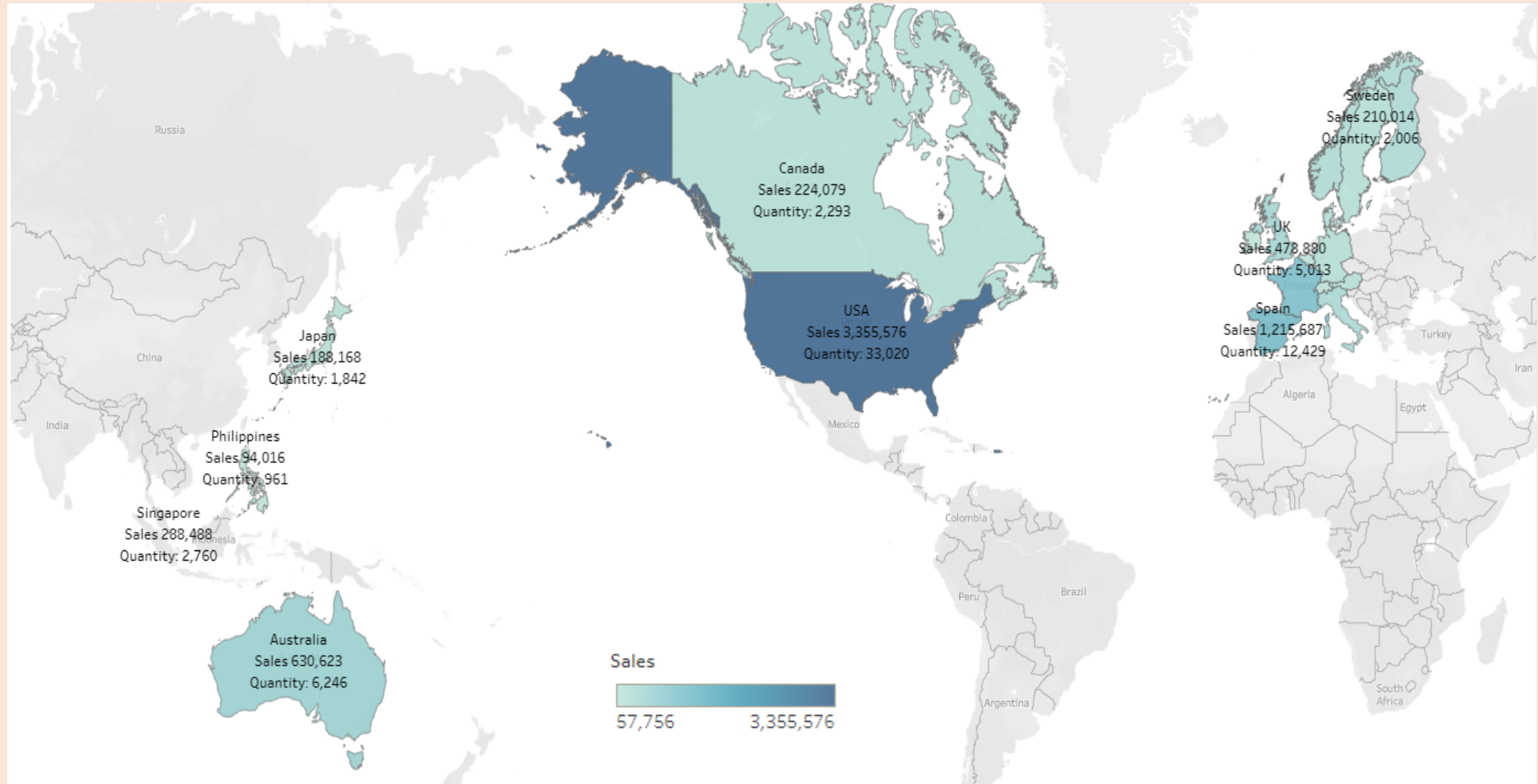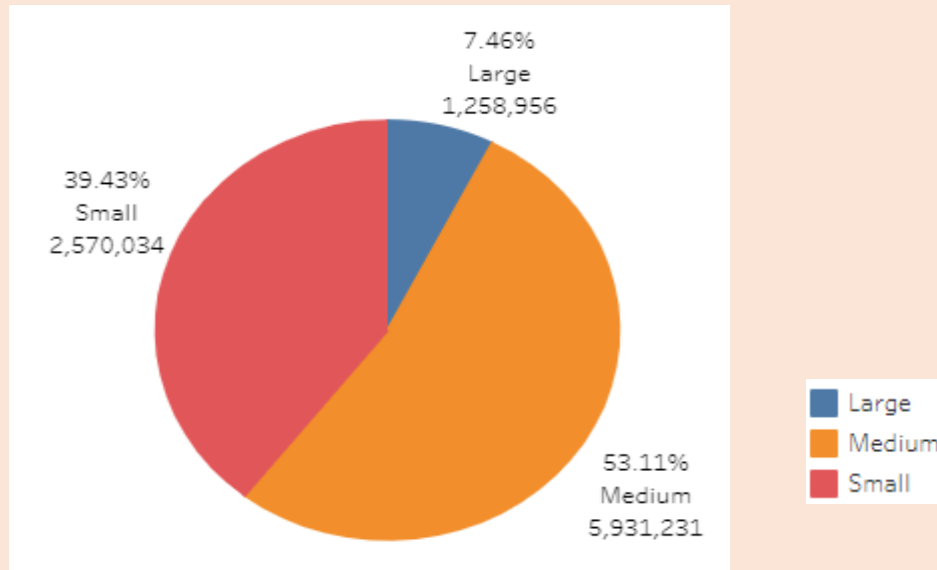
Most sales and products have been ordered in Medium deal size with 53.1% sales out of which Classic cars are having the highest sales.

Median can be measured as a central measure of tendency.
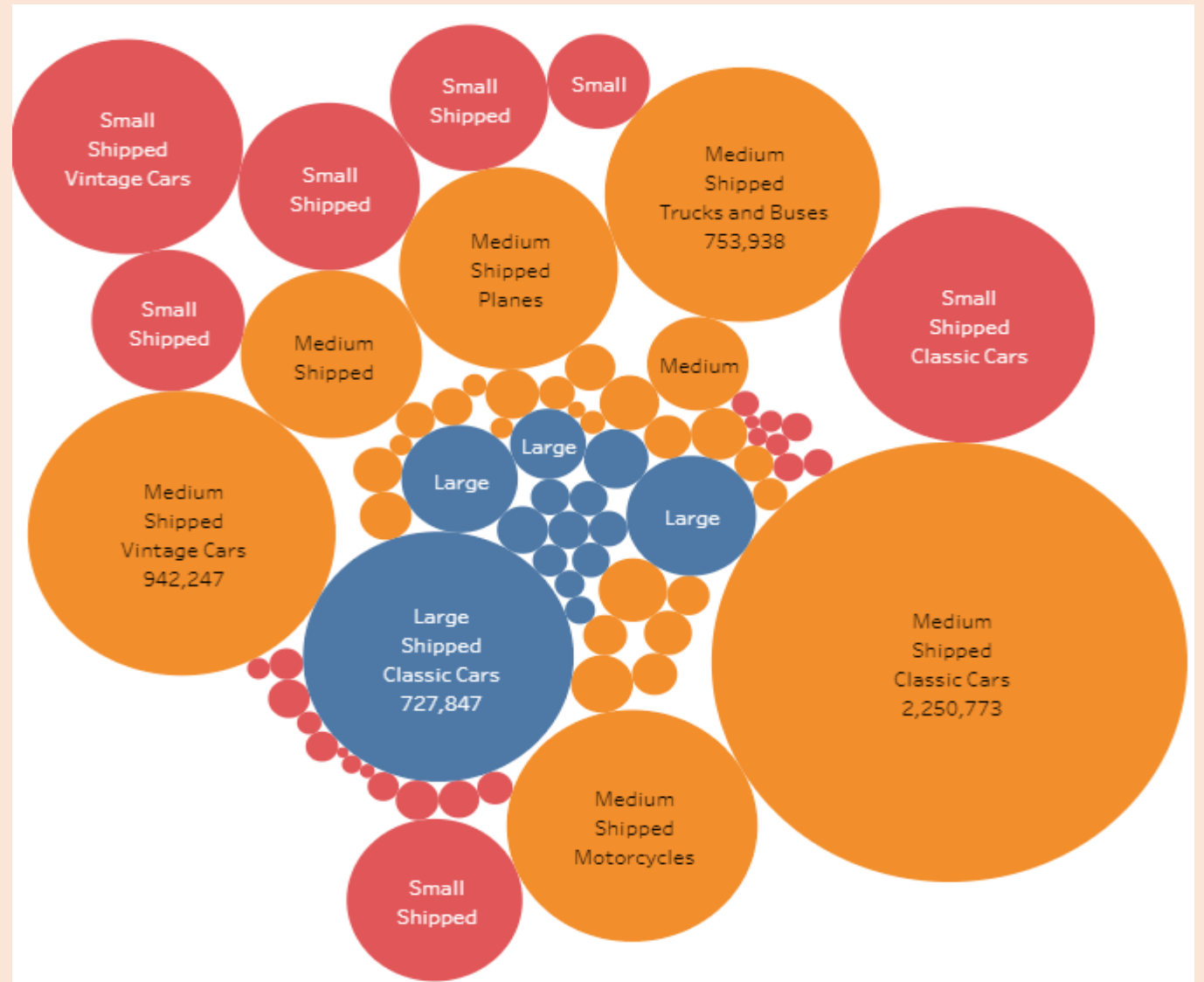
Trend analysis:

- Looking at the trend, we could see that week 44 to 48 has seen the highest sale in the November month Q4.

- Yearly sales were highest in 2019 and gradually reducing

- Sales are in the decreasing trend from Q1 to Q3 whereas Q4 is seen to be in peak
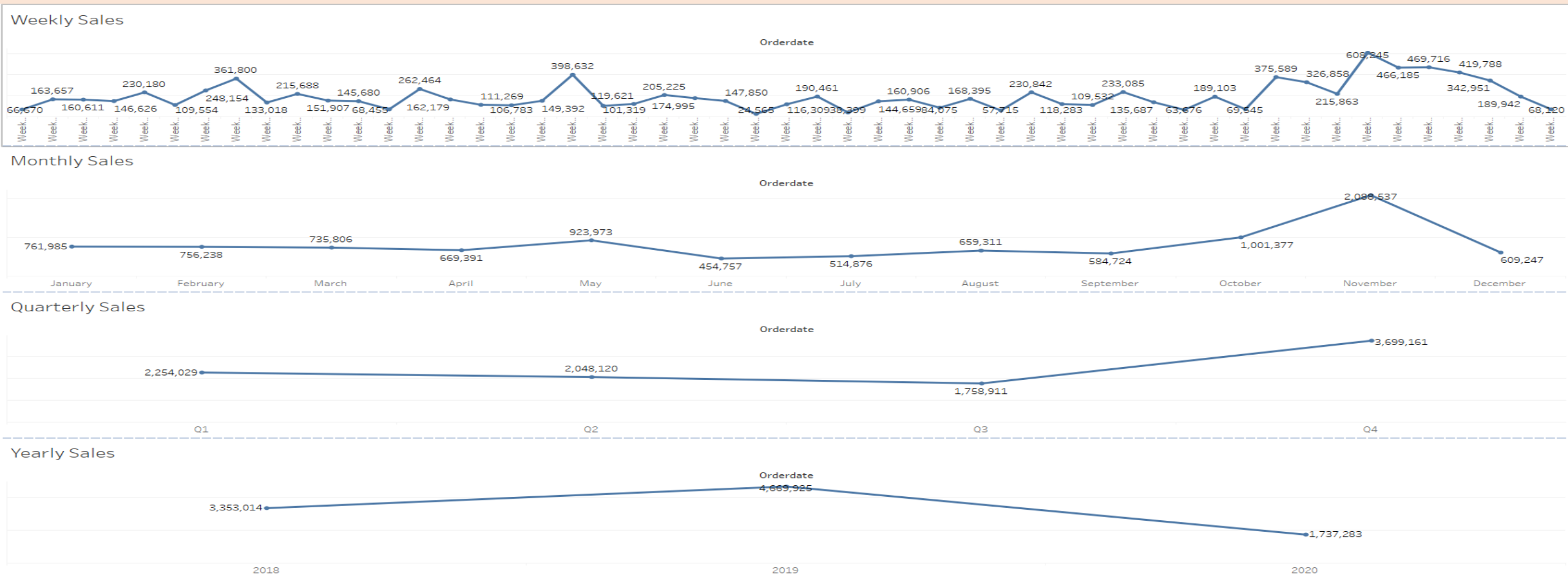
Multivariate Analysis:

Top 10 sales

Product	Sales
S18_3232	284,249
S10_1949	179,815
S12_1108	168,585
S10_4698	158,202
S18_2238	154,624
S12_3891	145,332
S18_1662	139,422
S24_3856	135,859
S18_4027	133,779
S12_1666	130,467
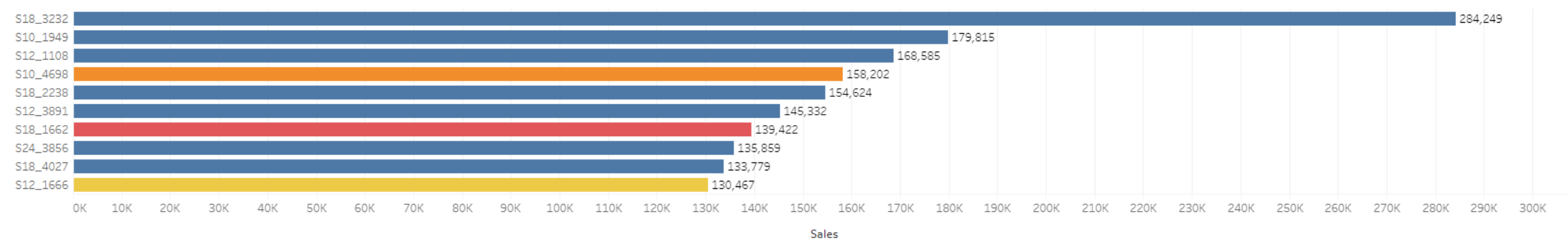
Top 10 purchased products

Product	Quantityordered
S18_3232	1,754
S18_4600	1,031
S700_4002	1,029
S12_4473	1,024
S24_3856	1,017
S24_3949	1,008
S50_1341	999
S24_2300	996
S18_2319	993
S700_2610	990

Legend:
- Classic Cars
- Planes
- Ships
- Trucks and Buses
- Vintage Cars
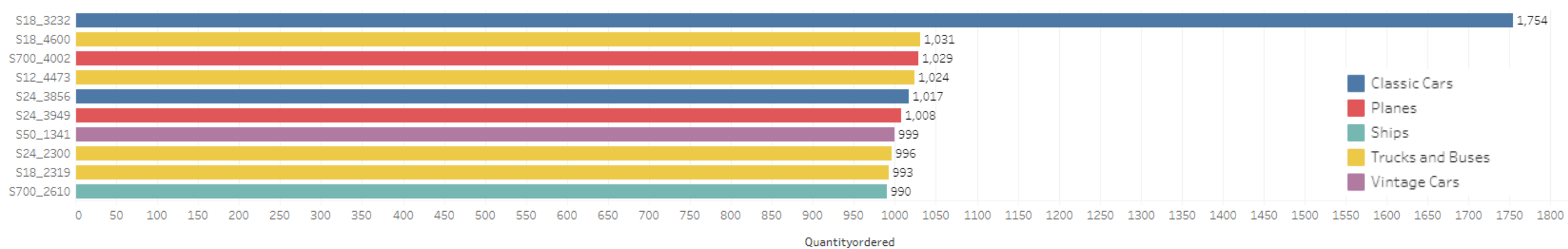
We could see that S18_3232 Classic cars have made the most sales and quantities have been ordered. However, when we look into the top 10 sales Classic cars are the most whereas when we look into the quantity ordered Trucks and buses are the most this defines the cost of the trucks and buses are less.

Multivariate Analysis:

Product line/code vs Sales/quantity
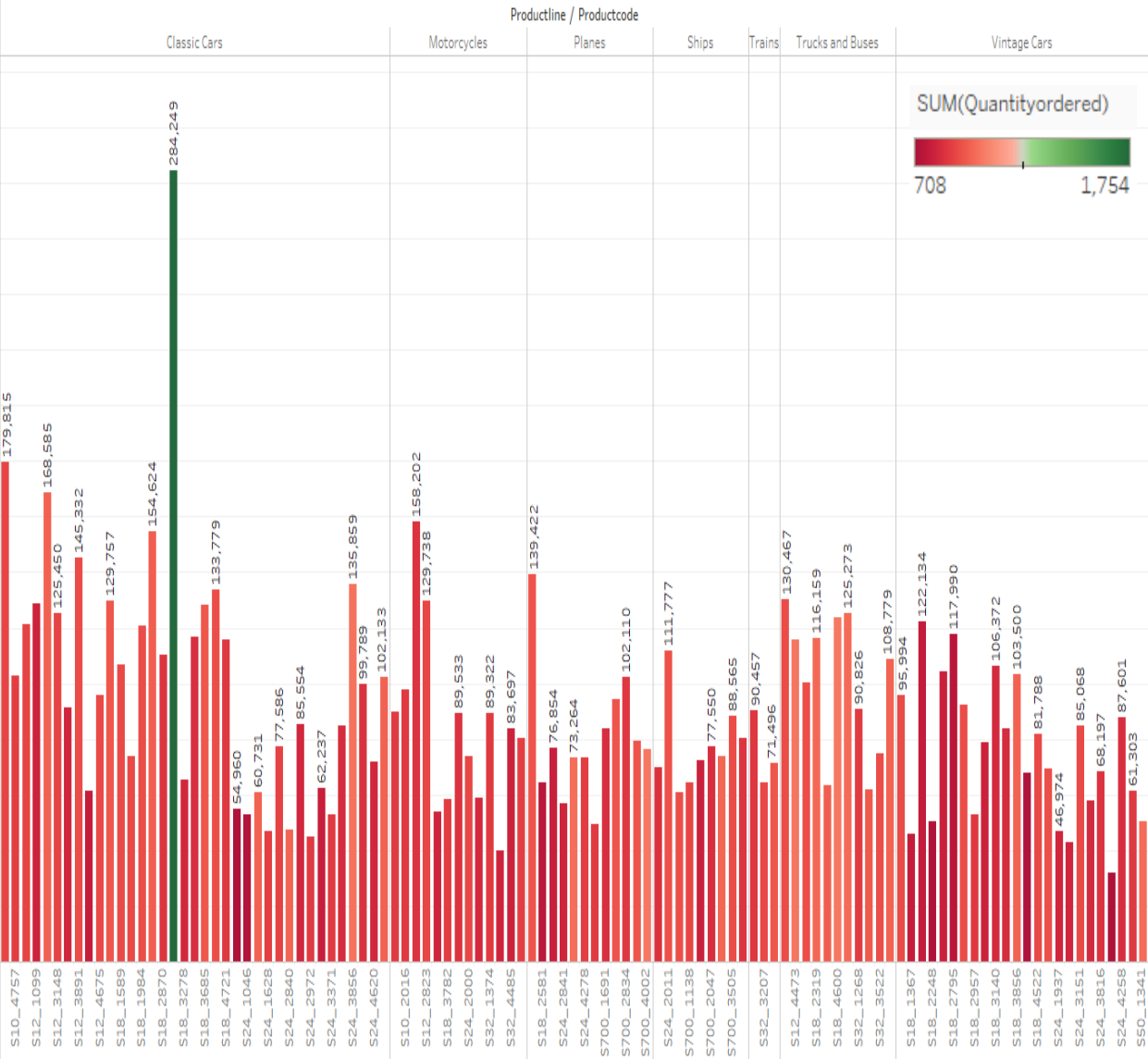
Product line vs Sales/quantity
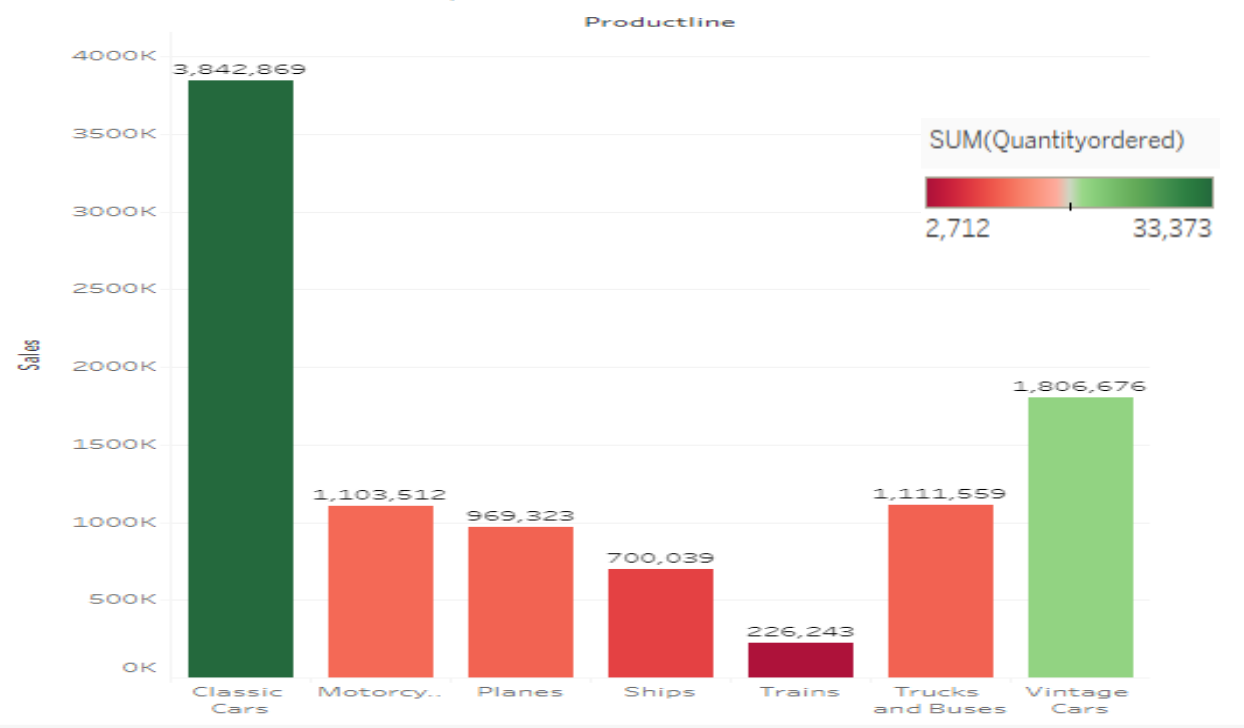
- Classic cars and Vintage cars are the highest quantity ordered with highest sales
- Trains with least quantity ordered with least sales
- Truck and buses, Motorcycle and planes are with below 50 percentile ordered quantitatively
- S18_3232 is one of the specific products with the highest sales and quantity ordered

Top 20 Customer with sales and quantity ordered:

Sales/ Order quantity vs Customer name - Top 20

Customer	Value
Euro Shopping Channel	912,294
Mini Gifts Distributors Ltd.	654,858
Australian Collectors, Co.	200,995
Muscle Machine Inc	197,737
La Rochelle Gifts	180,125
Dragon Souveniers, Ltd.	172,990
Land of Toys Inc.	164,069
The Sharp Gifts Warehouse	160,010
AV Stores, Co.	157,808
Anna's Decorations, Ltd	153,996
Souveniers And Things Co.	151,571
Salzburg Collectables	149,799
Danish Wholesale Imports	145,042
Saveley & Henriot, Co.	142,874
L'ordine Souveniers	142,601
Rovelli Gifts	137,956
Reims Collectables	135,043
Scandinavian Gift Ideas	134,259
Online Diecast Creations C..	131,685
Diecast Classics Inc.	122,138

SUM(Quantityordered)

1,111 9,327

Country wise sales and quantity ordered:

Sales/Quatity Vs Country

USA		France	Australia	UK
Qut: 33,020 Sales: 3,355,576		Qut: 11,090 Sales: 1,110,917	Qut: 6,246 Sales: 630,623	Qut: 5,013 Sales: 478,880

Italy	Singapore	Denmark	Canada	Germany
Qut: 3,773 Sales: 374,674	Qut: 2,760 Sales: 288,488	Qut: 2,197 Sales: 245,637	Qut: 2,293 Sales: 224,079	Qut: 2,148 Sales: 220,472

Finland

Qut: 3,192
Sales: 329,582

Spain	Sweden	Japan	Belgium
Qut: 12,429 Sales: 1,215,687	Qut: 2,006 Sales: 210,014	Qut: 1,842 Sales: 188,168	Qut: 1,074 Sales: 108,413

Norway

Qut: 2,842
Sales: 307,464

Austria

Qut: 1,974
Sales: 202,063

Switzerland

Qut: 1,078
Sales: 117,714

Ireland

Country and city wise sales and quantity ordered:

Sales/Quatity Vs Country/City

USA San Rafael Qut: 6,366 Sales: 654,858	**USA** Brickhaven Qut: 1,654 Sales: 165,255	**USA** San Jose Qut: 1,656 Sales: 160,010	**USA** Boston Qut: 1,628 Sales: 154,070	**France** Paris Qut: 2,521 Sales: 268,945	**France** Reims Qut: 1,433 Sales: 135,043	**France**

- **USA** San Rafael — Qut: 6,366 — Sales: 654,858
- **USA** NYC — Qut: 5,294 — Sales: 560,788
- **USA** New Bedford — Qut: 2,043 — Sales: 207,875
- **USA** Brickhaven — Qut: 1,654 — Sales: 165,255
- **USA** San Jose — Qut: 1,656 — Sales: 160,010
- **USA** Boston — Qut: 1,628 — Sales: 154,070
- **USA** Philadelphia — Qut: 1,398 — Sales: 151,189
- **USA** Cambridge — Qut: 1,333 — Sales: 139,244
- **USA** Nashua — Qut: 1,248 — Sales: 131,685
- **USA** Allentown — Qut: 1,111 — Sales: 122,138
- **USA** Burlingame — Qut: 1,179 — Sales: 120,783
- **USA** San Diego — Qut: 954 — Sales: 87,489
- **USA** White Plains — Qut: 929 — Sales: 85,556
- **USA** Newark — Qut: 906 — Sales: 83,228
- **USA** Pasadena — Qut: 1,060 — Sales: 104,562
- **USA** Las Vegas — Qut: 929
- **USA**
- **USA**
- **USA** Bridgewater — Qut: 903 — Sales: 101,895
- **USA** New Haven — Qut: 636
- **USA**

- **France** Paris — Qut: 2,521 — Sales: 268,945
- **France** Nantes — Qut: 2,102 — Sales: 204,305
- **France** Lyon — Qut: 1,428
- **France** Reims — Qut: 1,433 — Sales: 135,043
- **France** Marseille — Qut: 804 — Sales:
- **France** Toulouse — Qut: 687 — Sales:
- **France** Lille — Qut: 699 — Sales:
- **France**
- **France**

- **Australia** Melbourne — Qut: 1,926 — Sales: 200,995
- **Australia** Chatswood — Qut: 1,601 — Sales: 151,571
- **Australia**
- **Australia** North Sydney — Qut: 1,469 — Sales: 153,996

- **UK**
- **UK** London — Qut: 1,294 — Sales:
- **UK** Liverpool — Qut: 1,046
- **UK** Cowes

- **Italy** Reggio Emilia — Qut: 1,280
- **Italy** Bergamo — Qut: 1,650
- **Italy** Torino — Qut: 843 — Sales: 94,117

- **Singapore** Singapore — Qut: 2,760 — Sales: 288,488

- **Denmark** Kobenhavn — Qut: 1,315 — Sales: 145,042
- **Denmark** Aaarhus — Qut: 882 — Sales: 100,596

- **Canada** Montreal — Qut: 717

- **Germany** Koln — Qut: 936 — Sales:
- **Germany** Frankfurt — Qut: 811 — Sales: 85,172
- **Germany**

- **Finland** Espoo — Qut: 1,031
- **Finland** Helsinki — Qut: 1,051
- **Finland** Oulu — Qut: 1,110 — Sales: 104,370

- **Norway** Stavern — Qut: 1,082
- **Norway** Bergen

- **Sweden** Boras — Qut: 1,359 — Sales: 134,259
- **Sweden** Lule — Qut: 647 — Sales: 75,755

- **Japan** Minato-ku — Qut: 1,150 — Sales: 120,563
- **Japan**

- **Austria** Salzburg — Qut: 1,442 — Sales: 149,799

- **Switzerland** Gensve — Qut: 1,078

- **Ireland**

- **Spain** Madrid — Qut: 10,958 — Sales: 1,082,551
- **Spain**
- **Spain** Sevilla — Qut: 589

Inferences:

- Euro shopping channel is the customer with high sales

- Classic cars are the most ordered and highest sold product

- The USA has the highest sales and quantity ordered when compared to countries

- City Madrid in Spain has the highest sales and quantity ordered when compared to cities

- S18_3232 Classic cars have made the most sales and quantities have been ordered

- Trains with least quantity ordered with least sales with 2712 quantity ordered with least unit price 6476

- The sales of large size deals remain stagnant over the years and it can be presumed that the company should focus on getting large size projects

- The company is customer-driven as the major chunk of sales comes from 5 customers. So the company should focus on customer scouting in a rational way because if there is a customer churn among these top then it will impact the sales in huge manner

The EDA and visualization is done using the Tableau tool

https://public.tableau.com/app/profile/pavan.kumar1123/viz/Milestone1_16392526679750/ProductlinevsSalesquantity?publish=yes

Customer Segmentation – RFM analysis:

- KNIME Tool has been used to do the analysis
- DAYS_SINCE_LASTORDER, ORDERNUMBER, and SALES are the parameters used to get RFM
- DAYS_SINCE_LASTORDER (Min) – Recency
- ORDERNUMBER(Unique Count) – Frequency
- Sales(Sum) – Monetary

Segmentation:

		Monetary		
Recency	Frequency	H	M	L
H	H M L			
M	H M L			
L	H M L			

Output head– RFM analysis:

Row ID	I PRODU...	S PHONE	S ADDRESSL...	S CITY	S POSTA...	S COUNTRY	S CONTA...	S CONTA...	S DEALSIZE	S ORDER...	S SALES [...	S DAYS_...	S Recency	S Freque...	S Monetary
Row9	3	(1) 47.55.6555	25, rue Lauriston	Paris	75016	France	Perrier	Dominique	Medium	Bin 1	Bin 2	Bin 1	H	L	M
Row88	5	(198) 555-8888	Garden House ...	Cowes	PO31 7PJ	UK	Bennett	Helen	Small	Bin 1	Bin 2	Bin 2	M	L	M
Row87	3	3105553722	3675 Furth Circle	Burbank	94019	USA	Thompson	Steve	Medium	Bin 1	Bin 1	Bin 3	L	L	L
Row86	9	0921-12 3555	Berguvsv□gen 8	Lule	S-958 22	Sweden	Berglund	Christina	Medium	Bin 3	Bin 1	Bin 2	M	H	L
Row85	5	2125551500	2678 Kingston Rd.	NYC	10022	USA	Frick	Michael	Small	Bin 1	Bin 2	Bin 2	M	L	M
Row84	7	0897-034555	Grenzacherweg...	Gensve	1203	Switzerland	Holz	Michael	Medium	Bin 1	Bin 3	Bin 2	M	L	H
Row83	9	(171) 555-2282	Berkeley Garde...	Liverpool	WX1 6LT	UK	Devon	Elizabeth	Medium	Bin 1	Bin 3	Bin 1	H	L	H
Row82	9	6265557265	78934 Hillside Dr.	Pasadena	90003	USA	Young	Julie	Medium	Bin 1	Bin 2	Bin 3	L	L	M
Row81	0	90-224 8555	Keskuskatu 45	Helsinki	21240	Finland	Karttunen	Matti	Medium	Bin 1	Bin 3	Bin 1	H	L	H
Row80	7	0221-5554327	Mehrheimerstr. ...	Koln	50739	Germany	Pfalzheim	Henriette	Medium	Bin 1	Bin 2	Bin 2	M	L	M
Row8	7	30.59.8555	67, avenue de l...	Versailles	78000	France	Tonini	Daniel	Medium	Bin 1	Bin 1	Bin 1	H	L	L
Row79	2	+81 3 3584 0...	2-2-8 Roppongi	Minato-ku	106-0032	Japan	Shimamura	Akiko	Small	Bin 3	Bin 3	Bin 1	H	H	H
Row78	9	4085553659	3086 Ingle Ln.	San Jose	94217	USA	Frick	Sue	Small	Bin 3	Bin 3	Bin 1	H	H	H
Row77	1	2015559350	7476 Moss Rd.	Newark	94019	USA	Brown	William	Medium	Bin 1	Bin 2	Bin 2	M	L	M
Row76	0	6505556809	9408 Furth Circle	Burlingame	94217	USA	Hirano	Juri	Medium	Bin 3	Bin 3	Bin 1	H	H	H
Row75	7	2035559545	567 North Pend...	New Haven	97823	USA	Murphy	Leslie	Medium	Bin 1	Bin 2	Bin 2	M	L	M
Row74	0	+358 9 8045 ...	Software Engin...	Espoo	FIN-02271	Finland	Suominen	Kalle	Medium	Bin 1	Bin 3	Bin 2	M	L	H
Row73	5	(171) 555-0297	35 King George	London	WX3 6FW	UK	Brown	Ann	Medium	Bin 1	Bin 2	Bin 3	L	L	M
Row72	4	+61 2 9495 8...	Monitor Money ...	Chatswood	2067	Australia	Huxley	Adrian	Medium	Bin 3	Bin 3	Bin 1	H	H	H
Row71	4	7025551838	8489 Strong St.	Las Vegas	83030	USA	King	Sue	Small	Bin 1	Bin 2	Bin 3	L	L	M
Row70	5	4155554312	2793 Furth Circle	Brisbane	94217	USA	Taylor	Sue	Small	Bin 1	Bin 1	Bin 3	L	L	L
Row7	5	61-7-3844-6555	31 Duncan St. ...	South Brisbane	4101	Australia	Calaghan	Tony	Medium	Bin 1	Bin 1	Bin 1	H	L	L
Row69	3	0695-34 6555	?kergatan 24	Boras	S-844 67	Sweden	Larsson	Maria	Medium	Bin 1	Bin 3	Bin 1	H	L	H
Row68	1	78.32.5555	2, rue du Comm...	Lyon	69004	France	Saveley	Mary	Medium	Bin 1	Bin 3	Bin 3	L	L	H
Row67	5	6562-9555	Geislweg 14	Salzburg	5020	Austria	Pipps	Georg	Medium	Bin 3	Bin 3	Bin 1	H	H	H
Row66		(071) 23 67 2...	Boulevard Tirou...	Charleroi	B-6000	Belgium	Cartrain	Pascale	Medium	Bin 3	Bin 1	Bin 3	L	H	L
Row65	5	(604) 555-4555	23 Tsawassen ...	Tsawassen	T2F 8M4	Canada	Lincoln	Elizabeth	Small	Bin 1	Bin 1	Bin 2	M	L	L
Row64	4	035-640555	Via Ludovico il ...	Bergamo	24100	Italy	Rovelli	Giovanni	Small	Bin 1	Bin 3	Bin 3	L	L	H
Row63	7	26. 47. 1555	59 rue de l'Abb...	Reims	51100	France	Henriot	Paul	Small	Bin 3	Bin 3	Bin 1	H	H	H
Row62	2	(514) 555-8054	43 rue St. Laurent	Montreal	H1J 1C3	Canada	Fresnisre	Jean	Medium	Bin 1	Bin 1	Bin 1	H	L	L
Row61	0	(02) 5554 67	Rue Joseph-Be...	Bruxelles	B-1180	Belgium	Dewey	Catherine	Small	Bin 1	Bin 1	Bin 2	M	L	L
Row60	1	981-443655	Torikatu 38	Oulu	90110	Finland	Koskitalo	Pirkko	Medium	Bin 1	Bin 2	Bin 1	H	L	M
Row6	0	03 9520 4555	636 St Kilda Road	Melbourne	3004	Australia	Ferguson	Peter	Medium	Bin 3	Bin 3	Bin 1	H	H	H
Row59	0	+81 06 6342 ...	Dojima Avanza ...	Osaka	530-0003	Japan	Kentary	Mory	Small	Bin 1	Bin 1	Bin 3	L	L	L
Row58	5	6175557555	7635 Spinnaker...	Brickhaven	58339	USA	Barajas	Miguel	Medium	Bin 1	Bin 1	Bin 2	M	L	L
Row57	4	6035558647	2304 Long Airp...	Nashua	62005	USA	Young	Valarie	Medium	Bin 1	Bin 3	Bin 1	H	L	H
Row56	4	+47 2212 1555	Drammensveien...	Oslo	N 0106	Norway	Klaeboe	Jan	Small	Bin 1	Bin 2	Bin 3	L	L	M

KNIME Workflow– RFM analysis:
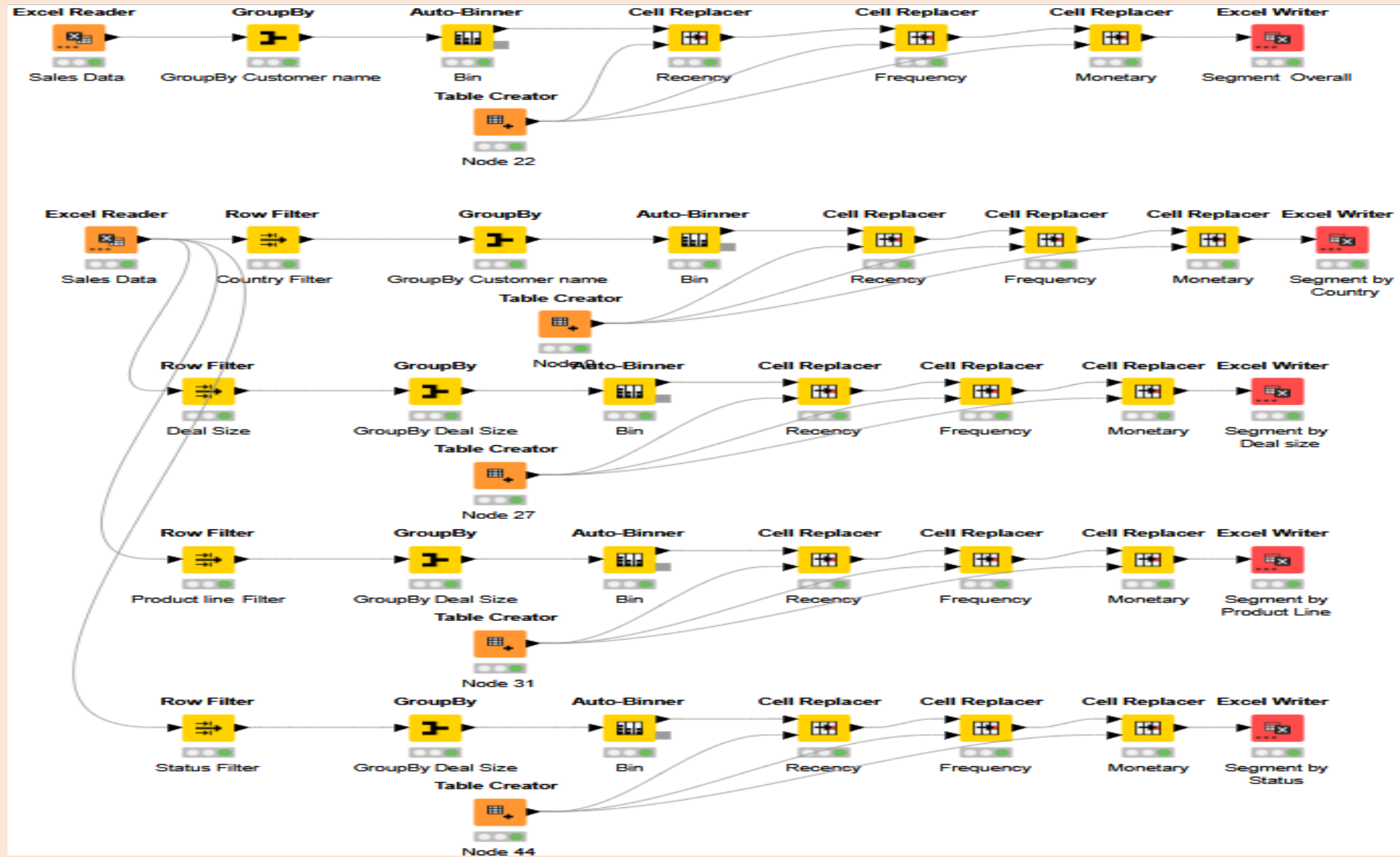
Inference from RFM analysis:

- Overall –H - above 67 %, M - 33.3< and 67% > , L - <33.3%

Gold Customer name	Silver Customer name	Bronze Customer name
Anna's Decorations, Ltd	Auto Canal Petit	Australian Gift Network, Co
Australian Collectors, Co.	Collectables For Less Inc.	Auto Assoc. & Cie.
Baane Mini Imports	FunGiftIdeas.com	Quebec Home Shopping Network
Diecast Classics Inc.	Gift Depot Inc.	
Euro Shopping Channel	Gifts4AllAges.com	
La Rochelle Gifts	Lyon Souveniers	
Land of Toys Inc.	Marta's Replicas Co.	
Mini Gifts Distributors Ltd.	Mini Classics	
Reims Collectables	Oulu Toy Supplies, Inc.	
Salzburg Collectables		
Souveniers And Things Co.		
Technics Stores Inc.		
The Sharp Gifts Warehouse		
Tokyo Collectables, Ltd		

- These 14 customers are given the scores on HHH – RFM
- Company needs to make sure in retain these customers by giving extra benefits like gift hampers/discounts (Best Customers)

- 9 Customers are considered to be the silver customer who requires attention in focusing to improve to the next level
- These customers are with HMM- RFM scores (Loyal Customers)

- 3 customers are considered to be the bronze customer with HLL who are very frequently visiting the store with low monetary and low purchasing who needs a special attention

Inference from RFM analysis:

- Overall –H - above 67 %, M - 33.3< and 67% > , L - <33.3%

Ignore	Needs Focus	MHH
Alpha Cognac	Amica Models & Co.	Danish Wholesale Imports
Australian Collectables, Ltd	Classic Legends Inc.	Handji Gifts& Co
Auto-Moto Classics Inc.	giftsbymail.co.uk	Muscle Machine Inc
Bavarian Collectables Imports, Co.	Heintze Collectables	
CAF Imports	La Corne D'abondance, Co.	
Clover Collections, Co.	Mini Caravy	
Daedalus Designs Imports	Motor Mint Distributors Inc.	
Diecast Collectables	Super Scale Inc.	
Double Decker Gift Stores, Ltd	Tekni Collectables Inc.	
Gift Ideas Corp.	Toms Spezialitten, Ltd	
Iberia Gift Imports, Corp.	Vitachrome Inc.	
Marseille Mini Autos		
Mini Auto Werke		
Osaka Souveniers Co.		
Signal Collectibles Ltd.		
West Coast Collectables Co.		

- These are 17 customers who are waste for the company with LLL – RFM scores (Lost Customers)
- Who needs to be completely ignored without wasting any company resource

- 3 customers are with MHH value who require special attention who give more sales and more frequency rate with medium recency rate

- 11 customers with MMM scores need to figure out how can we increase all the rates by taking customer feedback (Churning out)

Inference from RFM analysis:

- Overall –H - above 67 %, M - 33.3< and 67% > , L - <33.3%

- Similar analysis can be done using the filtering option for Country-wise, Deal size, product line and status

HHH - USA ▼
Diecast Classics Inc.
Land of Toys Inc.
Mini Gifts Distributors Ltd.
Technics Stores Inc.
The Sharp Gifts Warehouse

HHH - Classic cars ▼
Anna's Decorations, Ltd
Euro Shopping Channel
Mini Gifts Distributors Ltd.
Salzburg Collectables
Suominen Souveniers

HHH - Large ▼
Diecast Classics Inc.
Euro Shopping Channel
Mini Gifts Distributors Ltd.
Saveley & Henriot, Co.
The Sharp Gifts Warehouse
Tokyo Collectables, Ltd

- The customers with LLL RFM scores to be ignored and not utilize any company resource to retain

Thank you.
Pavan Kumar