Marketing and Retail Analytics

Project - Milestone 1

Agenda & Executive Summary of the data

- **Contents**
  * Data set and problem statement
  * Exploratory Data Analysis and Inferences
  * Customer Segmentation using RFM analysis
  * Inferences from RFM analysis and 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

About data set

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2747 entries, 0 to 2746
Data columns (total 21 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    object
 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
 20  CONTACT NAME          2747 non-null    object
dtypes: float64(2), int64(5), object(14)
memory usage: 450.8+ KB
```

- The given dataset contains 2747 rows and 21 columns of data.
- It has two float data type, five integer data type and 14 object data type of columns.
- The dataset has both no null values and no duplicate rows of data
- There is presence of outliers in variables such as Quantityordered,Price and Sales.
- Variable 'Sales' has highest positive skewness(0.784) and Variable 'Days_since_lastorder' has lowest negative skewness(-0.002)
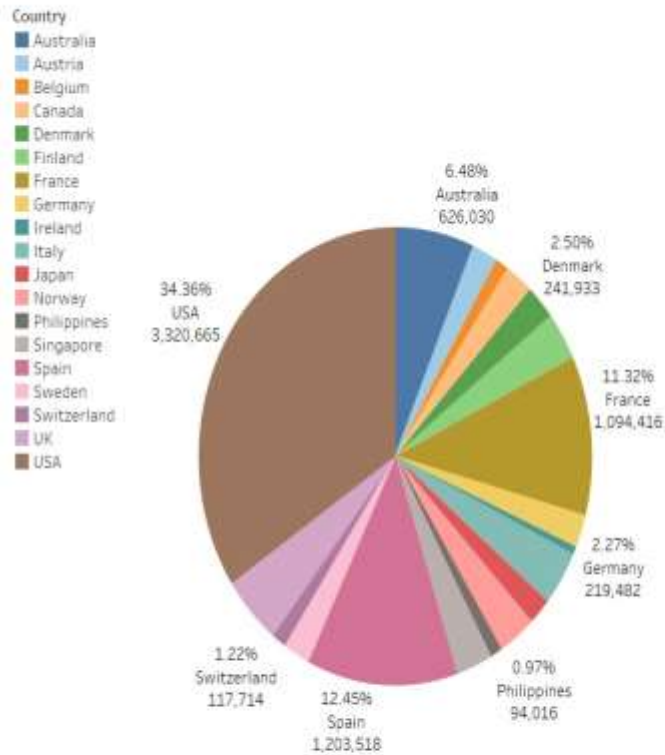
	ORDERNUMBER	QUANTITYORDERED	PRICEEACH	ORDERLINENUMBER	SALES	ORDERDATE	DAYS_SINCE_LASTORDER
count	2747.000000	2747.000000	2747.000000	2747.000000	2747.000000	2747	2747.000000
unique	NaN	NaN	NaN	NaN	NaN	246	NaN
top	NaN	NaN	NaN	NaN	NaN	14-11-2018	NaN
freq	NaN	NaN	NaN	NaN	NaN	38	NaN
mean	10259.761558	35.103021	101.098952	6.491081	3553.047583	NaN	1757.085912
std	91.877521	9.762135	42.042549	4.230544	1838.953901	NaN	819.280576
min	10100.000000	6.000000	26.880000	1.000000	482.130000	NaN	42.000000
25%	10181.000000	27.000000	68.745000	3.000000	2204.350000	NaN	1077.000000
50%	10264.000000	35.000000	95.550000	6.000000	3184.800000	NaN	1761.000000
75%	10334.500000	43.000000	127.100000	9.000000	4503.095000	NaN	2436.500000
max	10425.000000	97.000000	252.870000	18.000000	14082.800000	NaN	3562.000000

	STATUS	PRODUCTLINE	MSRP	CUSTOMERNAME	CITY	COUNTRY	DEALSIZE	CONTACT NAME
count	2747	2747	2747.000000	2747	2747	2747	2747	2747
unique	6	7	NaN	89	71	19	3	89
top	Shipped	Classic Cars	NaN	Euro Shopping Channel	Madrid	USA	Medium	Diego Freyre
freq	2541	949	NaN	259	304	928	1349	259
mean	NaN	NaN	100.691664	NaN	NaN	NaN	NaN	NaN
std	NaN	NaN	40.114802	NaN	NaN	NaN	NaN	NaN
min	NaN	NaN	33.000000	NaN	NaN	NaN	NaN	NaN
25%	NaN	NaN	68.000000	NaN	NaN	NaN	NaN	NaN
50%	NaN	NaN	99.000000	NaN	NaN	NaN	NaN	NaN
75%	NaN	NaN	124.000000	NaN	NaN	NaN	NaN	NaN
max	NaN	NaN	214.000000	NaN	NaN	NaN	NaN	NaN

- **Maximum** and **Minimum** values for respective numerical variables are highlighted in **Green** and **Red** colors
- **Top** and **Freq** for respective string variables are highlighted in **Orange** color
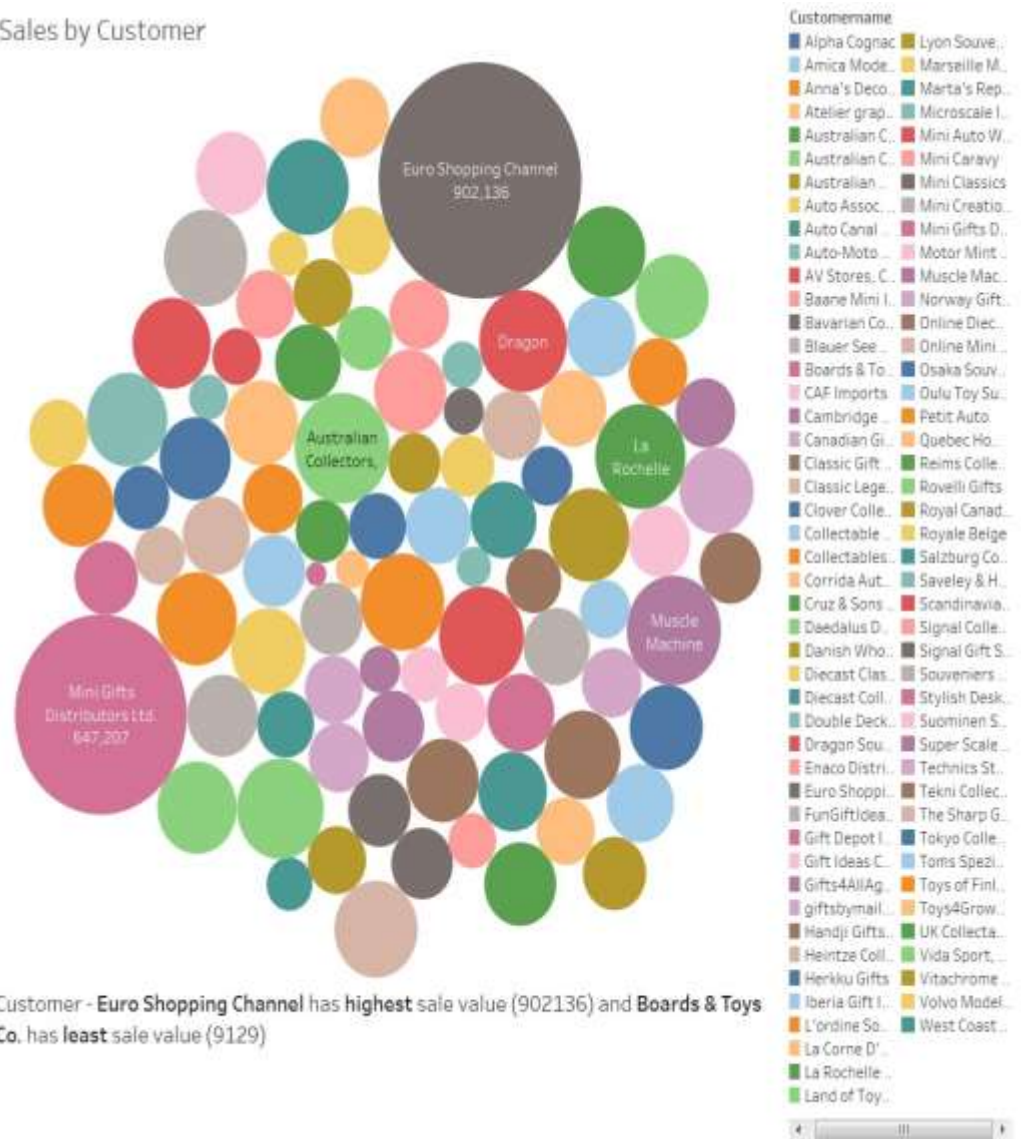
Exploratory Analysis and Inferences
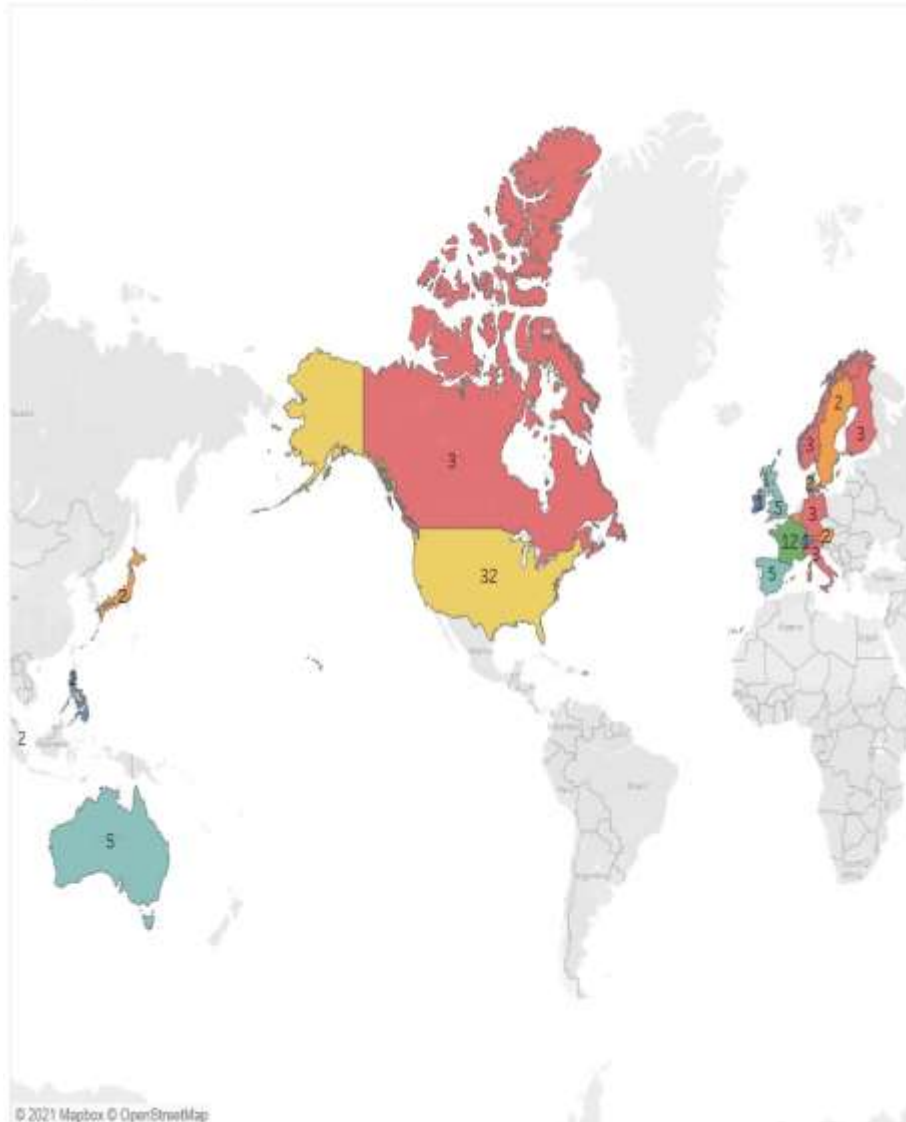
Sales by country

Sales by Customer

Customername

Country
- Australia
- Austria
- Belgium
- Canada
- Denmark
- Finland
- France
- Germany
- Ireland
- Italy
- Japan
- Norway
- Philippines
- Singapore
- Spain
- Sweden
- Switzerland
- UK
- USA

34.36%
USA
3,320,665

6.48%
Australia
626,030

2.50%
Denmark
241,933

11.32%
France
1,094,416

2.27%
Germany
219,482

0.97%
Philippines
94,016

12.45%
Spain
1,203,518

1.22%
Switzerland
117,714

Euro Shopping Channel
902,136

Dragon

Australian
Collectors,

La
Rochelle

Muscle
Machine

Mini Gifts
Distributors Ltd.
647,207

Customer - **Euro Shopping Channel** has **highest** sale value (902136) and **Boards & Toys Co.** has **least** sale value (9129)

Alpha Cognac — Lyon Souve..
Amica Mode.. — Marseille M..
Anna's Deco.. — Marta's Rep..
Atelier grap.. — Microscale I..
Australian C.. — Mini Auto W..
Australian C.. — Mini Caravy
Australian .. — Mini Classics
Auto Assoc.. — Mini Creatio..
Auto Canal .. — Mini Gifts D..
Auto-Moto .. — Motor Mint ..
AV Stores. C.. — Muscle Mac..
Baane Mini I.. — Norway Gift..
Bavarian Co.. — Online Diec..
Blauer See .. — Online Mini ..
Boards & To.. — Osaka Souv ..
CAF Imports — Oulu Toy Su..
Cambridge .. — Petit Auto
Canadian Gi.. — Quebec Ho..
Classic Gift .. — Reims Colle..
Classic Lege.. — Rovelli Gifts
Clover Colle.. — Royal Canad..
Collectable .. — Royale Beige
Collectables.. — Salzburg Co..
Corrida Aut.. — Saveley & H..
Cruz & Sons .. — Scandinavia..
Daedalus D .. — Signal Colle..
Danish Who.. — Signal Gift S..
Diecast Clas.. — Souveniers
Diecast Coll.. — Stylish Desk..
Double Deck.. — Suominen S..
Dragon Sou.. — Super Scale ..
Enaco Distri.. — Technics St..
Euro Shoppi.. — Tekni Collec..
FunGiftIdea.. — The Sharp G..
Gift Depot I .. — Tokyo Colle..
Gift Ideas C.. — Toms Spezi..
Gifts4AllAg.. — Toys of Finl..
giftsbymail.. — Toys4Grow..
Handji Gifts.. — UK Collecta..
Heintze Coll.. — Vida Sport, ..
Herkku Gifts — Vitachrome ..
Iberia Gift I.. — Volvo Model..
L'ordine So.. — West Coast ..
La Corne D' ..
La Rochelle ..
Land of Toy..

Country -**USA** has the **Highest** sales(3320665) and **Ireland** has the **least** sales(57450)
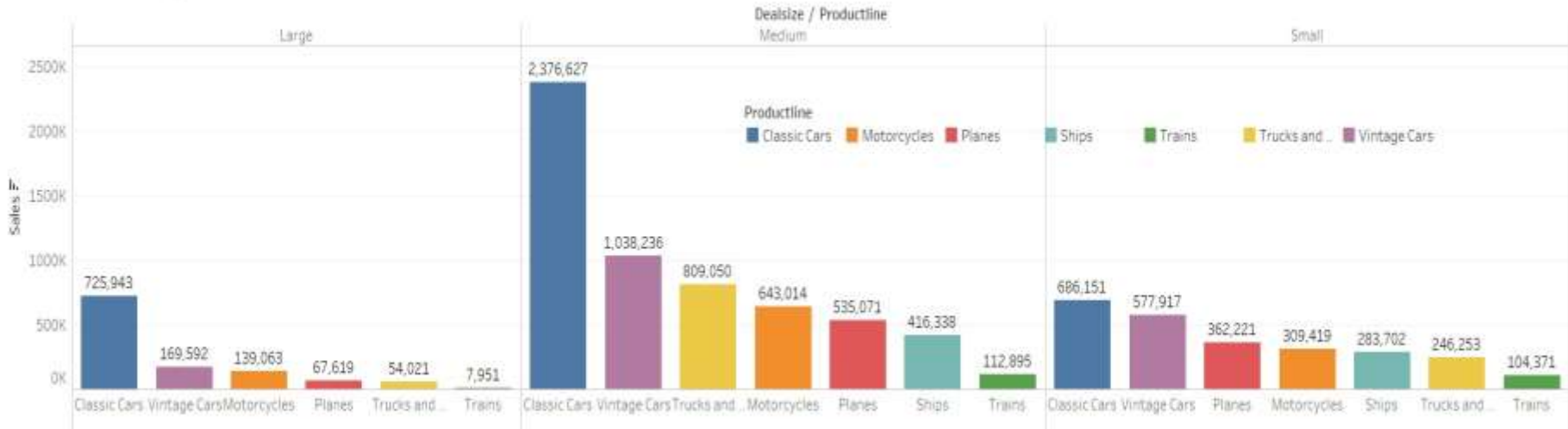
Country wise customer

Country wise ordered qty

Quantityordered
490 — 32,974

USA has the **Highest** count of customers(32) and **Ireland** has the **least** count of customers(1)

© 2021 Mapbox © OpenStreetMap
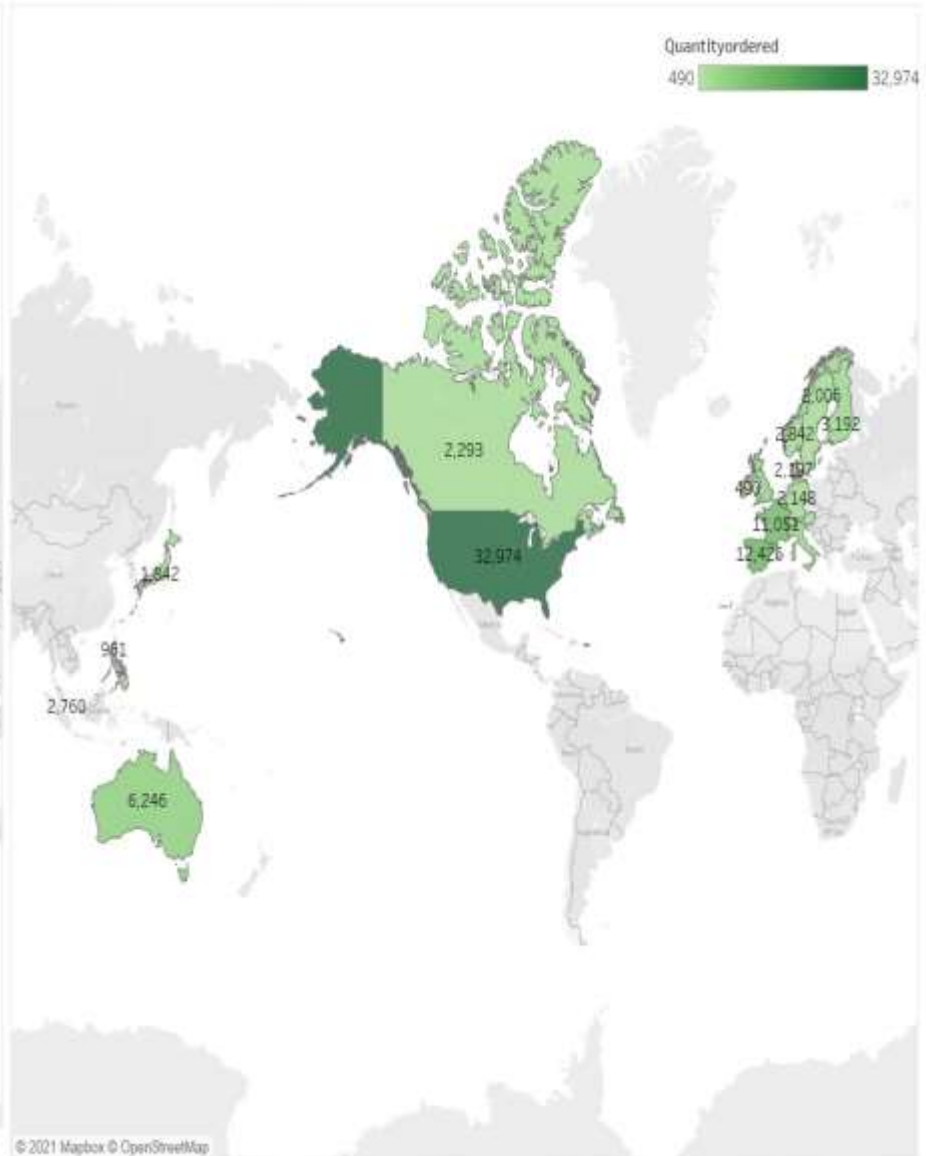
USA has the **Highest** ordered quantity(32974) and **Ireland** has the **least** ordered quantity(490)

Sales by Deal size/Productline

Deal size - **'Medium'** and Product line -**'Classic Cars'** has the **Highest** sales amount(2376627) and Deal size - **'Large'** and Product line -**'Ships'** has the **zero** sales amount(0)

# Ordered qtys by Deal size/Productline

Deal size - **'Medium'** and Product line -**'Classic Cars'** has the **Highest** ordered quantity(18746) and Deal size - **'Large'** and Product line -**'Ships'** has the **zero** ordered quantity(0)

Sales Productline/year

Productline / Year of Orderdate
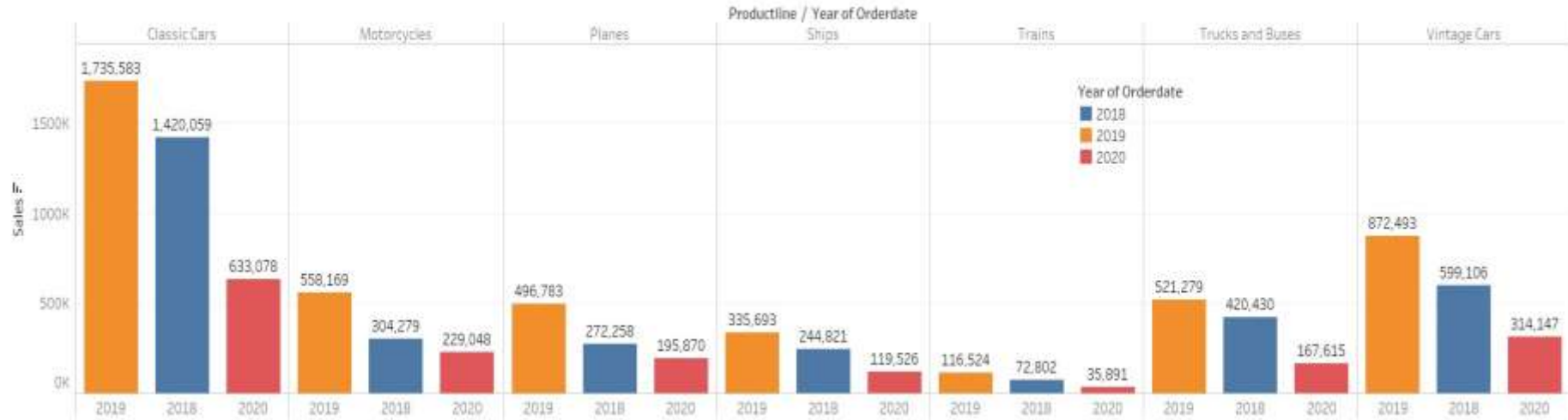
'Classic Cars' has the Highest increase in Sales from 2018 to 2019 and 'Trains' has the Lowest increase in Sales from 2018 to 2019

Sales by Month
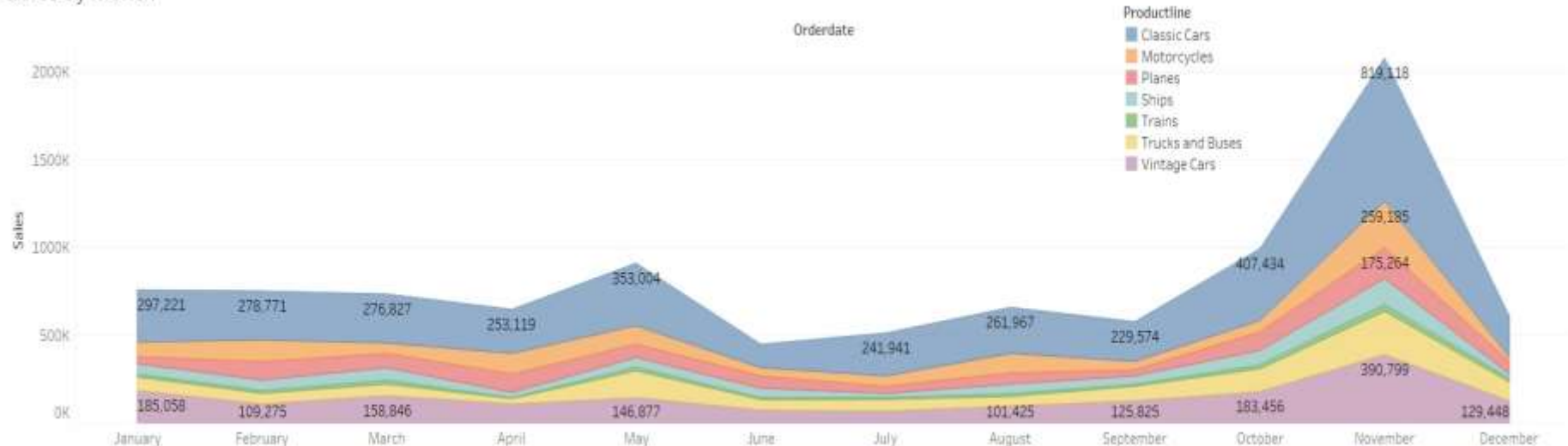
'November Month' is the Highest in Sales and 'June' has the Lowest in Sales for all Product line. Overall , 'Classic Cars' has highest sales amount and 'Trains' has lowest sales amount.

Distplot and Boxplot distribution
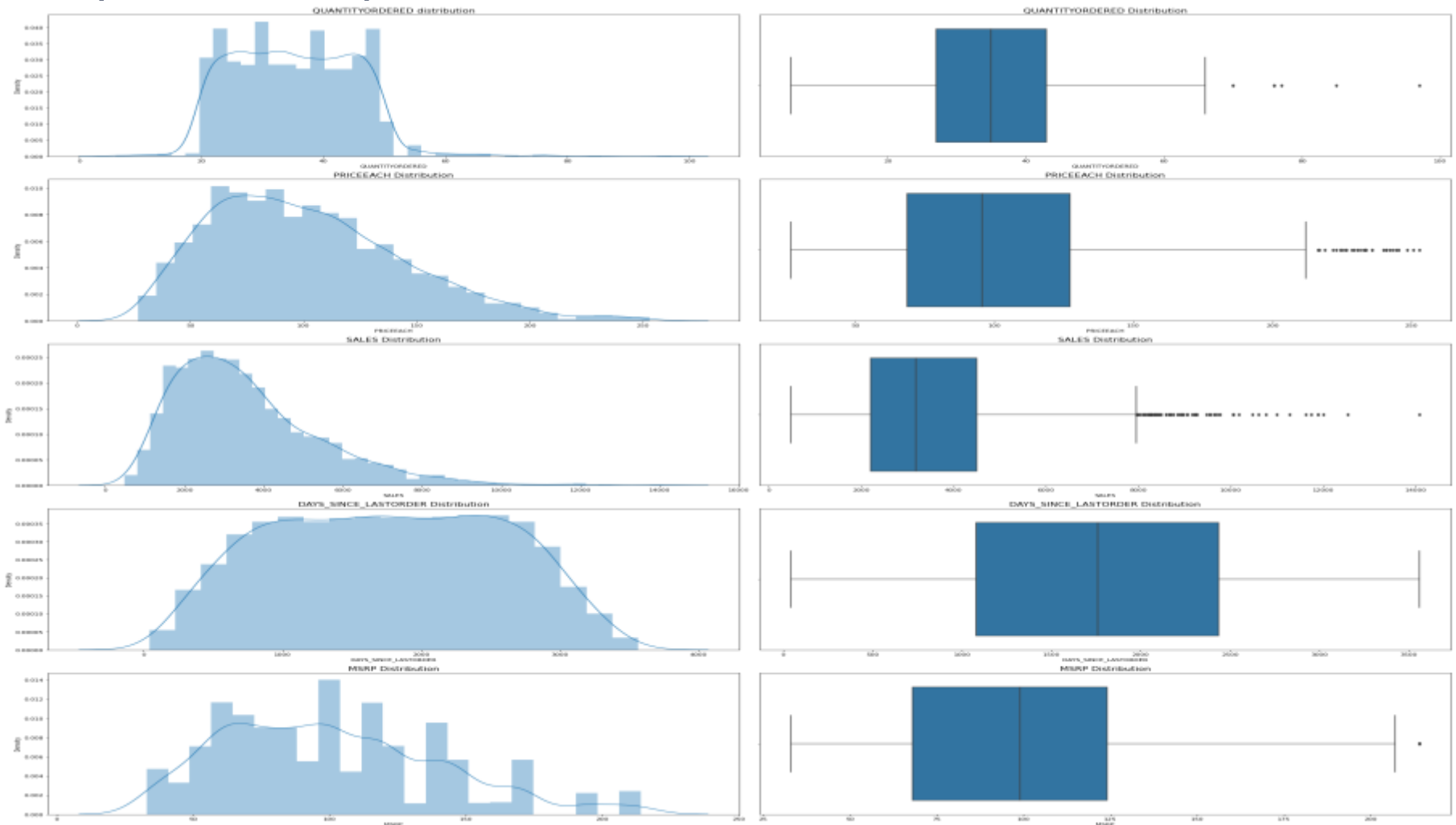
• There is presence of **outliers** in variables such as **Quantityordered,Price and Sales**.
• Variable **'Sales'** has highest **positive** skewness(0.784) and Variable '**Days_since_lastorder'** has lowest **negative** skewness(-0.002)

Heat Map

- Variables **'Sales'** and **'Priceeach'** have **highest positive** Correlation(0.81) and Variables '**Days_since_lastorder**' and '**MSRP**' have highest **negative** Correlation(-0.52)
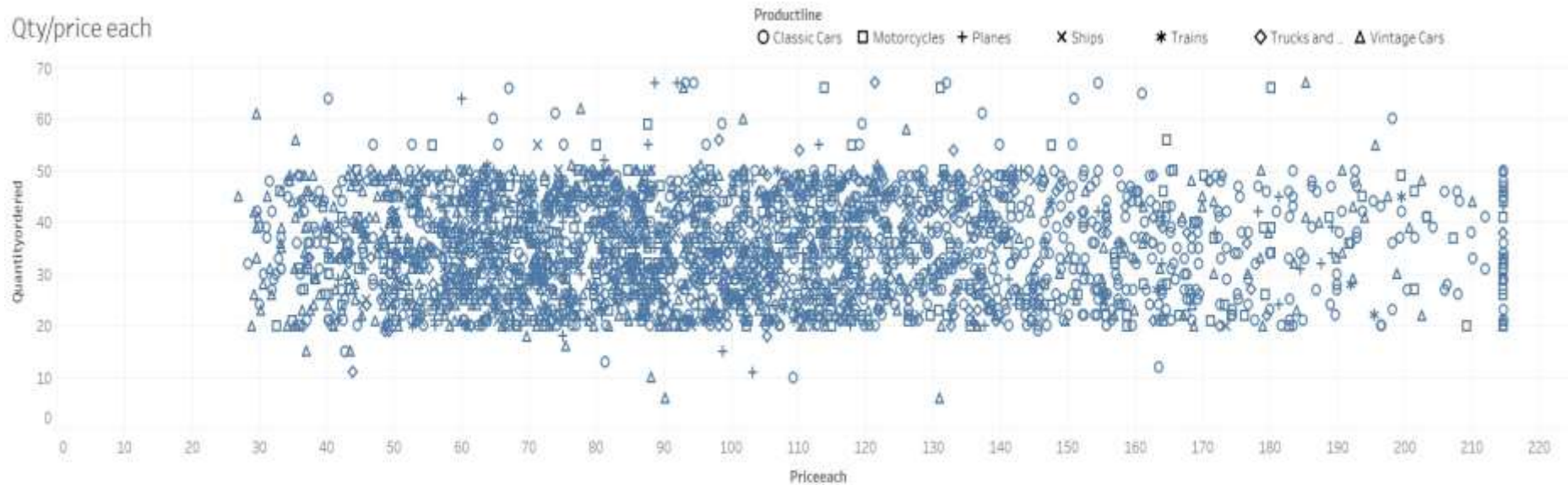
Pair Plot
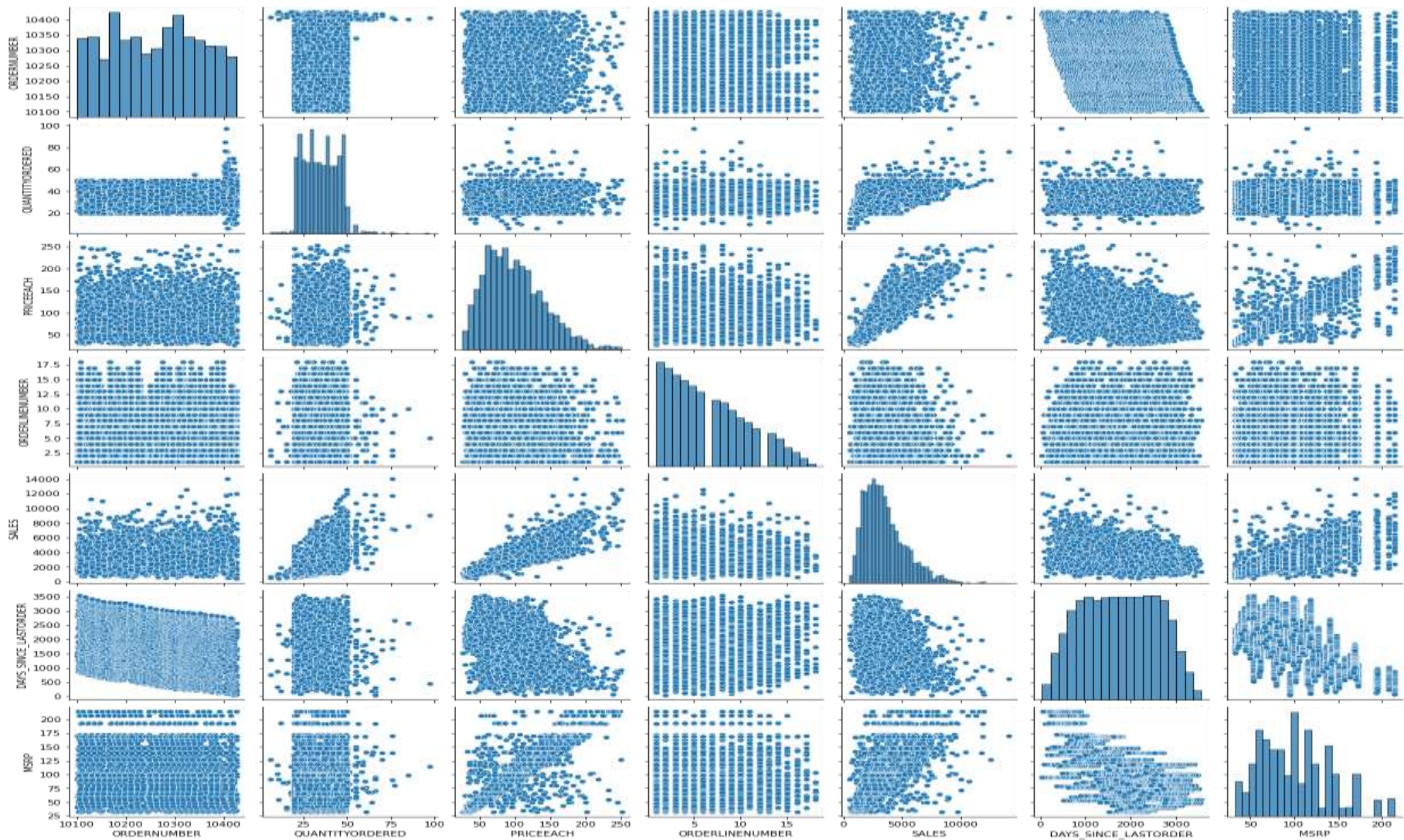
•From the plot , we can see that none of the variables are symmetric
•Variables '**Sales**' and '**Priceeach**' have almost linear realtionship between them.

Sales/Qty ordered

Productline

Quantityordered
| 2,712
▨ 10,000
▨ 20,000
▨ 33,322

Sales
225,218   3,788,721
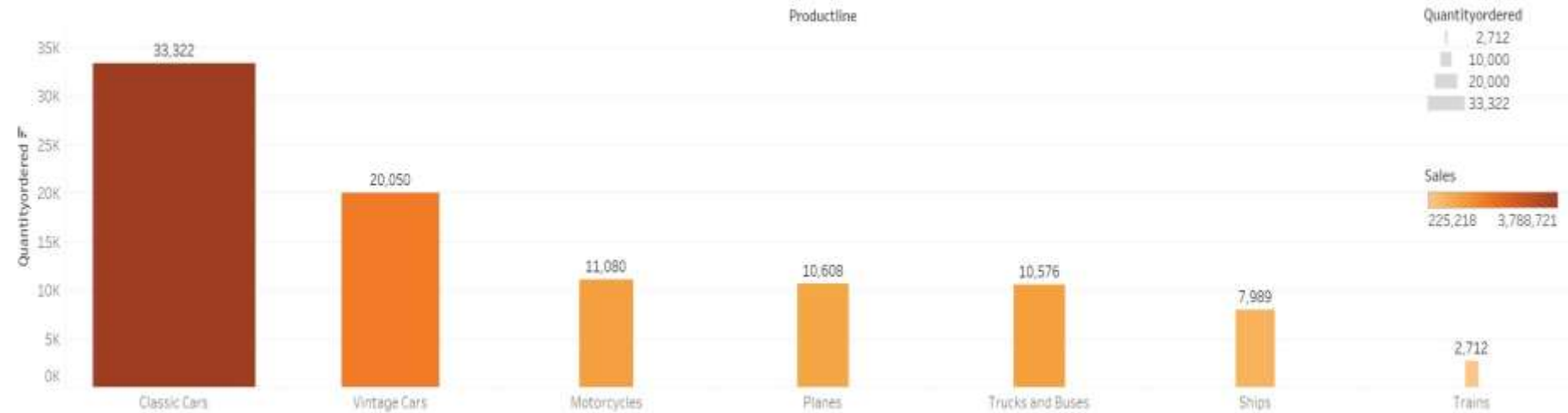
'Classic Cars' are the **Highest** in both 'Sales' & 'Quantity' and 'Trains' are the **Lowest** in both 'Sales' & 'Quantity'

Qty/price each

Productline
O Classic Cars   □ Motorcycles   + Planes   ✕ Ships   ✱ Trains   ◇ Trucks and ..   △ Vintage Cars

Priceeach

Majority of the units are in the range of **(20-50)** for **Quantity ordered** and **(50-120)** for the **Price**.

Sales diff by Msrp-Price

We can **increase Sales amount** for almost all the customers by **changing prices** for items for which **MSRP is greater than Price**

Sales diff by Product Line

Increase Sales amount for **highly ordered** Product line items to get the **maximum benefit**

Sales-Price each/Msrp

Sales-Price each/Msrp
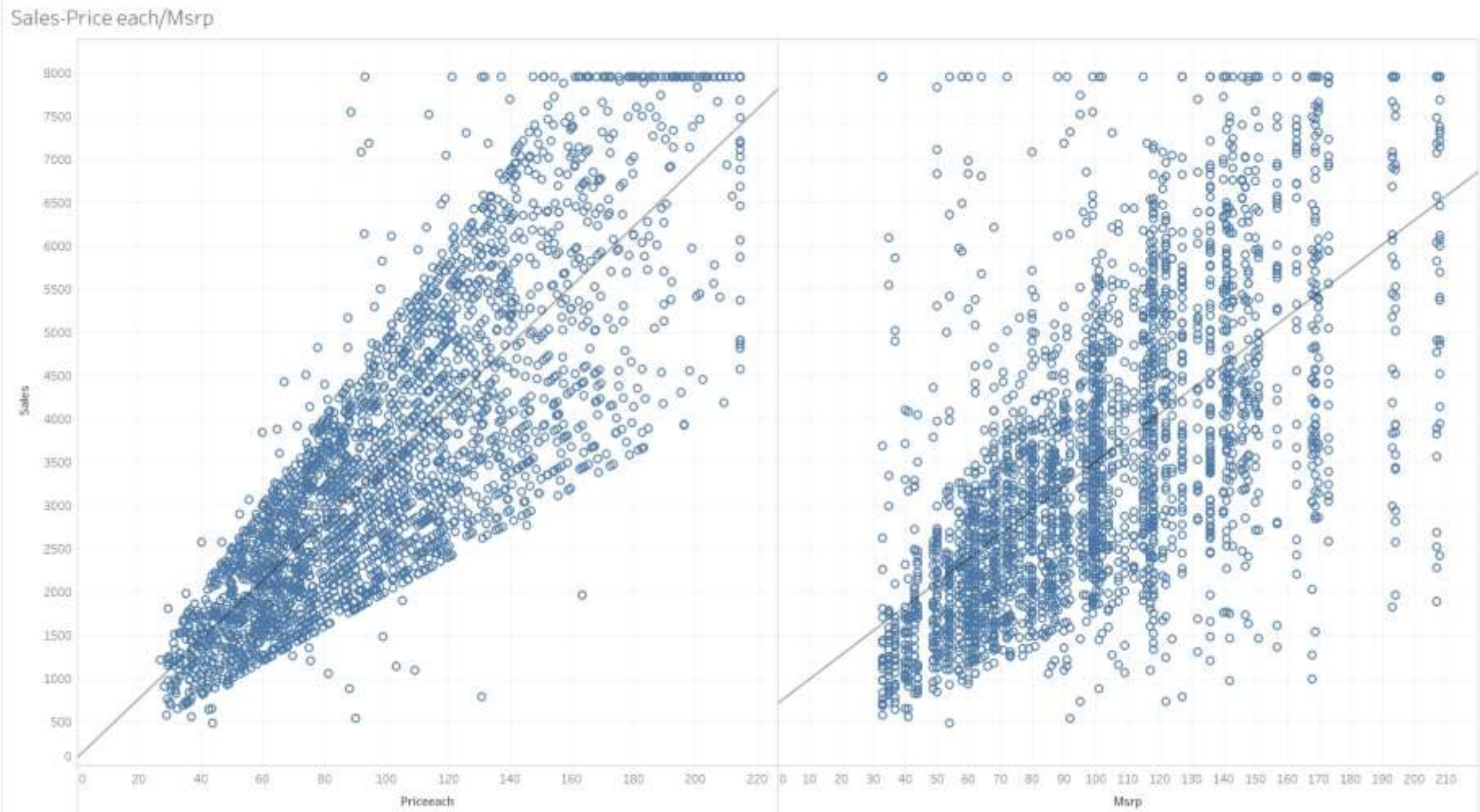
From the plots , we can see there is higher spread of data along the trend line for Msrp compared to Price each. So we need to maximize sales by identifying respective items for which there is higher price change.

Customer Segmentation using RFM analysis

Customer Segmentation done by using KNIME and MS Excel by dividing the data based on Receny,Frequency and Monetary variables through grouping data by variable 'CustomerName'

Recency – **Min** of variable **'DAYS_SINCE_LASTORDER'**
Frequency – **Unique count** of variable **'ORDERNUMBER'**
Monetary – **Sum** of variable **'SALES'**

Receny , Frequency and Monetary variables are divided into three bins based on percentile of respective column data.

Data	Recency	Frequency	Monetary
Below 30 percentile	High	Low	Low
30-70 percentile	Medium	Medium	Medium
Above 70 percentile	Low	High	High

Customer vs Order Frequency
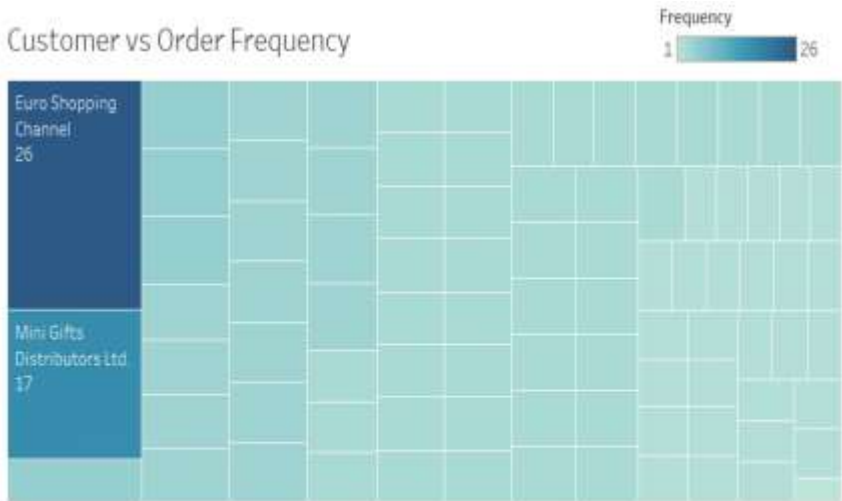
Frequency
1 — 26

European Shopping Channel has highest Order Frequency and Bavrian Collectables has lowest Order Frequency

Customer vs Min Days since last order

Min. Days Since Lastorder
42 — 1,353

European Shopping Channel has highest Recency and Auto-Moto Classics Inc. has lowest Recency

Sales by Customer

Customer - Euro Shopping Channel has highest sale value (902136) and Boards & Toys Co. has least sale value (9129)

KNIME WorkFlow Image and Output Table Head

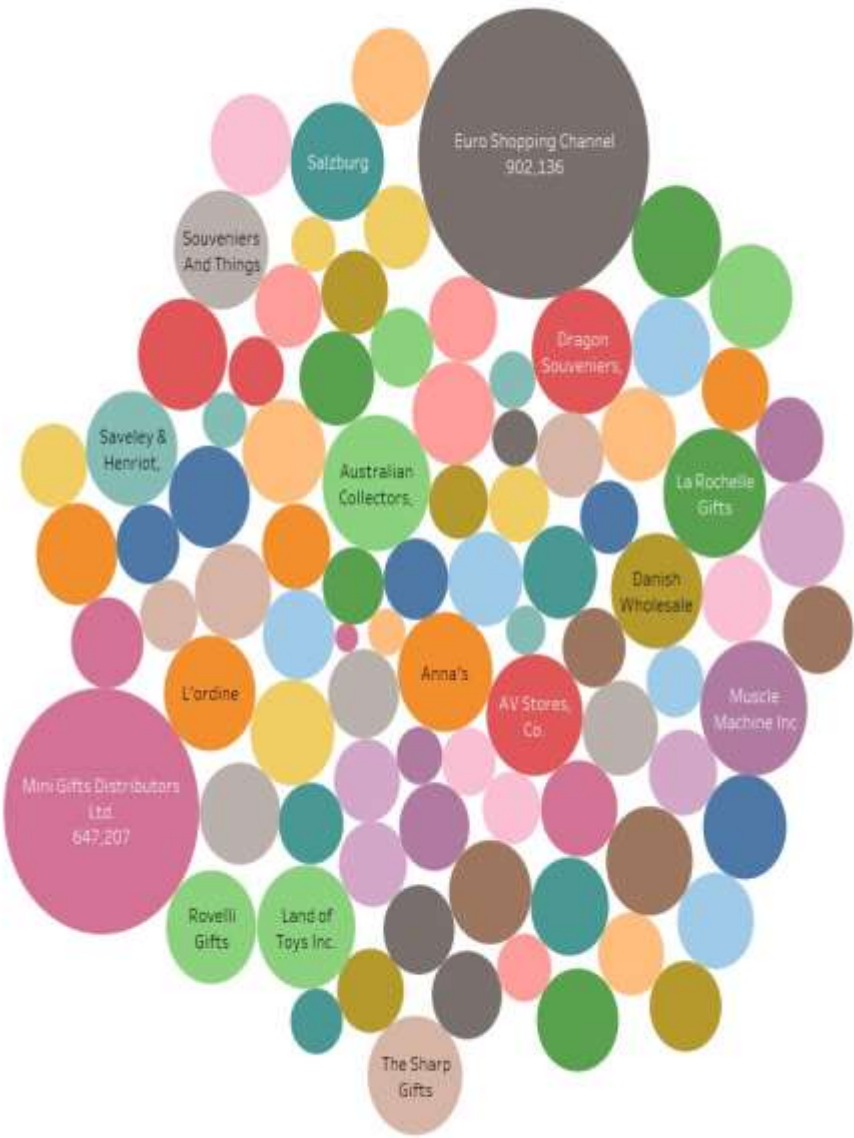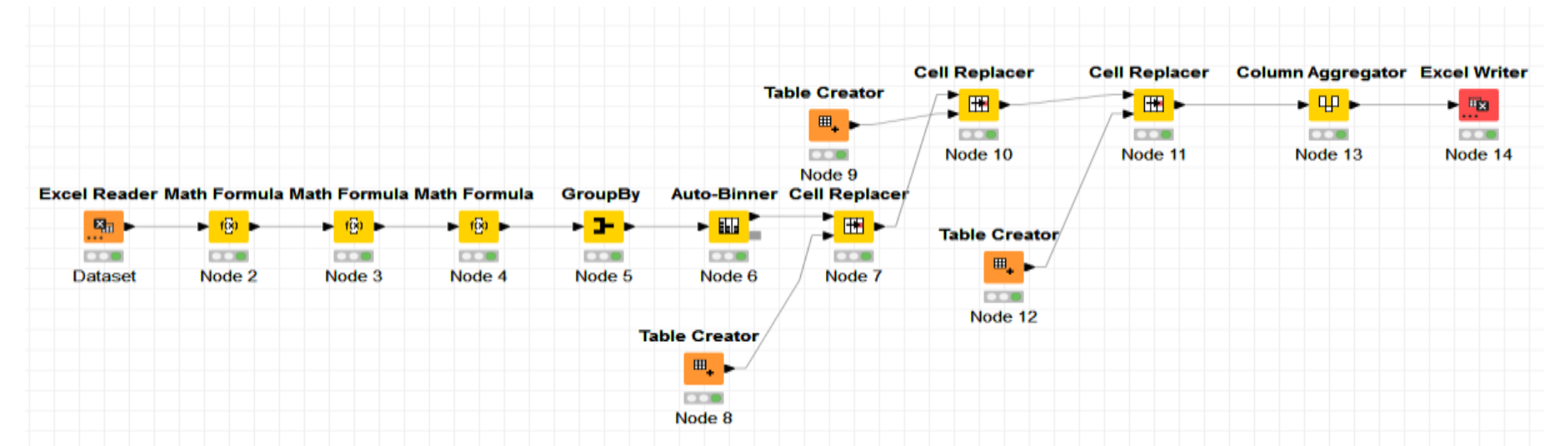| CUSTOMERNAME	QUANTITYORDERED	PRICEEACH	DAYS_SINCE_LASTORDER	RECENCY	Frequency	Monetary	RECENCY_HML	FREQUENCY_HML	MONETARY_HML	Concatenate
AV Stores, Co.	1778	91.0845098	421	421	3	157807.81	M	M	H	MMH
Alpha Cognac	687	101.16	675	675	3	70108.0425	L	M	L	LML
Amica Models & Co.	843	110.04125	328	328	2	93291.5175	M	L	M	MLM
Anna's Decorations, Ltd	1469	105.5928804	131	131	4	153083.705	H	H	H	HHH
Atelier graphique	270	92.23857143	312	312	3	24179.96	M	M	L	MML
Australian Collectables, Ltd	705	90.04173913	1018	1018	3	64115.6525	L	M	L	LML
Australian Collectors, Co.	1926	104.3235	229	229	5	197858.945	H	H	H	HHH
Australian Gift Network, Co	545	109.1475	190	190	3	59405.5125	H	M	L	HML
Auto Assoc. & Cie.	637	99.38736111	275	275	2	62612.8325	H	L	L	HLL
Auto Canal Petit	1001	94.25518519	127	127	3	93170.66	H	M	M	HMM
Auto-Moto Classics Inc.	287	92.8	1353	1353	3	26479.26	L	M	L	LML
Baane Mini Imports	1082	108.0078906	245	245	4	115706.2825	H	H	H	HHH
Bavarian Collectables Imports, Co.	401	84.28928571	801	801	1	34993.92	L	L	L	LLL
Blauer See Auto, Co.	811	108.0313636	705	705	4	85171.59	L	H	M	LHM
Boards & Toys Co.	102	89.80666667	410	410	2	9129.35	M	L	L	MLL
CAF Imports	468	104.9630769	625	625	2	49214.6825	L	L	L	LLL
Cambridge Collectables Co.	357	100.5802273	484	484	2	36163.62	M	L	L	MLL
Canadian Gift Exchange Network	703	105.3409091	364	364	2	74125.2425	M	L	L	MLL
Classic Gift Ideas, Inc	668	102.4177381	344	344	2	67506.97	M	L	L	MLL

Inferences from RFM Analysis and identified segments

RFM Count

Count of Concatenate

Recency Hml	Frequency Hml	Monetary Hml H	Monetary Hml M	Monetary Hml L
H	H	13		
	M	3	8	2
	L			1
M	H	4	1	
	M	3	9	1
	L	1	7	9
L	H	1	1	1
	M	2	7	5
	L		2	8

HHH has **highest** Count(13) and (**HHM,HHL,HLH,HLM,MHL and LLH**) have **lowest** Count of value zero (0)

We have got customers in three type of groups based upon most important aspect –'Receny'

Active : Customers with High Recency – **27 Nos**

These are the Most active customers as they belong to top 30 percentile of the customers and they play most important role in the business.

At-Risk :Customers with Medium Recency -**35Nos**

These are the Fairly active customers as they belong to middle 40 percentile of the customers. Need to concentrate on the customers otherwise we are at risk of losing them.

Inactive :Customers with Low Recency -**27Nos**

These are the least active customers as they belong to bottom 30 percentile of the customers. We can think of these the customers only after satisfying the needs of Active and At-Risk group customers

RFM Customer

Recency Hml	Frequency Hml	Monetary Hml		
		H	M	L

Recency Hml: H, Frequency Hml: H
- The Sharp Gifts Warehouse
- Australian Collectors, Co.
- Souveniers And Things Co.
- Euro Shopping Channel
- Mini Gifts Distributors Ltd.

Recency Hml: H, Frequency Hml: M
- UK Collectables, Ltd.
- Online Diecast Creations Co.
- Scandinavian Gift Ideas
- Oulu Toy Supplies, Inc.
- Collectables For Less Inc.
- FunGiftIdeas.com
- Gifts4AllAges.com
- Quebec Home Shopping Network
- Australian Gift Network, Co

Recency Hml: H, Frequency Hml: L
- Auto Assoc. & Cie.

Recency Hml: M, Frequency Hml: H
- Reims Collectables
- Danish Wholesale Imports
- Muscle Machine Inc
- Handji Gifts& Co
- Volvo Model Replicas, Co

Recency Hml: M, Frequency Hml: M
- L'ordine Souveniers
- AV Stores, Co.
- Corrida Auto Replicas, Ltd
- Tekni Collectables Inc.
- Classic Legends Inc.
- Suominen Souveniers
- La Corne D'abondance, Co.
- Motor Mint Distributors Inc.
- Atelier graphique

Recency Hml: M, Frequency Hml: L
- Vida Sport, Ltd
- Toms Spezialitten, Ltd
- Amica Models & Co.
- Marta's Replicas Co.
- Heintze Collectables
- Royal Canadian Collectables, Ltd.
- Cambridge Collectables Co.
- Online Mini Collectables
- Canadian Gift Exchange Network

Recency Hml: L, Frequency Hml: H
- Dragon Souveniers, Ltd.
- BlauerSee Auto, Co.
- Royale Belge

Recency Hml: L, Frequency Hml: M
- Saveley & Henriot, Co.
- Rovelli Gifts
- Stylish Desk Decors, Co.
- Enaco Distributors
- Mini Creations Ltd.
- Marseille Mini Autos
- Mini Auto Werke
- Alpha Cognac
- Auto-Moto Classics Inc.
- Australian Collectables, Ltd

Recency Hml: L, Frequency Hml: L
- Norway Gifts By Mail, Co.
- Collectable Mini Designs Co.
- Iberia Gift Imports, Corp.
- Bavarian Collectables Imports, Co.
- Double Decker Gift Stores, Ltd
- Daedalus Designs Imports

Best Customers – HHH-13 Nos

These are the most valuable customers as they are with High level in all the three aspects Receny, Frequency and Monetary.We should not loose these customers at any cost as they are the biggest contributors of the business.

Anna's Decorations, Ltd Australian Collectors, Co. Baane Mini Imports Diecast Classics Inc. Euro Shopping Channel La Rochelle Gifts Land of Toys Inc. Mini Gifts Distributors Ltd. Salzburg Collectables Souveniers And Things Co. Technics Stores Inc. The Sharp Gifts Warehouse Tokyo Collectables, Ltd

Better Customers – HHM,HMH,HMM-11Nos

These are the still valuable customers as they are with better level in all aspects Frequency and Monetary. We have to look after the needs of these customers to bring them into the pool of best.

Auto Canal Petit Collectables For Less Inc. FunGiftIdeas.com Gift Depot Inc. Gifts4AllAges.com Lyon Souveniers Oulu Toy Supplies, Inc. Toys of Finland, Co., Online Diecast Creations Co. Scandinavian Gift Ideas UK Collectables, Ltd.

Good Customers – HLH,HLM,HLL,HHL,HML-3 Nos

These are customers as average level in all the aspects Frequency and Monetary.We can still concentrate on them to increase business as they have got High Receny level.

Australian Gift Network, Co Quebec Home Shopping Network, Auto Assoc. & Cie.

Bad Customers –LLL-8Nos

These are the most avoidable customers as they are with lowest level in all aspects Receny, Frequency and Monetary. There is no point in spending time and effort to maintain business with these customers.

Bavarian Collectables Imports, Co. CAF Imports Clover Collections, Co. Daedalus Designs Imports Diecast Collectables Double Decker Gift Stores, Ltd Iberia Gift Imports, Corp. Signal Collectibles Ltd.