

MRA Project

PARTHASARATHI BEHURA

DSBA APR-B

Table of Content

- Problem Statement and Business context
- Data Info
- Basic EDA: univariate, bivariate and trend analysis
- RFM analysis
- Customer segmentation and RFM insights

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.

CHECKING NULL VALUES: -

- Here is the basic info on the data: -
- The data consists of 2747 rows and 20 columns.
- There are no null values and no duplicate values in the data.

Null Values	
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

- DATA INFO: -

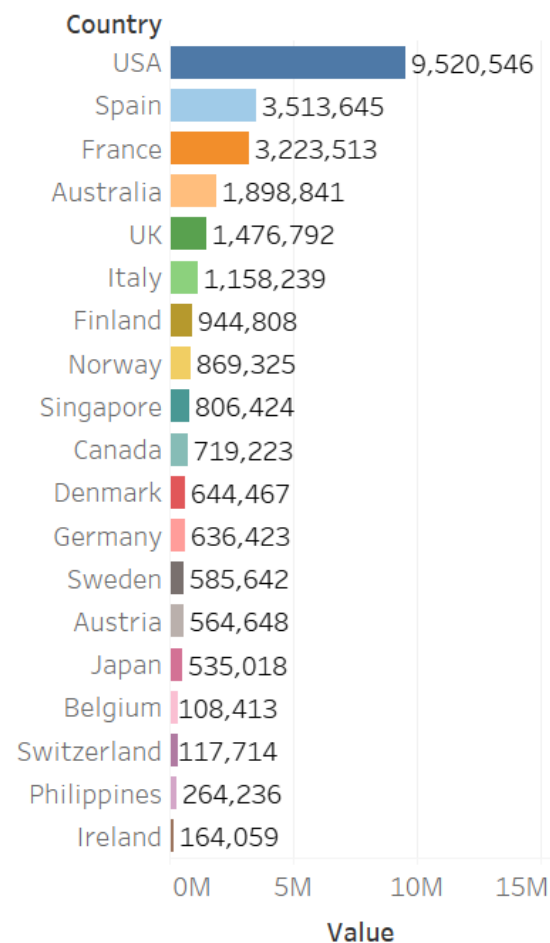
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2747 entries, 0 to 2746
Data columns (total 20 columns):
#   Column                Non-Null Count  Dtype
---  -
0   ORDERNUMBER           2747 non-null   int64
1   QUANTITYORDERED       2747 non-null   int64
2   PRICEEACH             2747 non-null   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 DESCRIPTION: -

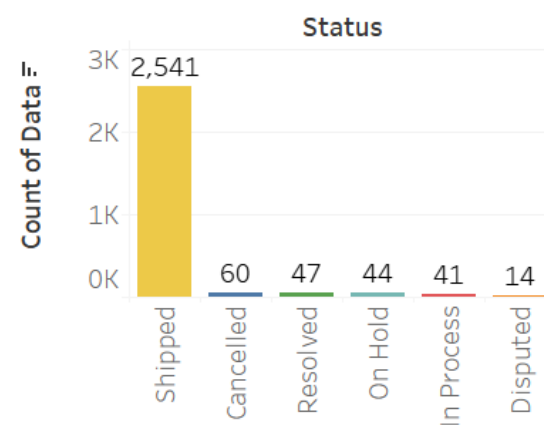
	count	mean	min	25%	50%	75%	max	std
ORDERNUMBER	2747.0	10259.761558	10100.0	10181.0	10264.0	10334.5	10425.0	91.877521
QUANTITYORDERED	2747.0	35.103021	6.0	27.0	35.0	43.0	97.0	9.762135
PRICEEACH	2747.0	101.098951	26.88	68.745	95.55	127.1	252.87	42.042548
ORDERLINENUMBER	2747.0	6.491081	1.0	3.0	6.0	9.0	18.0	4.230544
SALES	2747.0	3553.047583	482.13	2204.35	3184.8	4503.095	14082.8	1838.953901
ORDERDATE	2747	2019-05-13 21:56:17.211503360	2018-01-06 00:00:00	2018-11-08 00:00:00	2019-06-24 00:00:00	2019-11-17 00:00:00	2020-05-31 00:00:00	NaN
DAYS_SINCE_LASTORDER	2747.0	1757.085912	42.0	1077.0	1761.0	2436.5	3562.0	819.280576
MSRP	2747.0	100.691664	33.0	68.0	99.0	124.0	214.0	40.114802

UNIQUE VALUES AND THEIR COUNT

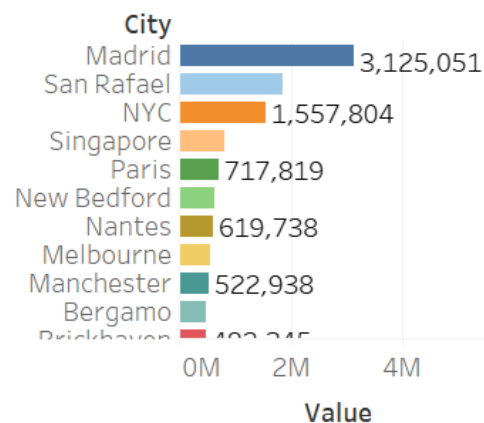
Country Count



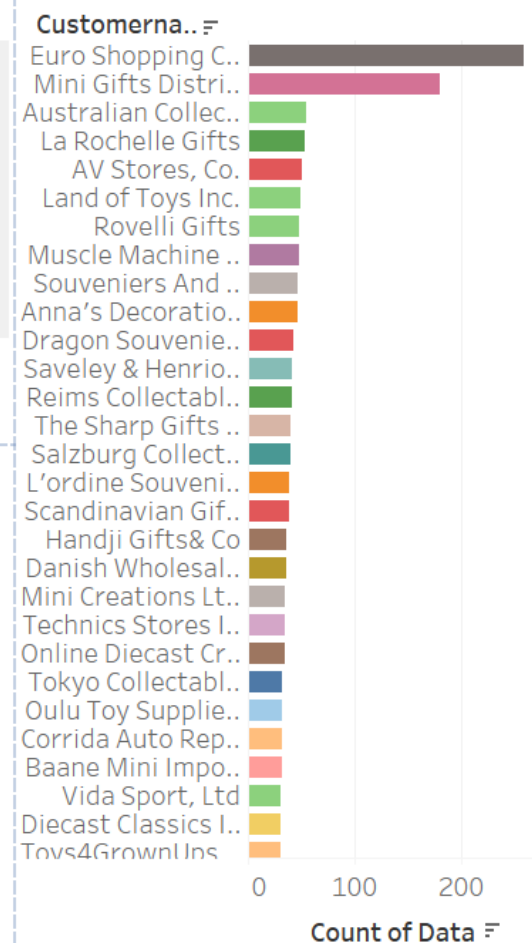
Status Count



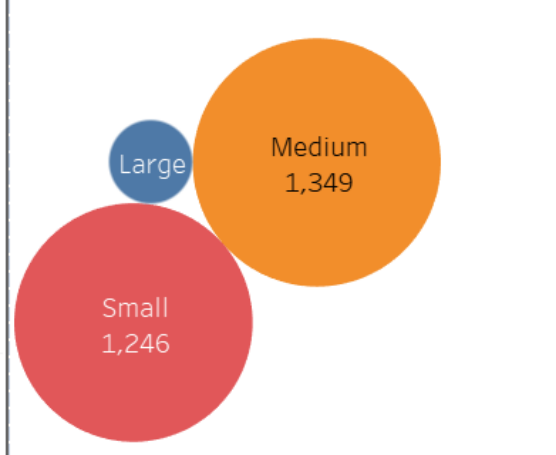
City Count



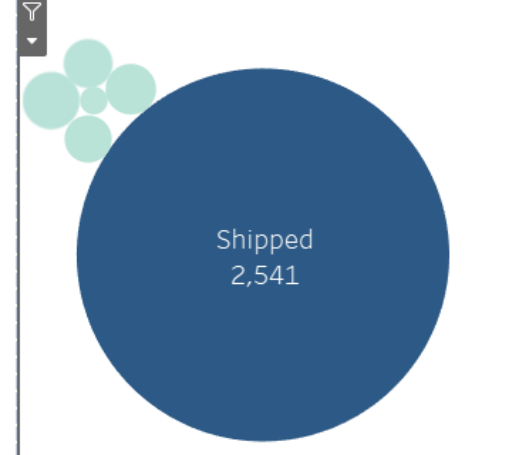
Customer Count



Deal size Count



Status



UNIVARAITE ANALYSIS

1. Ordernumber

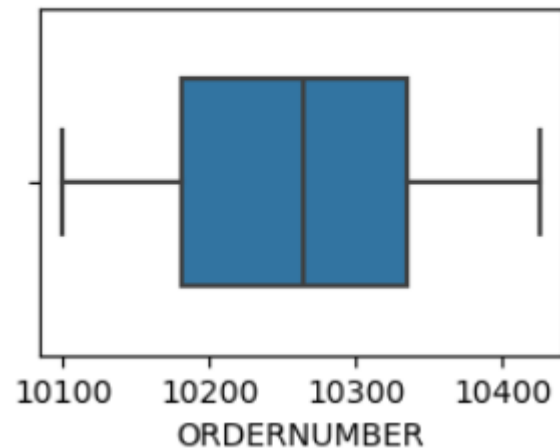
Description of ORDERNUMBER is: -

count	2747.000000
mean	10259.761558
std	91.877521
min	10100.000000
25%	10181.000000
50%	10264.000000
75%	10334.500000
max	10425.000000

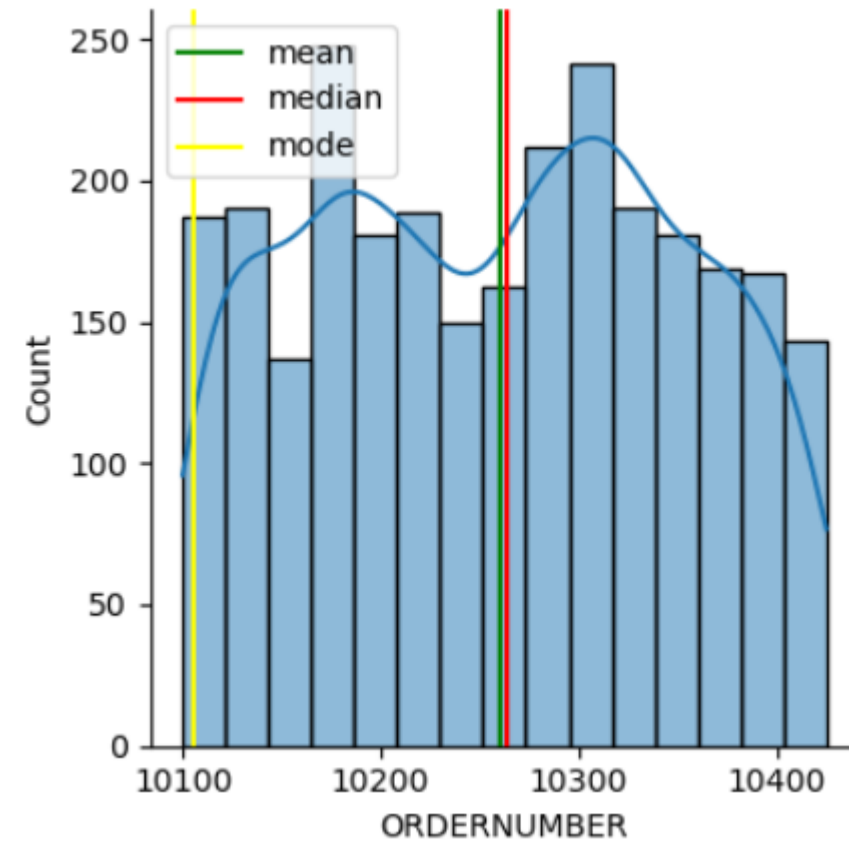
Name: ORDERNUMBER, dtype: float64

Mean is: 10259.761558063341
Median is: 10264.0
Mode is: 10106

Boxplot of ORDERNUMBER is: -



Distribution of ORDERNUMBER is: -



UNIVARAITE ANALYSIS

2. Quantityordered

Description of QUANTITYORDERED is: -

count	2747.000000
mean	35.103021
std	9.762135
min	6.000000
25%	27.000000
50%	35.000000
75%	43.000000
max	97.000000

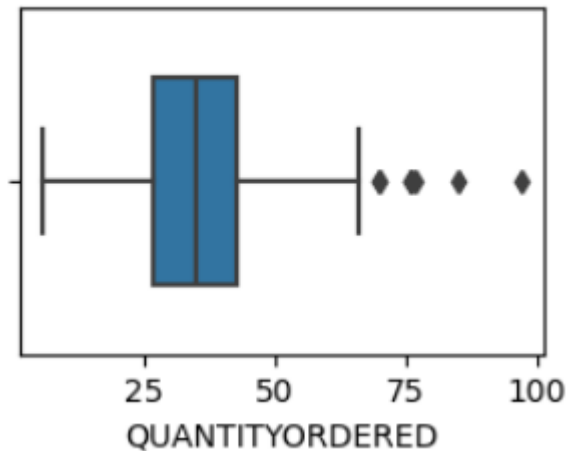
Name: QUANTITYORDERED, dtype: float64

Mean is: 35.10302147797597

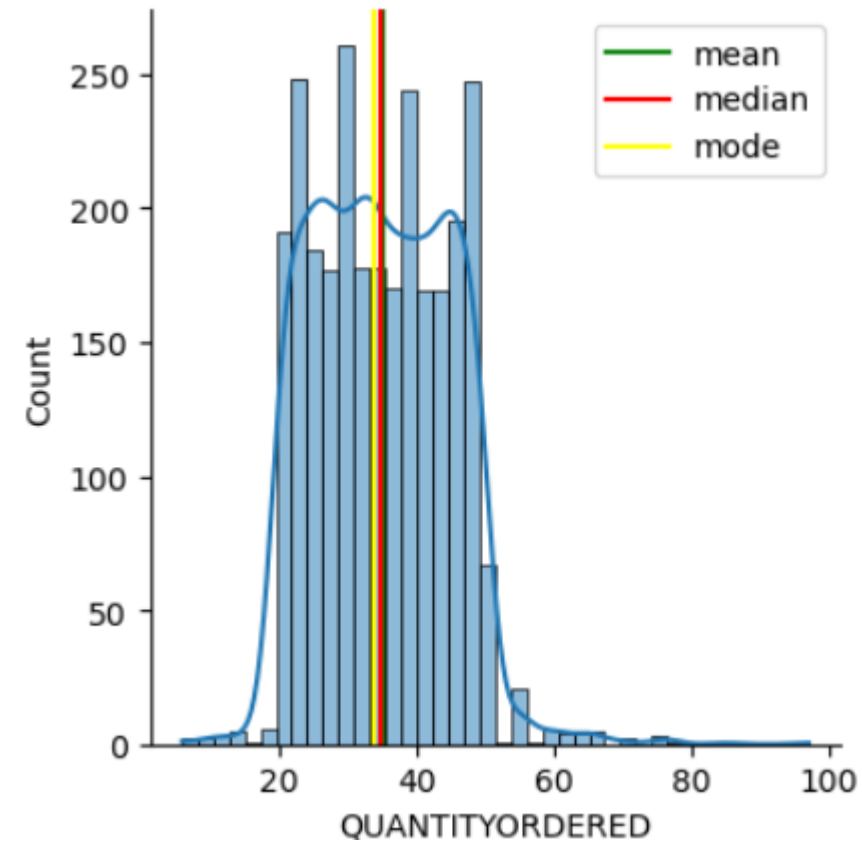
Median is: 35.0

Mode is: 34

Boxplot of QUANTITYORDERED is: -



Distribution of QUANTITYORDERED is: -



3. Priceeach

Description of PRICE EACH is: -

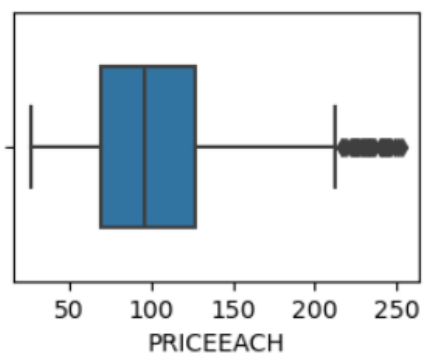
```
count    2747.000000
mean      101.098951
std       42.042548
min       26.880000
25%       68.745000
50%       95.550000
75%      127.100000
max       252.870000
Name: PRICE EACH, dtype: float64
```

Mean is: 101.09895113914965

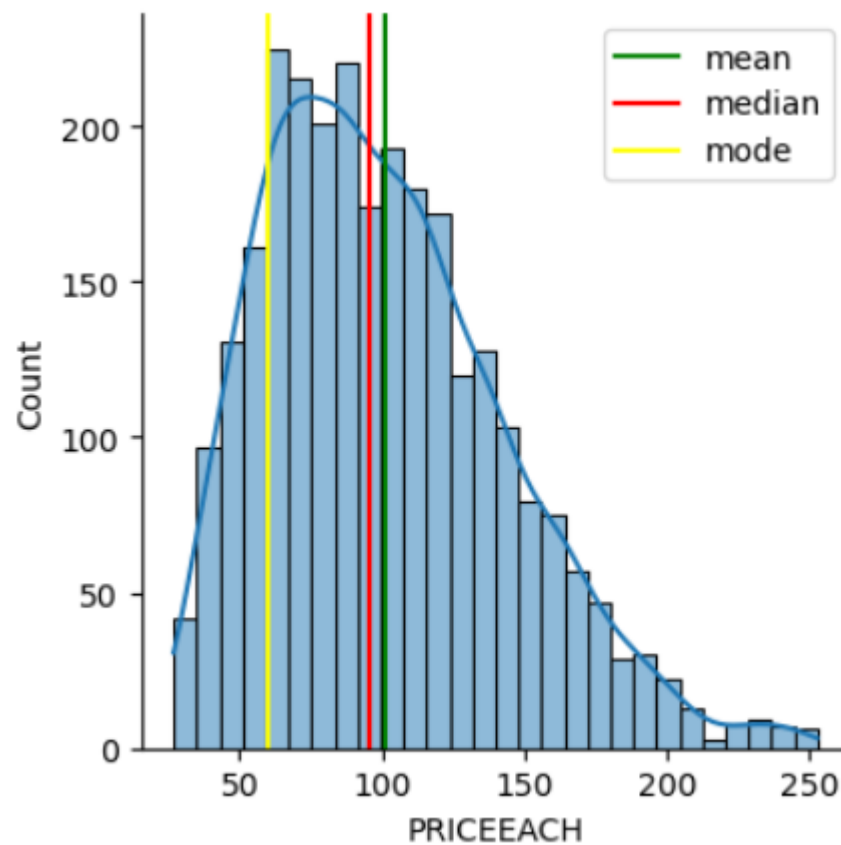
Median is: 95.55

Mode is: 59.87

Boxplot of PRICEEACH is: -



Distribution of PRICEEACH is: -



4. Orderlinenumber

Description of ORDERLINENUMBER is: -

```
count    2747.000000
mean       6.491081
std        4.230544
min        1.000000
25%        3.000000
50%        6.000000
75%        9.000000
max       18.000000
```

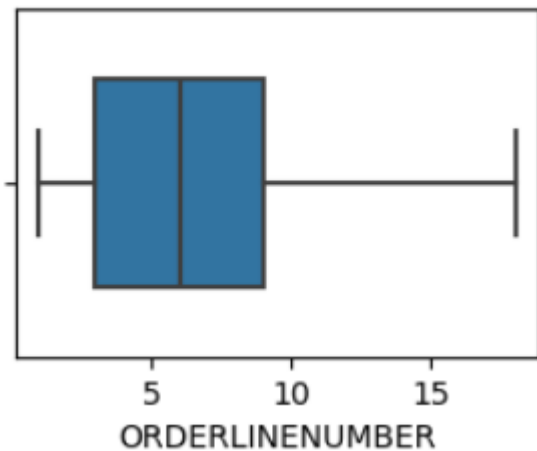
Name: ORDERLINENUMBER, dtype: float64

Mean is: 6.4910811794685115

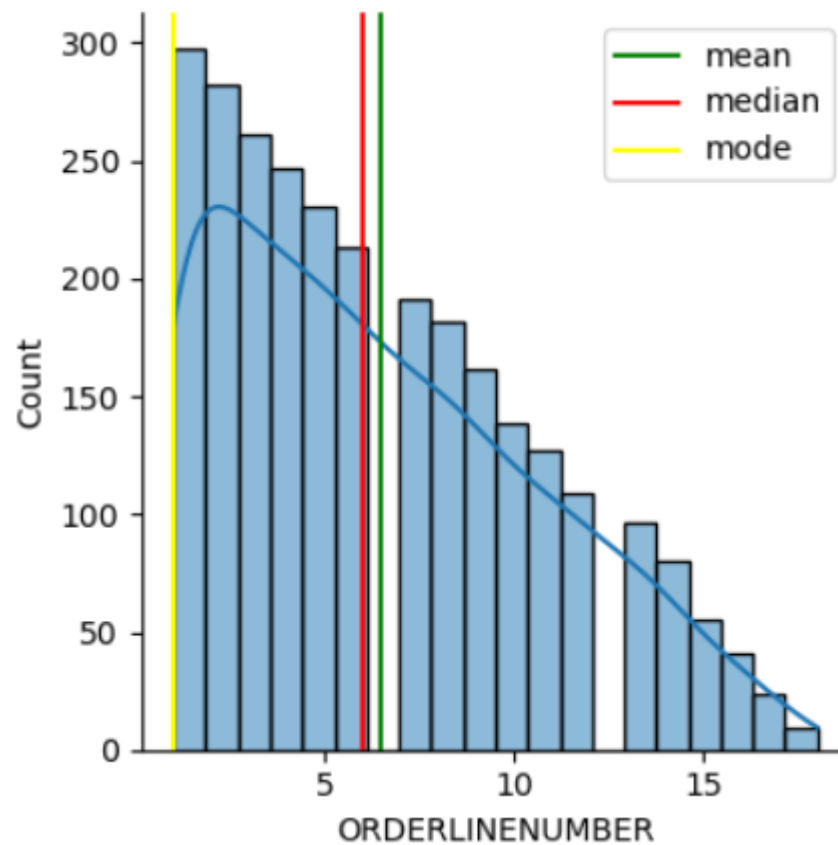
Median is: 6.0

Mode is: 1

Boxplot of ORDERLINENUMBER is: -



Distribution of ORDERLINENUMBER is: -



5. Sales

Description of SALES is: -

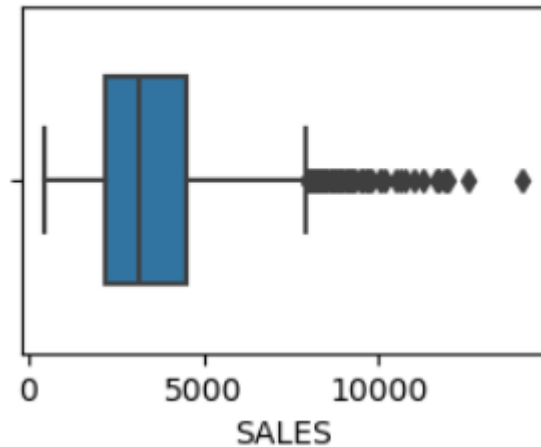
```
count    2747.000000
mean     3553.047583
std      1838.953901
min       482.130000
25%      2204.350000
50%      3184.800000
75%      4503.095000
max     14082.800000
Name: SALES, dtype: float64
```

Mean is: 3553.0475828176195

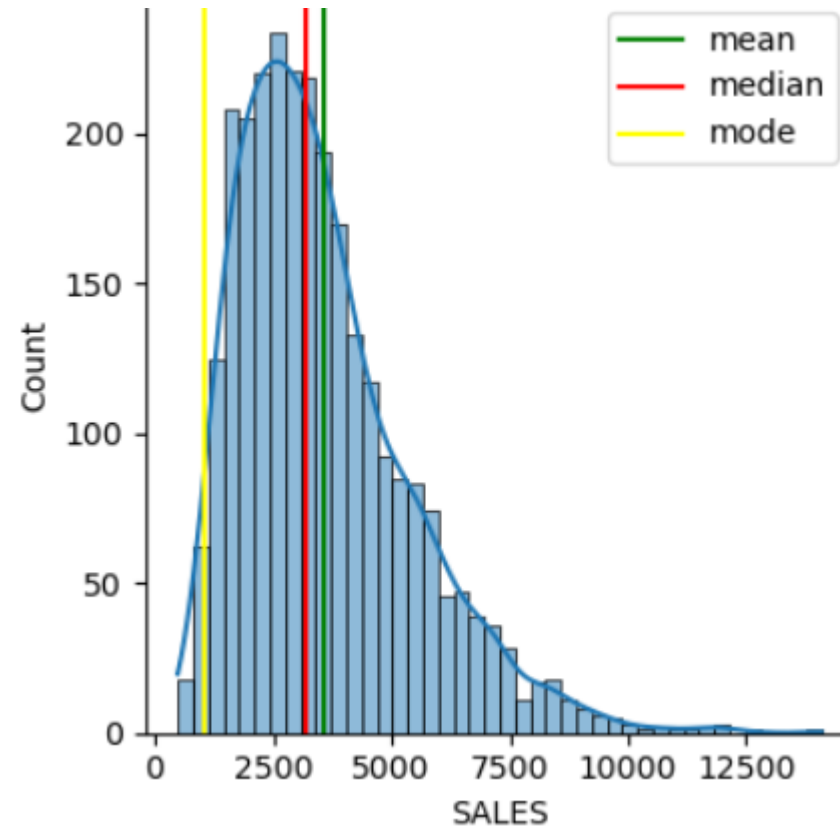
Median is: 3184.8

Mode is: 1030.44

Boxplot of SALES is: -



Distribution of SALES is: -



6. MSRP

Description of MSRP is: -

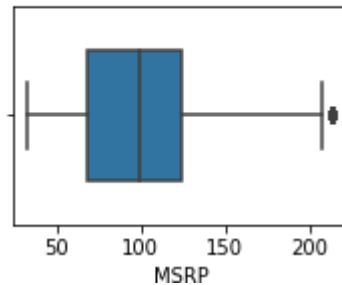
```
count    2747.000000
mean      100.691664
std       40.114802
min       33.000000
25%       68.000000
50%       99.000000
75%      124.000000
max      214.000000
Name: MSRP, dtype: float64
```

Mean is: 100.69166363305425

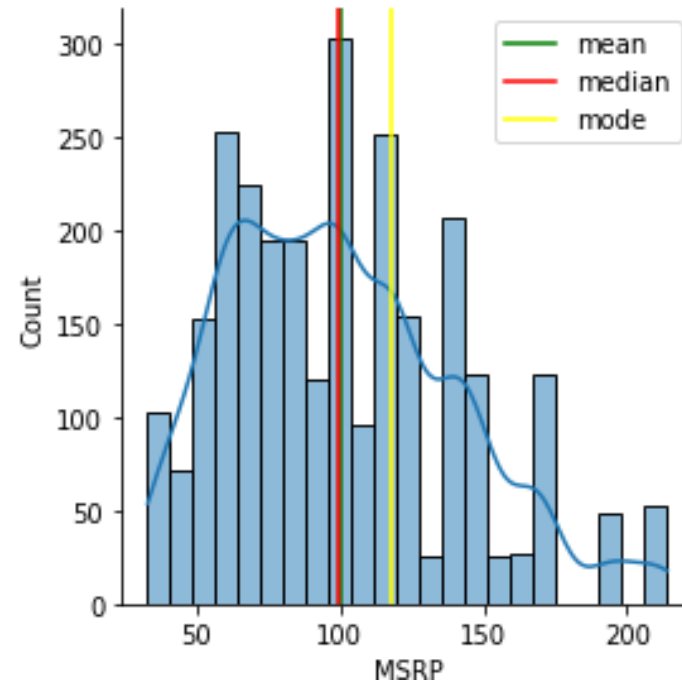
Median is: 99.0

Mode is: 118

Boxplot of MSRP is: -

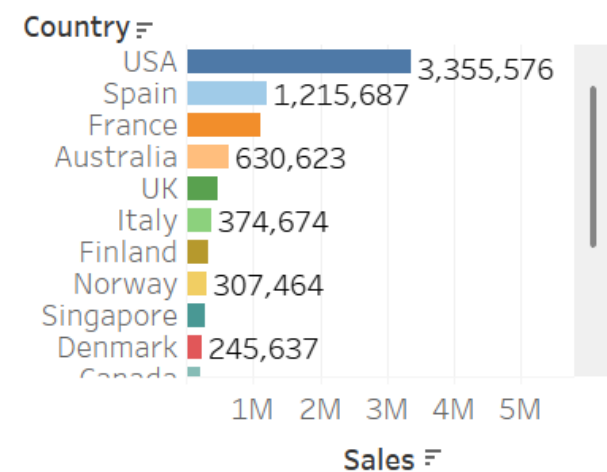


Distribution of MSRP is: -

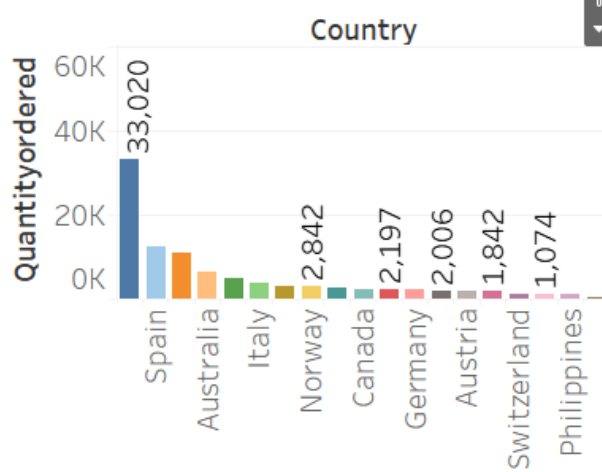


BIVARAITE ANALYSIS

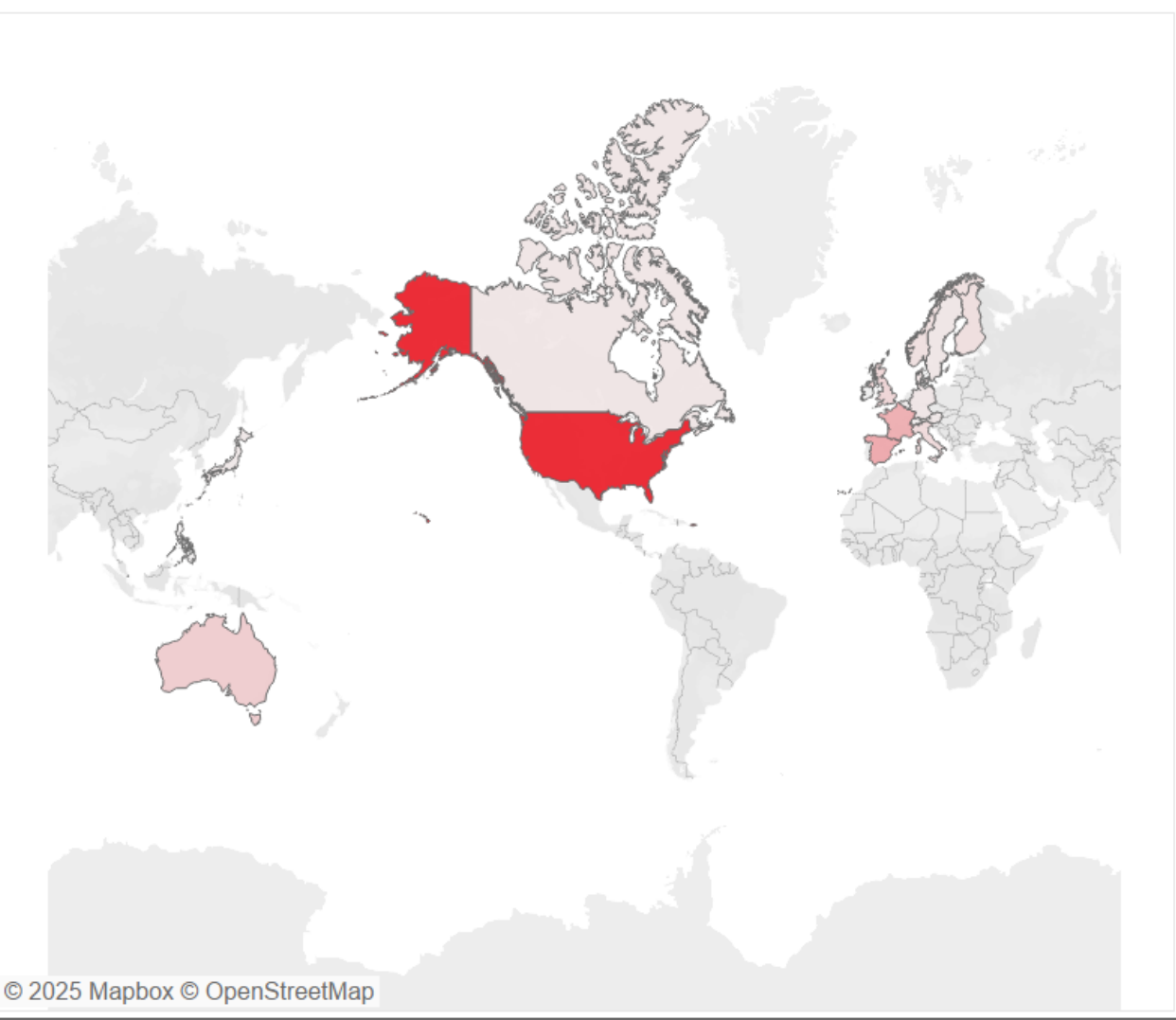
Country Vs Sales



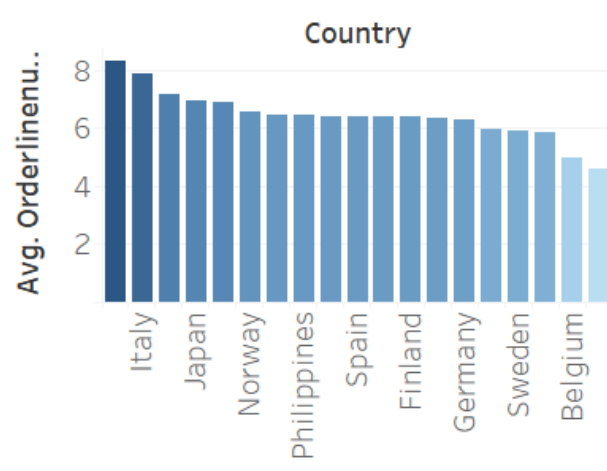
Country Vs Qty ordered



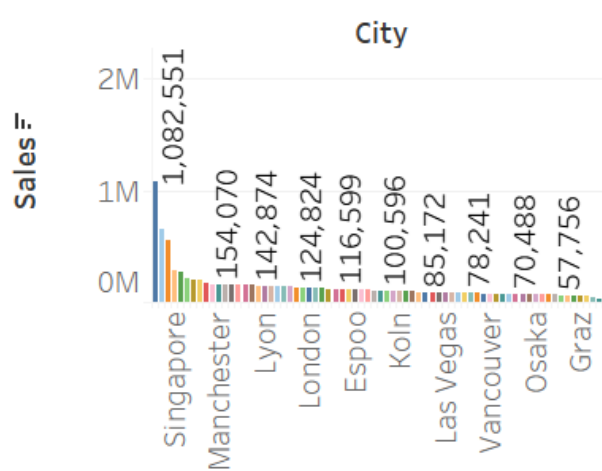
Sales on Map



Country Vs Orderline



Sales Vs City



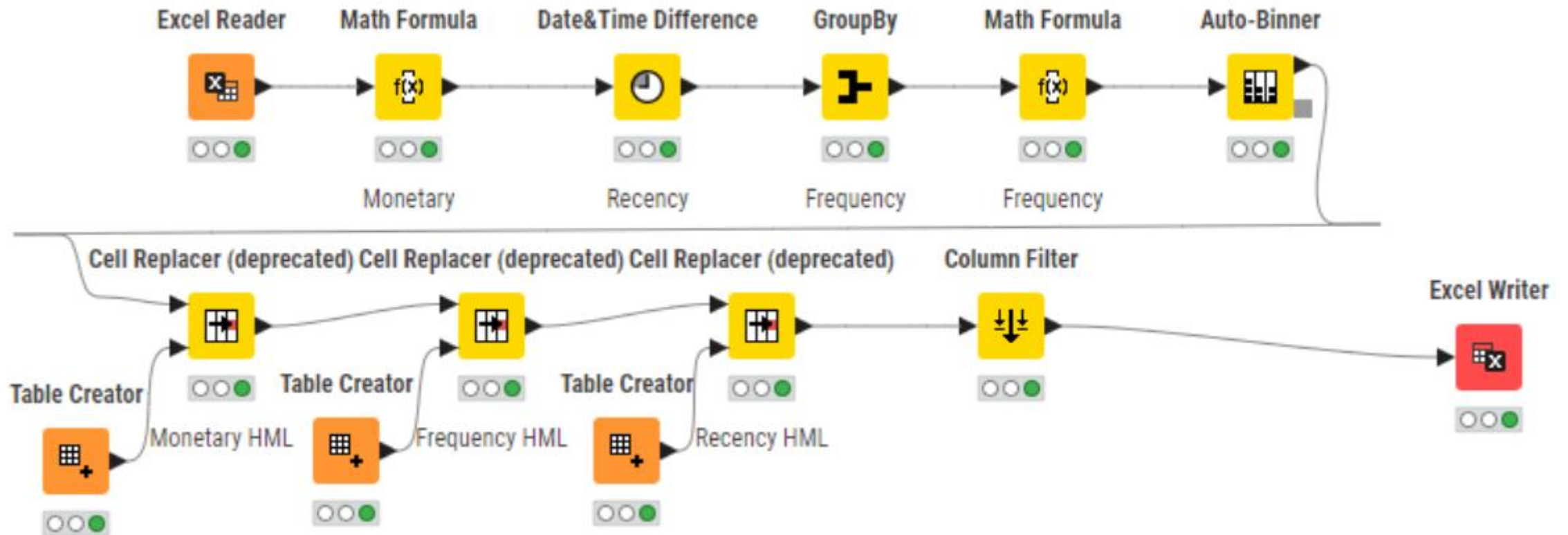
EDA SUMMARY AND INFERENCES

- Most of the products are already shipped.
- Classic car & Vintage cars are highly selling productlines.
- USA and Spain are the two countries with max sales. In US San Rafael and NYC are the cities that cater to max sales, while in Spain, Madrid is the largest Market.
- Euro Shopping channel and Mini gifts distributor, are the largest customers.
- Switzerland and Italy are having the highest orderline.

RFM

- Using RFM the customers are divided into 3 segments, namely, High valued, Mid valued or At Risk/Potential and Churn. Knime workflow was used for the same.
- To start off, Recency, Frequency and Monetary columns were defined and customers segmented based upon it.
- It is assumed that the high and mid valued customers are loyal. The at risk customers might have some potential left and they could be converted to mid valued by different promotional schemes
- However, the ones with Churn could be allowed to let go.

Knime Workflow



Segmented Table

	CUSTOMERNAME	ORDERNUMBER	QUANTITYORDERED	PRICEEACH	ORDERLINENUMBER	SALES		Monetary	Recency	Frequency	Monetary_HML	Frequency_HML	Recency_HML
0	AV Stores, Co.	51	34.862745	91.084510	51	157807.81	157807.81	1904	51	High Valued	High Valued	Risk/Potential	
1	Alpha Cognac	20	34.350000	101.160000	20	70488.44	70488.44	1772	20	Churn	Churn	High Valued	
2	Amica Models & Co.	26	32.423077	110.852692	26	94117.26	94117.26	1973	26	Mid Valued	Risk/Potential	Churn	
3	Anna's Decorations, Ltd	46	31.934783	106.424130	46	153996.13	153996.13	1791	46	High Valued	High Valued	Risk/Potential	
4	Atelier graphique	7	38.571429	92.238571	7	24179.96	24179.96	1896	7	Churn	Churn	Risk/Potential	

RFM Inferences

Best Customers: -

	CUSTOMERNAME	Recency_HML	Frequency_HML	Monetary_HML
26	Danish Wholesale Imports	High Valued	High Valued	High Valued
32	Euro Shopping Channel	High Valued	High Valued	High Valued
41	L'ordine Souvenirs	High Valued	High Valued	High Valued
43	La Rochelle Gifts	High Valued	High Valued	High Valued
53	Mini Gifts Distributors Ltd.	High Valued	High Valued	High Valued
63	Reims Collectables	High Valued	High Valued	High Valued
67	Salzburg Collectables	High Valued	High Valued	High Valued
72	Souvenirs And Things Co.	High Valued	High Valued	High Valued
78	The Sharp Gifts Warehouse	High Valued	High Valued	High Valued

RFM Inferences

Verge of Churning: -

CUSTOMERNAME	Recency_HML	Frequency_HML	Monetary_HML
Alpha Cognac	High Valued	Churn	Churn
Australian Collectables, Ltd	High Valued	Risk/Potential	Churn
Auto Canal Petit	High Valued	Risk/Potential	Mid Valued
Danish Wholesale Imports	High Valued	High Valued	High Valued
Diecast Classics Inc.	High Valued	Risk/Potential	High Valued
Euro Shopping Channel	High Valued	High Valued	High Valued
Gift Depot Inc.	High Valued	Risk/Potential	Mid Valued
Gifts4AllAges.com	High Valued	Risk/Potential	Mid Valued
Handji Gifts& Co	High Valued	High Valued	Mid Valued
L'ordine Souvenirs	High Valued	High Valued	High Valued
La Rochelle Gifts	High Valued	High Valued	High Valued
Lyon Souvenirs	High Valued	Churn	Mid Valued
Mini Caravy	High Valued	Churn	Mid Valued
Mini Gifts Distributors Ltd.	High Valued	High Valued	High Valued
Petit Auto	High Valued	Risk/Potential	Mid Valued
Quebec Home Shopping Network	High Valued	Risk/Potential	Mid Valued
Reims Collectables	High Valued	High Valued	High Valued
Salzburg Collectables	High Valued	High Valued	High Valued
Souvenirs And Things Co.	High Valued	High Valued	High Valued
Tekni Collectables Inc.	High Valued	Risk/Potential	Mid Valued
The Sharp Gifts Warehouse	High Valued	High Valued	High Valued
Tokyo Collectables, Ltd	High Valued	Risk/Potential	Mid Valued
UK Collectables Ltd.	High Valued	Risk/Potential	Mid Valued

RFM Inferences

Lost Customers: -

	CUSTOMERNAME	Recency_HML	Frequency_HML	Monetary_HML
8	Auto Assoc. & Cie.	Churn	Churn	Churn
12	Bavarian Collectables Imports, Co.	Churn	Churn	Churn
15	CAF Imports	Churn	Churn	Churn
16	Cambridge Collectables Co.	Churn	Churn	Churn
20	Clover Collections, Co.	Churn	Churn	Churn
25	Daedalus Designs Imports	Churn	Churn	Churn
29	Double Decker Gift Stores, Ltd	Churn	Churn	Churn
40	Iberia Gift Imports, Corp.	Churn	Churn	Churn
58	Online Mini Collectables	Churn	Churn	Churn
59	Osaka Souveniers Co.	Churn	Churn	Churn
70	Signal Collectibles Ltd.	Churn	Churn	Churn
87	West Coast Collectables Co.	Churn	Churn	Churn

RFM Inferences

Loyal Customers: -

Any customer with a decent recency matrix can be termed as loyal

CUSTOMERNAME	Recency_HML	Frequency_HML	Monetary_HML
Alpha Cognac	High Valued	Churn	Churn
Australian Collectables, Ltd	High Valued	Risk/Potential	Churn
Auto Canal Petit	High Valued	Risk/Potential	Mid Valued
Danish Wholesale Imports	High Valued	High Valued	High Valued
Diecast Classics Inc.	High Valued	Risk/Potential	High Valued
Euro Shopping Channel	High Valued	High Valued	High Valued
Gift Depot Inc.	High Valued	Risk/Potential	Mid Valued
Gifts4AllAges.com	High Valued	Risk/Potential	Mid Valued
Handji Gifts& Co	High Valued	High Valued	Mid Valued
L'ordine Souvenirs	High Valued	High Valued	High Valued
La Rochelle Gifts	High Valued	High Valued	High Valued
Lyon Souvenirs	High Valued	Churn	Mid Valued
Mini Caravy	High Valued	Churn	Mid Valued
Mini Gifts Distributors Ltd.	High Valued	High Valued	High Valued
Petit Auto	High Valued	Risk/Potential	Mid Valued
Quebec Home Shopping Network	High Valued	Risk/Potential	Mid Valued
Reims Collectables	High Valued	High Valued	High Valued
Salzburg Collectables	High Valued	High Valued	High Valued
Souvenirs And Things Co.	High Valued	High Valued	High Valued
Tekni Collectables Inc.	High Valued	Risk/Potential	Mid Valued
The Sharp Gifts Warehouse	High Valued	High Valued	High Valued
Tokyo Collectables, Ltd	High Valued	Risk/Potential	Mid Valued
UK Collectables, Ltd.	High Valued	Risk/Potential	Mid Valued