Business Report

Index- Page no.

Agenda & Executive Summary of the data -> Contents of the ppt -> Problem statement -> About Data (Info, Shape, Summary Stats, your assumptions about data)

- We are going to recommend customized marketing strategies for different segments of customers for automobile parts manufacturing company.

- In the provided data from the company we have 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	int64
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

- We don't have any null values.
- The shape of the data is **(2747, 20)**
- We have 2747 data points and 20 columns.
- The dataset contains 2 columns with floating-point numerical data, 6 columns with integer numerical data, and 12 columns with categorical/object data types.
- Order date is int data type , we have to convert this to datetime format

	ORDERNUMBER	QUANTITYORDERED	PRICEEACH	ORDERLINENUMBER	SALES	ORDERDATE	DAYS_SINCE_LASTORDER	MSRP
count	2747.000000	2747.000000	2747.000000	2747.000000	2747.000000	2747.000000	2747.000000	2747.000000
mean	10259.761558	35.103021	101.098951	6.491081	3553.047583	43598.914088	1757.085912	100.691664
std	91.877521	9.762135	42.042548	4.230544	1838.953901	230.231295	819.280576	40.114802
min	10100.000000	6.000000	26.880000	1.000000	482.130000	43106.000000	42.000000	33.000000
25%	10181.000000	27.000000	68.745000	3.000000	2204.350000	43412.000000	1077.000000	68.000000
50%	10264.000000	35.000000	95.550000	6.000000	3184.800000	43640.000000	1761.000000	99.000000
75%	10334.500000	43.000000	127.100000	9.000000	4503.095000	43786.000000	2436.500000	124.000000
max	10425.000000	97.000000	252.870000	18.000000	14082.800000	43982.000000	3562.000000	214.000000

ORDERNUMBER

- The column "ORDERNUMBER" represents unique order numbers for each transaction made by customers.
- The dataset contains a total of 2,747 unique order numbers.
- The minimum order number is 10100, and the maximum order number is 10425

QUANTITYORDERED

- The "QUANTITYORDERED" column indicates the quantity of items ordered in each transaction.
- The quantity ordered ranges from 6 to 97 units.
- The average quantity ordered per transaction is approximately 35 units.

PRICEEACH

- "PRICEEACH" refers to the unit price of each item in the order.
- The unit price varies from $26.88 to $252.87.
- The average unit price across all transactions is approximately $101.10.

ORDERLINENUMBER

- The "ORDERLINENUMBER" column denotes the position of the item in the order.

- The order line number ranges from 1 to 18, indicating the sequential arrangement of items in a transaction.

SALES

- The "SALES" column represents the total sales amount for each transaction, calculated as the product of "QUANTITYORDERED" and "PRICEEACH."
- The sales amount per order ranges from $482.13 to $14,082.80.
- The average sales value per order is approximately $3,553.05.

ORDERDATE:

- "ORDERDATE" is represented as an integer, suggesting that it might require conversion to a datetime format for further analysis.
- The order dates likely cover a period of three years, given that the data spans transactions for 3 years.

DAYS_SINCE_LASTORDER:

- This column shows the number of days since the last order made by each customer.
- The days since the last order range from 42 to 3562 days.
- The average number of days since the last order is approximately 1757 days.

MSRP

- "MSRP" stands for Manufacturer's Suggested Retail Price, which indicates the suggested retail price of the items in the order.
- The MSRP ranges from $33 to $214.
- The average MSRP across all transactions is approximately $100.69.

Value counts for categorical columns:

Value Counts for STATUS:

- Shipped 2541
- Cancelled 60
- Resolved 47
- On Hold 44
- In Process 41
- Disputed 14

Value Counts for PRODUCTLINE:

- Classic Cars 949
- Vintage Cars 579
- Motorcycles 313
- Planes 304
- Trucks and Buses 295
- Ships 230
- Trains 77

Value Counts for CUSTOMERNAME

- Euro Shopping Channel 259
- Mini Gifts Distributors Ltd. 180
- Australian Collectors, Co. 55
- La Rochelle Gifts 53
- AV Stores, Co. 51
- ...
- Microscale Inc. 10
- Royale Belge 8
- Auto-Moto Classics Inc. 8
- Atelier graphique 7
- Boards & Toys Co. 3

Length: 89

Value Counts for CITY:

- Madrid 304
- San Rafael 180
- NYC 152
- Singapore 79
- Paris 70
- ...
- Brisbane 15
- Sevilla 15
- Munich 14
- Burbank 13
- Charleroi 8

Length: 71

Value Counts for COUNTRY:

- USA 928
- Spain 342
- France 314
- Australia 185
- UK 144
- Italy 113
- Finland 92
- Norway 85
- Singapore 79
- Canada 70
- Denmark 63
- Germany 62
- Sweden 57
- Austria 55
- Japan 52

- Belgium 33
- Switzerland 31
- Philippines 26
- Ireland 16

Value Counts for DEALSIZE:

- Medium 1349
- Small 1246
- Large 152

Assumption for order date column -

- We determined the maximum number of days since the last order and then calculated the difference between this value and the corresponding order date to establish a reference point.
- We consider the order with the maximum days since the last order as the earliest and most distant entry in our data, serving as the reference for the first order in our dataset.

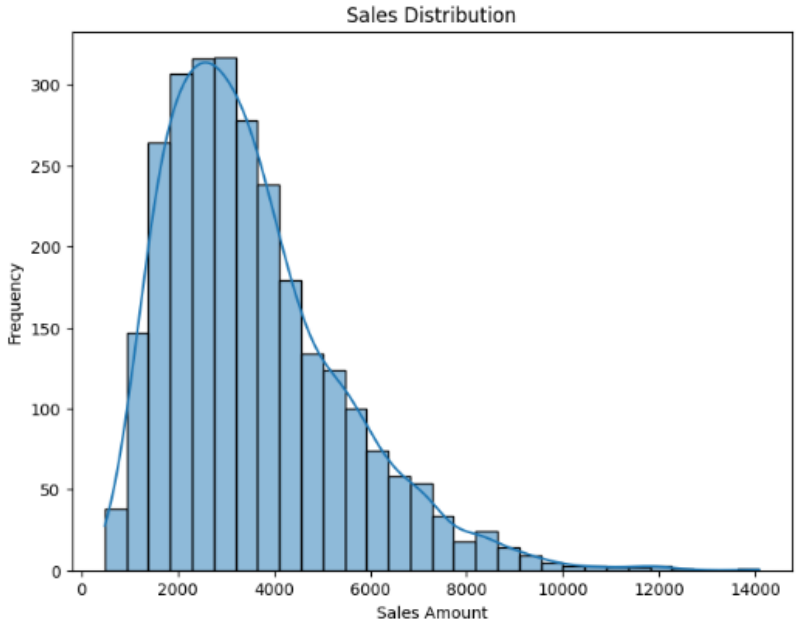
After conversion to datetime-

	ORDERNUMBER	QUANTITYORDERED	PRICEEACH	ORDERLINENUMBER	SALES	ORDERDATE	DAYS_SINCE_LASTORDER	STATUS	PRODUCTLINE	MSRP	PRODUCTCODE	CUSTOMERNAME
0	10107	30	95.70	2	2871.00	2013-10-14	828	Shipped	Motorcycles	95	S10_1678	Land of Toys Inc
1	10121	34	81.35	5	2765.90	2013-08-03	757	Shipped	Motorcycles	95	S10_1678	Reims Collectables
2	10134	41	94.74	2	3884.34	2013-06-09	703	Shipped	Motorcycles	95	S10_1678	Lyon Souveniers
3	10145	45	83.26	6	3746.70	2013-04-15	649	Shipped	Motorcycles	95	S10_1678	Toys4GrownUps.com
4	10168	36	96.66	1	3479.76	2013-02-10	586	Shipped	Motorcycles	95	S10_1678	Technics Stores Inc

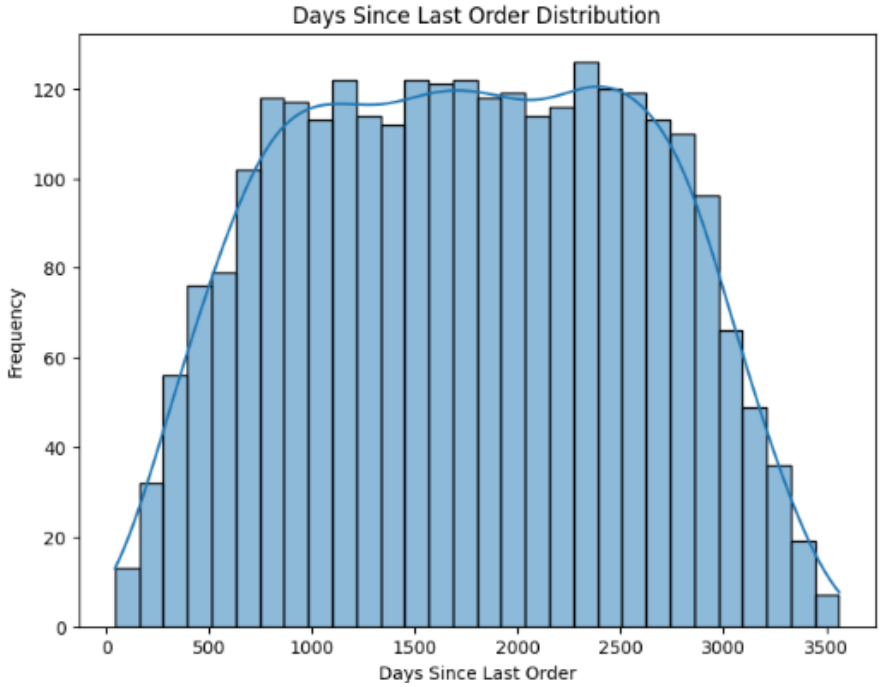
According to our order_date column the span of this data collection is 876 days.

Exploratory Analysis and Inferences -> Univariate, Bivariate, and multivariate analysis using data visualization (Weekly, Monthly, Quarterly, Yearly Trends in Sales and Sales Across different Categories of different features in the given data) -> Summarise the inferences

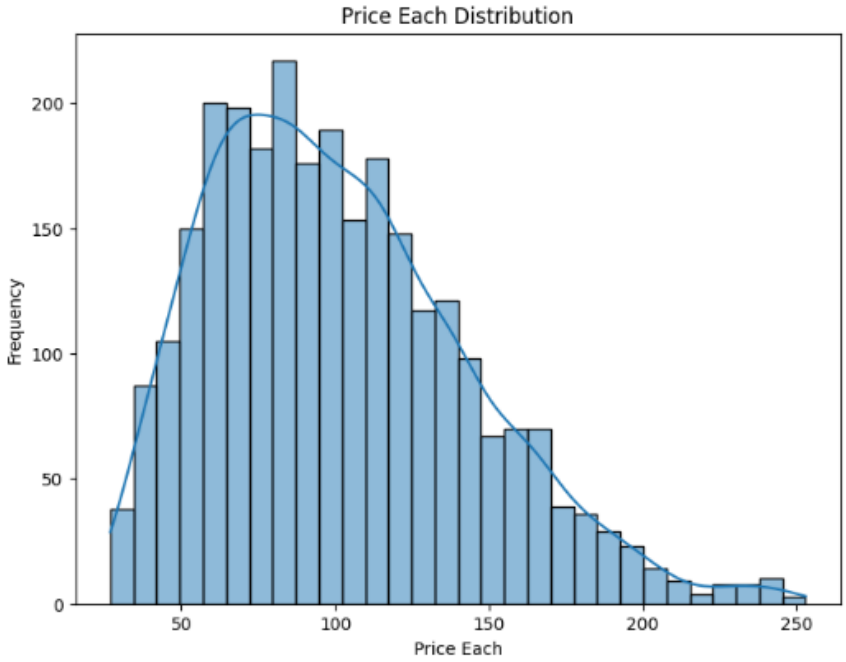
Univariate analysis



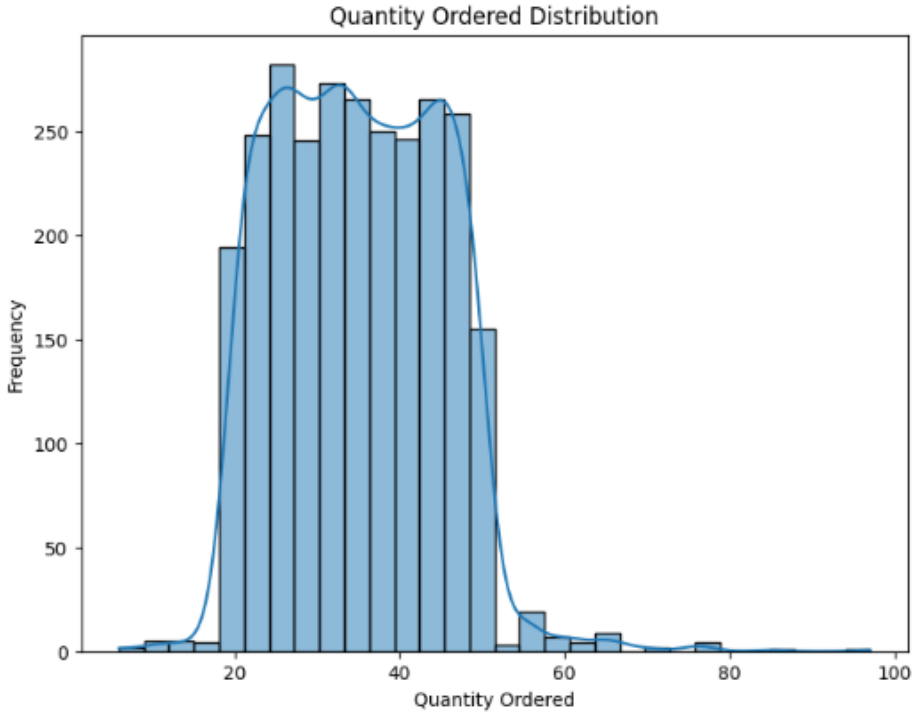
- The distribution of sales is positively skewed, indicating that the majority of transactions have lower sales values. However, there is a long tail on the right side of the distribution, which suggests the presence of some outliers with significantly higher sales values.



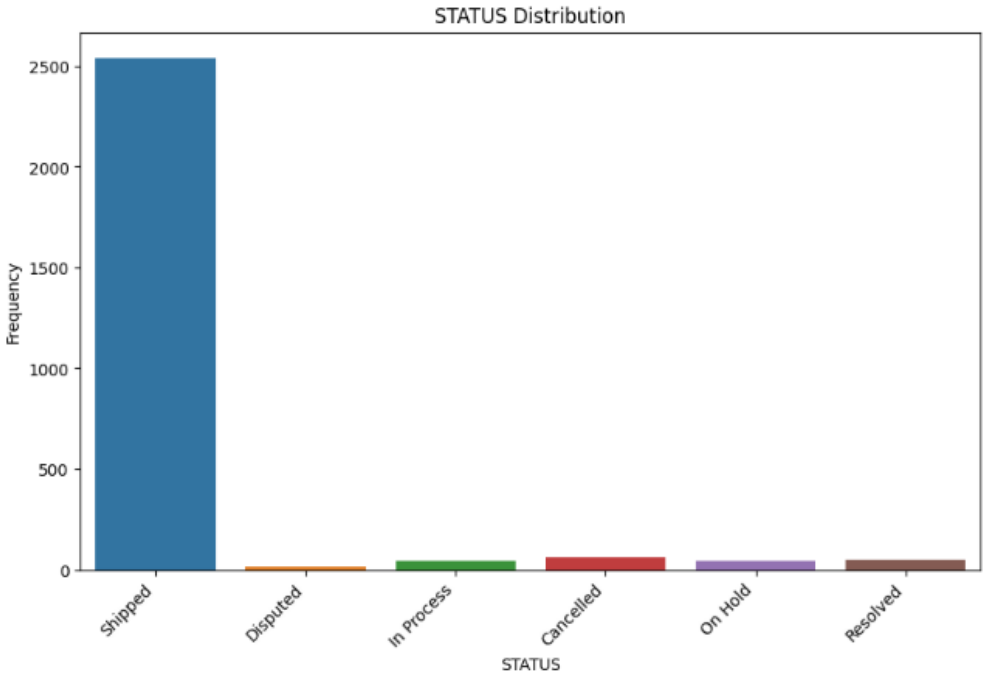
- In the dataset, the frequency of days since the last order is relatively uniform across the range of 800 to 2700. This means that there are similar occurrences of customers making purchases with varying intervals between 800 and 2700 days.



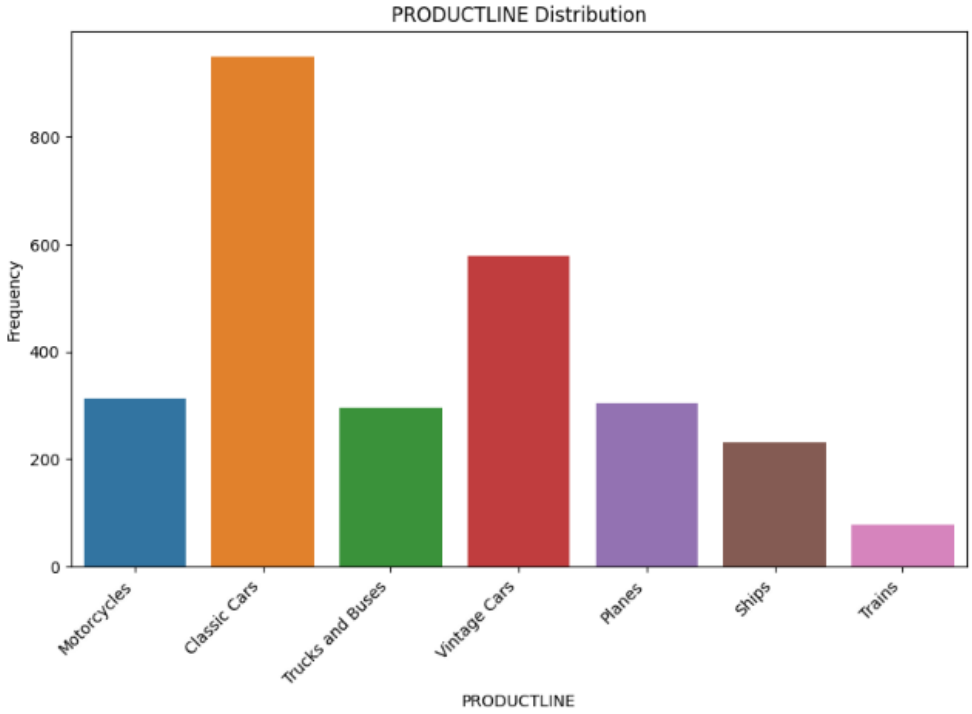
- The distribution of the "price-each" column is slightly right-skewed, indicating that most orders have relatively lower prices. As a result, there is a higher frequency of orders with lower prices compared to those with higher prices



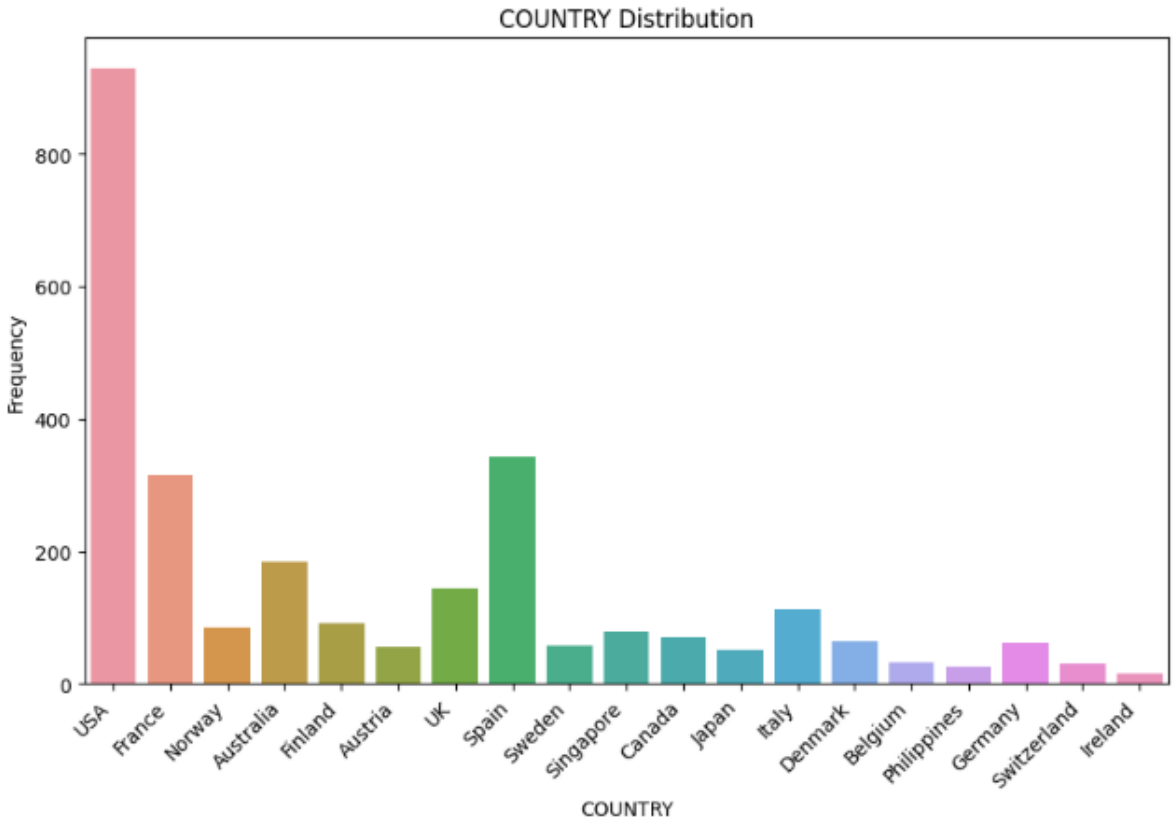
- In the dataset, quantities ranging from 20 to 50 exhibit a similar frequency, meaning that there are nearly equal occurrences of orders with quantities in this range.



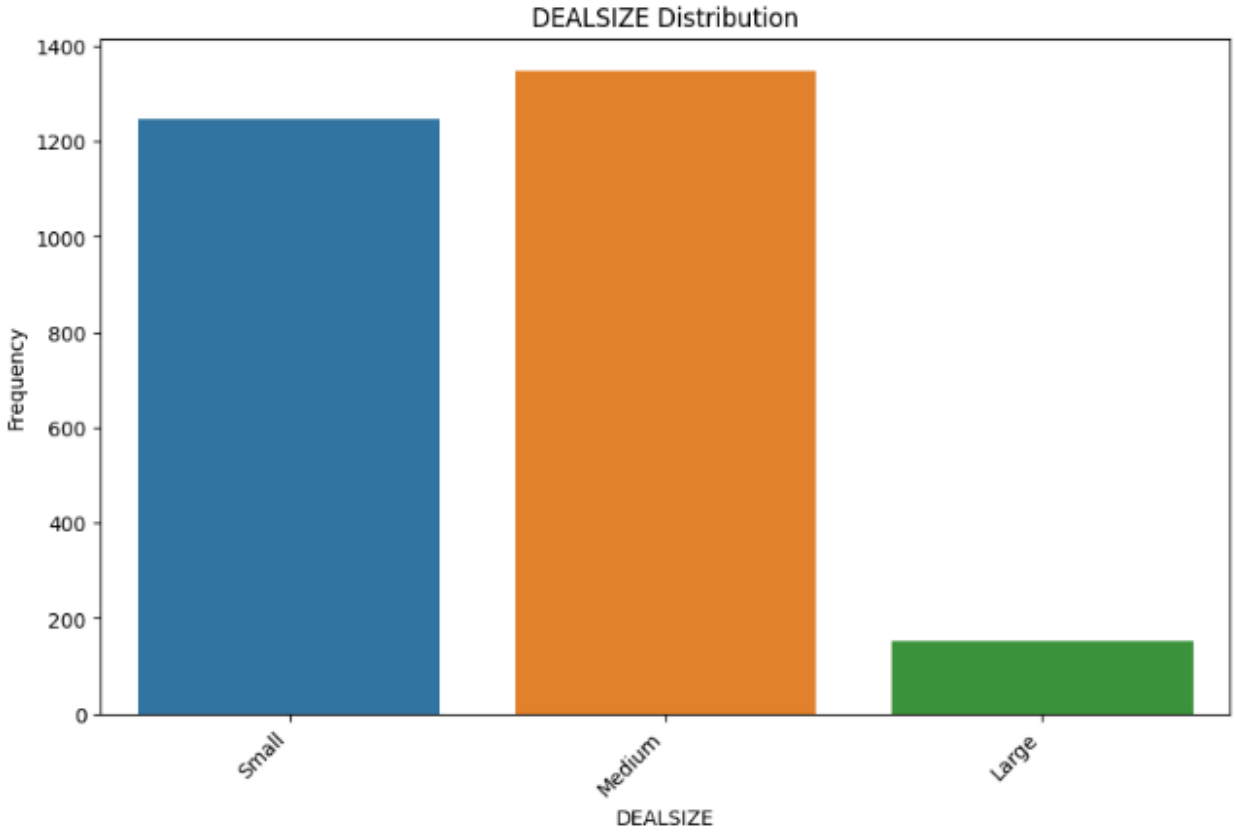
- The majority of the orders in the dataset have already been shipped.



- A large proportion of buyers in the dataset show a preference for vintage and classic cars.



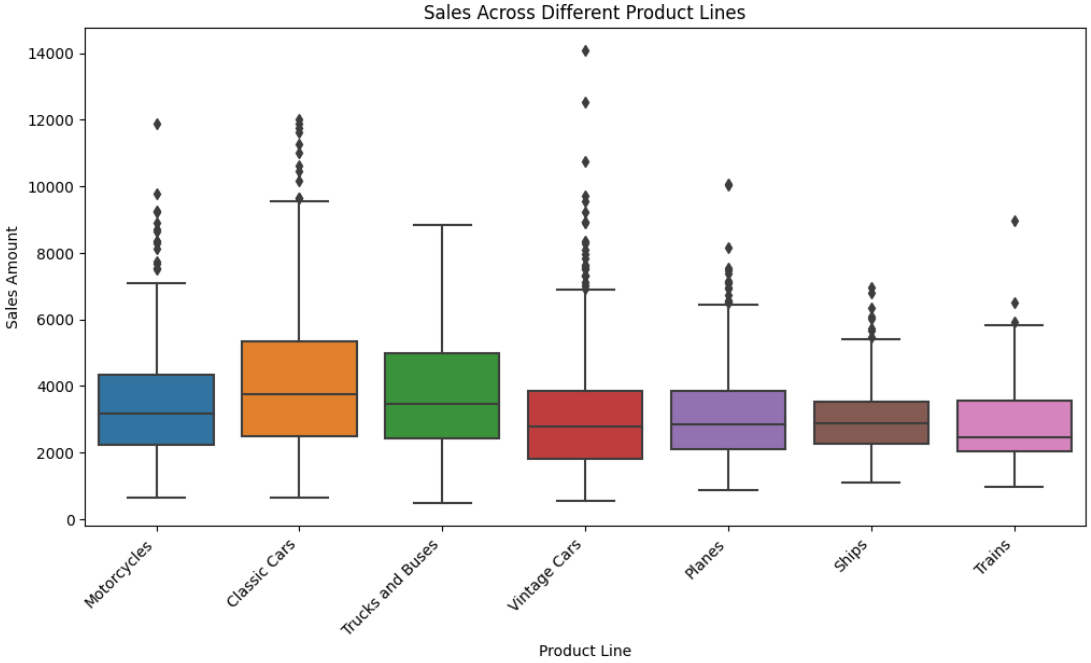
- Most of the customers are from USA



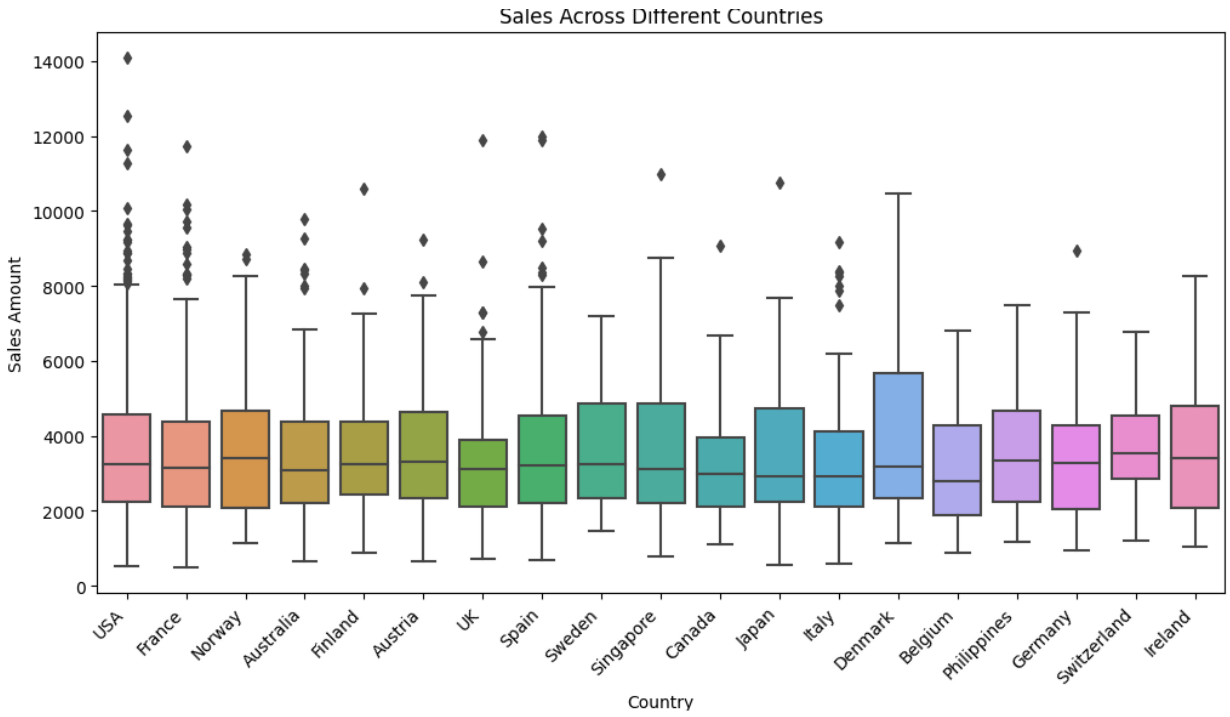
- The frequency of large deal sizes is lower compared to medium and small deal sizes in the dataset.

Bivariate analysis

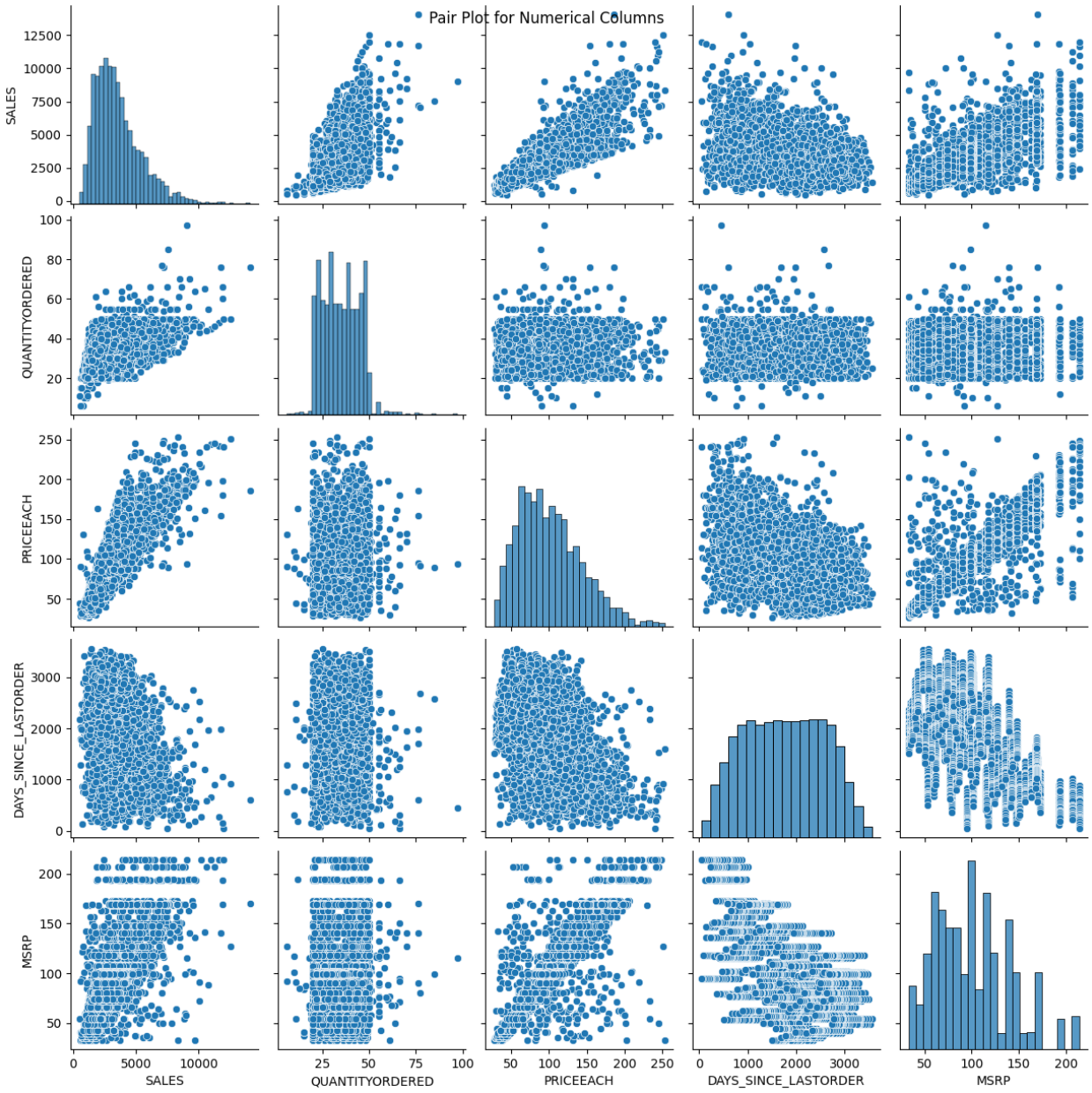
- There is a linear relationship between the "sales" and "price-each" columns in the dataset. As the price of each item increases, the sales also tend to increase proportionally.



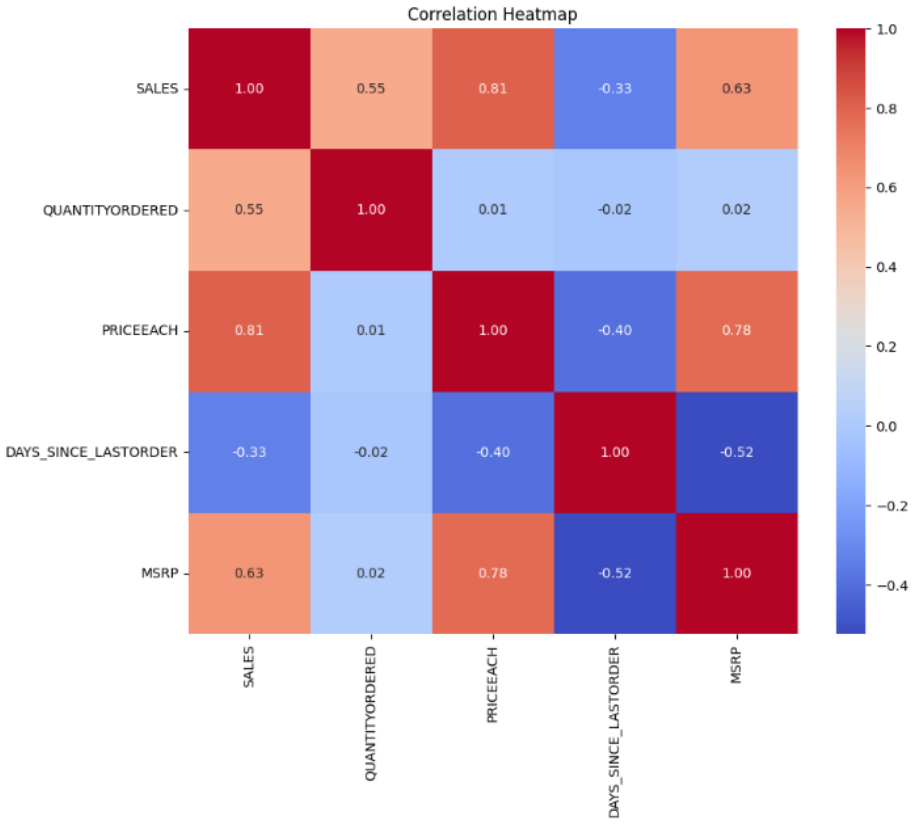
- Big sales sums can be seen for all product line categories.



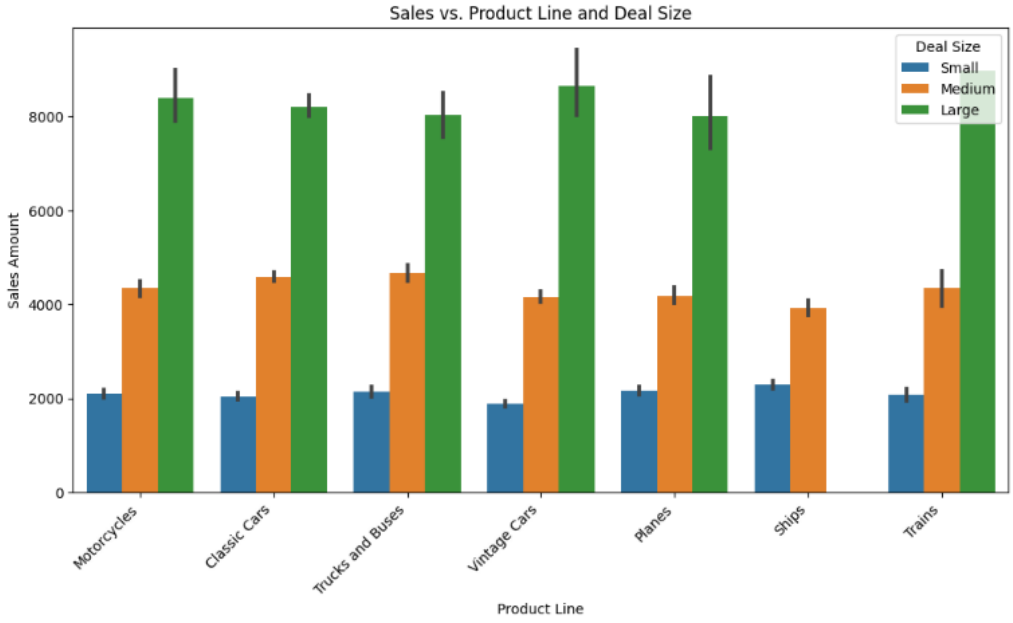
- Distribution of sales amount is almost similar in all countries.



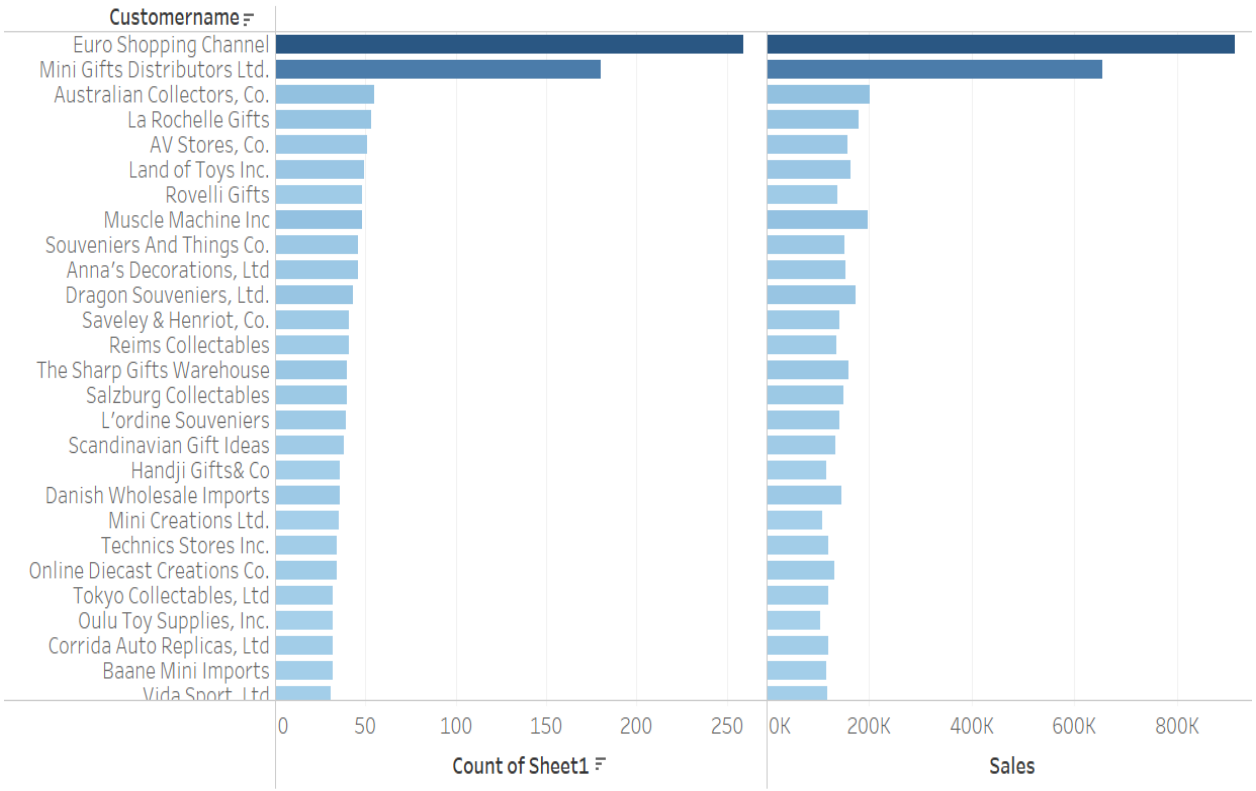
- All distributions can be seen here and patterns between columns too. Good linear patterns can be seen in price each and MSRP column.



- All four columns quality ordered, price each, days since last order and MSRP has a good correlation with sales column, and with each other too they have good correlation.

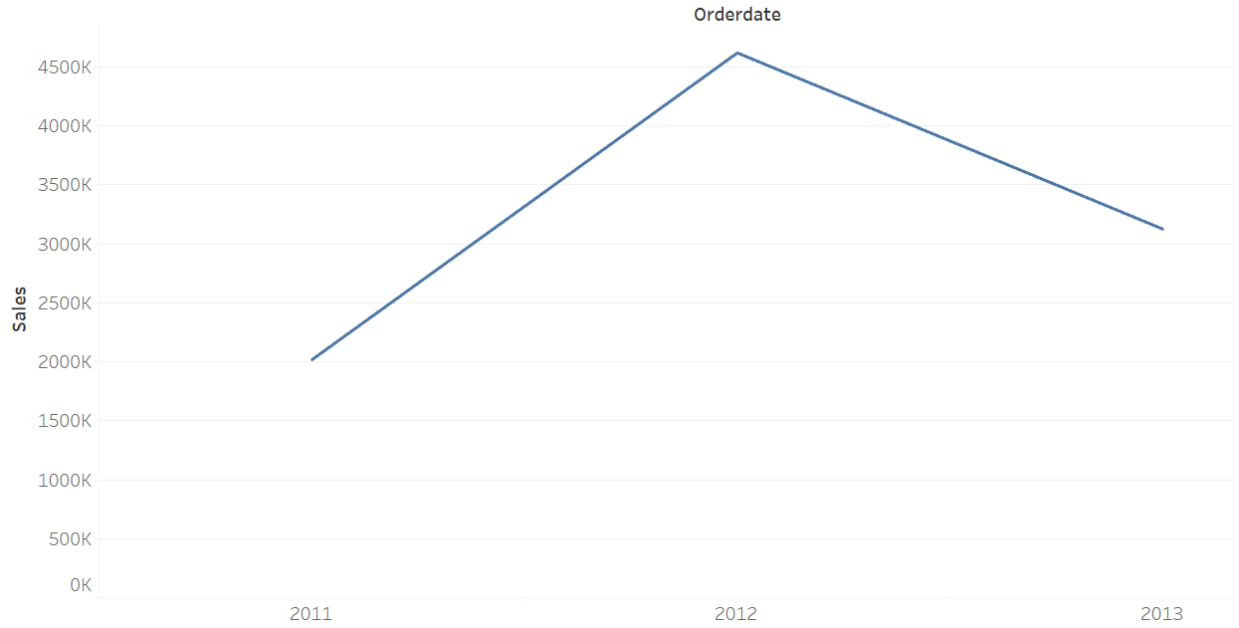


- Although the frequency of large orders is low, they still contribute to most of the sales.

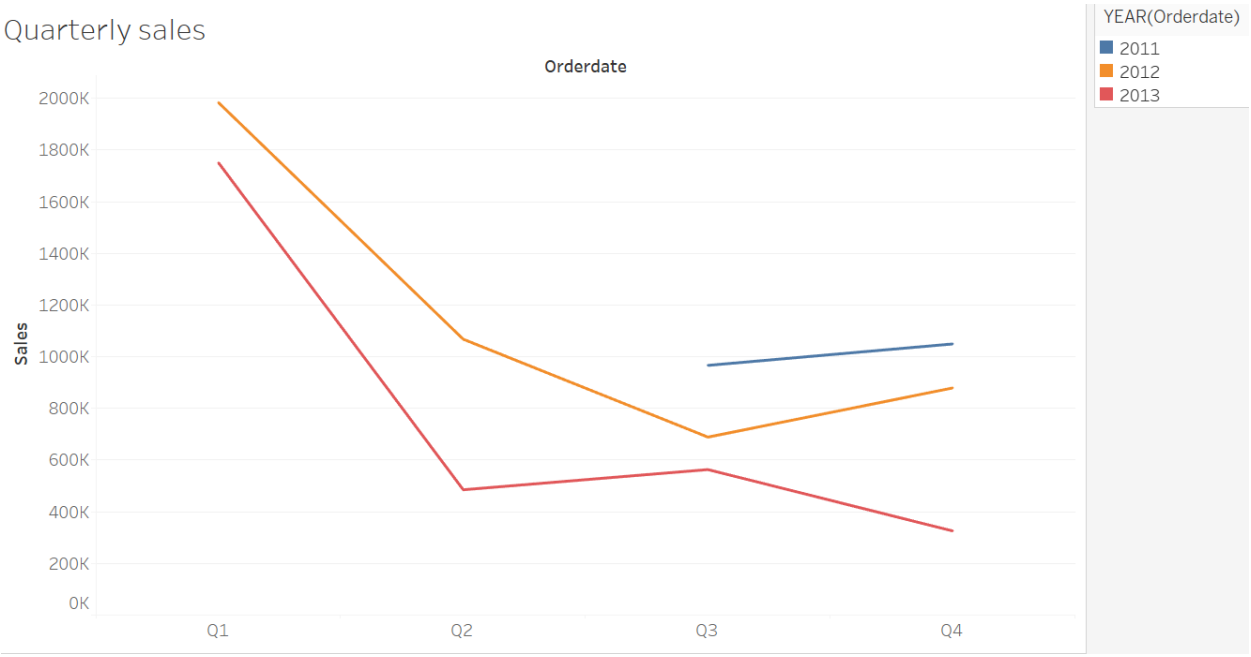


- The Euro Shipping channel and Mini Gifts Distribution Ltd. are the primary contributors to the maximum sales in the dataset.

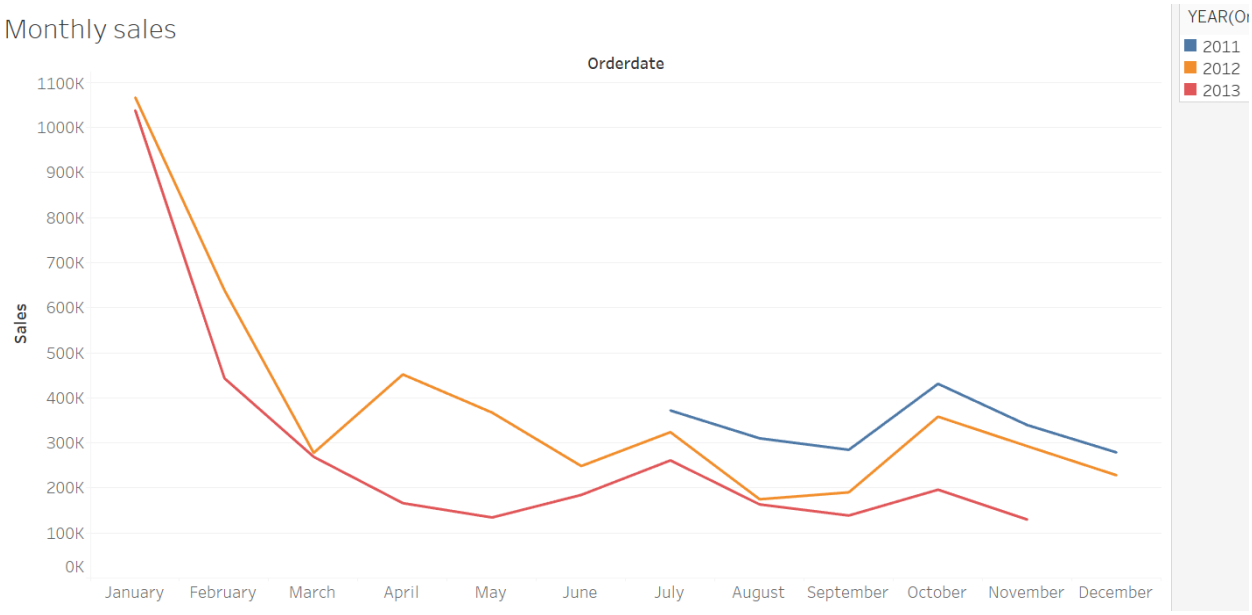
Yearly sales



- We got maximum sales in year 2012 , lowest in 2011 and medium sales in 2013



- The sales for both the years 2012 and 2013 are at their highest in Q1 (January to March), and then the trend shows a gradual decrease in Q2 (April to June), Q3 (July to September), and Q4 (October to December). This pattern indicates that the sales peak in the first quarter and gradually decline throughout the rest of the year for both years.



- The overall trend of sales has shown a decline over the passing years. Sales in each month of 2011 and 2012 were relatively higher compared to each month of 2013. The highest sales were recorded in January, followed by a significant drop in February. As the months pass by, the trend shows a slight flattening up, but overall, the sales have been decreasing year by year.

Summary-

1. Sales data is right-skewed, indicating some outliers in the dataset.

2. The distribution of "Days since last order" ranges from 800 to 2700 and shows similar frequency.

3. The "Price Each" column follows a bit right-skewed distribution, with higher frequency for orders with relatively lower prices.

4. Quantities ranging from 20 to 50 have almost the same frequency in the data.

5. Most of the orders have already been shipped, indicating a high fulfillment rate.

6. Vintage and classic cars are the most sought-after product line categories by the buyers.

7. The majority of customers are from the USA, indicating a significant customer base from this region.

8. Large deal sizes have a lower frequency compared to medium and small deal sizes.

9. There is a linear relationship between the "Sales" and "Price Each" columns.

10. Big sales sums are observed for all product line categories, and the distribution of sales amount is similar across different countries.

11. Euro Shipping Channel and Mini Gifts Distribution Ltd. contribute to the maximum sales.

12. Sales are highest in Q1 for both years 2012 and 2013, followed by a decline in Q2, Q3, and Q4.

13. The overall trend of sales has been decreasing over the years, with the highest sales in January, followed by a drop in February. The trend flattens out as the months progress.

Customer Segmentation using RFM analysis (4 segments) -> What is RFM? -> What all parameters used and assumptions made? -> Showcase the KNIME workflow image -> What results are there in the output table head?

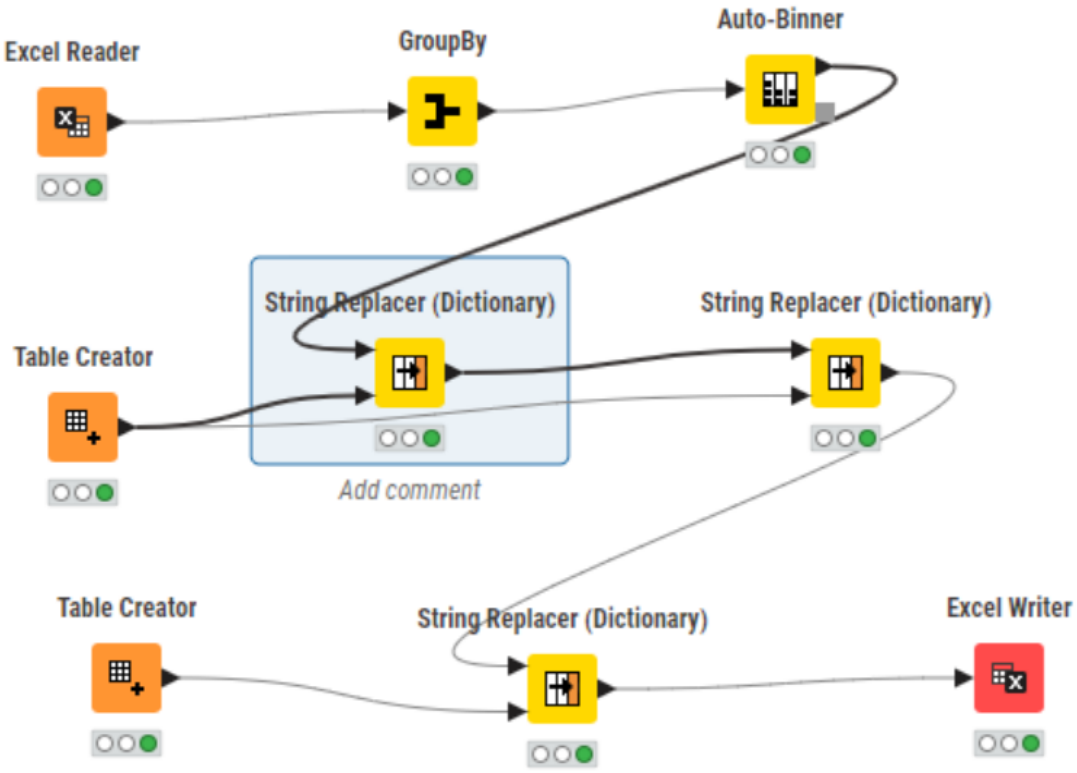
- **What is RFM?**
    - RFM stands for Recency, Frequency, and Monetary Value, and it is a popular marketing analysis technique used to segment customers based on their purchasing behavior. RFM analysis helps businesses identify different groups of customers and tailor marketing strategies accordingly.

- **Here's a brief explanation of each component in RFM:**
    - **Recency (R):** This refers to how recently a customer has made a purchase. It is a measure of how much time has elapsed since the customer's last purchase. Customers who have made recent purchases are generally considered more engaged and active.
    - **Frequency (F)**: This indicates how often a customer makes purchases within a specific time period. It is a measure of how frequently a customer interacts with the business. Customers who make frequent purchases are typically more loyal and valuable.
    - **Monetary Value (M)**: This represents the total monetary value of a customer's purchases within a specific time period. It is a measure of how much a customer spends on average. Customers with higher monetary value are considered high spenders and are valuable to the business.

- **What all parameters used and assumptions made?**
    - Sales column is used as a monetary value, order-number is used as frequency column and days_since_last_order is used as a recency column required for RFM analysis.
    - We have grouped data by customer name to get frequency.
    - We have segmentented the data as H- High , M- Medium and L- Low for all three columns, that are Recency , frequency and monetary.

- Also we have used the logic that values till 25th percentile will be considered as low for frequency and monetary column, 25th to 75th percentile as medium and above 75th percentile values are considered as high. .
- For the recency analysis, we classified the "Days since last order" column into three categories based on their respective percentile values. Orders with a value up to the 25th percentile were labeled as "high" recency, orders with values between the 25th and 75th percentile were labeled as "medium" recency, and orders with values above the 75th percentile were labeled as "low" recency.
- By converting the numerical values of "recency," "frequency," and "monetary" into categorical values (H, M, or L) for each customer name, we have effectively created three new columns.
- Then we clubbed these 3 columns together to give a tag as 'HHH','HMM','MLL' etc. to different clients and named the column as RFM.

- **Showcase the KNIME workflow image**



- We have used groupby to group data by customer_name , then we binned the data according to percentiles and then we converted those BIN1,BIN2,BIN3 tags to H, M and L with the help of table creator and string replacer and exported the data after wards.

- **What results are there in the output table head?**

Rows: 89 | Columns: 20 — Table | Statistics

#	Row...	CUSTOMERNAME String	ORDER... Number (int...	QUANTI... Number (int...	PRICEE... Number (do...	ORDERL... Number (int...	SALES Number (do...	ORDER... Number (int...	DAYS_S... Number (int...
1	Row0	AV Stores, Co.	51	1778	4,645.31	51	157,807.81	51	421
2	Row1	Alpha Cognac	20	687	2,023.2	20	70,488.44	20	675
3	Row2	Amica Models & Co.	26	843	2,882.17	26	94,117.26	26	328
4	Row3	Anna's Decorations, Ltd	46	1469	4,895.51	46	153,996.13	46	131
5	Row4	Atelier graphique	7	270	645.67	7	24,179.96	7	312
6	Row5	Australian Collectables, Ltd	23	705	2,070.96	23	64,591.46	23	1018
7	Row6	Australian Collectors, Co.	55	1926	5,752.46	55	200,995.41	55	229
8	Row7	Australian Gift Network, Co	15	545	1,658.31	15	59,469.12	15	190
9	Row8	Auto Assoc. & Cie.	18	637	1,790.78	18	64,834.32	18	275
10	Row9	Auto Canal Petit	27	1001	2,544.89	27	93,170.66	27	127
11	Row10	Auto-Moto Classics Inc.	8	287	742.4	8	26,479.26	8	1353

- **Output table - Grouped data by customer_name.**

- **Lets check out data after binning-**

ORDERNUMBER... String	SALES [Binned] String	DAYS_SINCE_LASTORDER String
Bin 3	Bin 3	Bin 2
Bin 1	Bin 1	Bin 3
Bin 2	Bin 2	Bin 2
Bin 3	Bin 3	Bin 1
Bin 1	Bin 1	Bin 2
Bin 2	Bin 1	Bin 3
Bin 3	Bin 3	Bin 1
Bin 1	Bin 1	Bin 1
Bin 1	Bin 1	Bin 2
Bin 2	Bin 2	Bin 1
Bin 1	Bin 1	Bin 3

- Bins have been made.

- **Lets convert then to H, M and L tags**

ORDERNUMBE... *String*	SALES [Binned]Monet... *String*	DAYS_SINCE_LA... *String*
H	H	M
L	L	L
M	M	M
H	H	H
L	L	M
M	L	L
H	H	H
L	L	H
L	L	M
M	M	H

    - We have required tags now for each customer

Inferences from RFM Analysis and identified segments -> Who are your best customers? (give at least 5) -> Which customers are on the verge of churning? (give at least 5) -> Who are your lost customers? (give at least 5) -> Who are your loyal customers? (give at least 5)

- **After segmentation this is the matrix we got-**

Recency	Frequency	Monetary H	Monetary L	Monetary M
H	H	11.0	0.0	0.0
	L	0.0	1.0	1.0
	M	1.0	0.0	9.0
L	H	2.0	0.0	0.0
	L	0.0	10.0	1.0
	M	0.0	1.0	8.0
M	H	7.0	0.0	2.0
	L	0.0	10.0	4.0
	M	1.0	1.0	19.0

MMM	19
HHH	11
MLL	10
LLL	10
HMM	9
LMM	8
MHH	7
MLM	4
LHH	2
MHM	2
HLL	1
LML	1
MML	1
MMH	1
HMH	1
LLM	1
HLM	1

- **MMM (Recency = Medium, Frequency = Medium, Monetary = Medium):**
    - Suggestion: This segment represents moderately engaged and moderately spending customers.
    - Marketing Strategy: Focus on maintaining their current level of engagement and spending. Offer personalized product recommendations based on their previous purchases to encourage repeat business.

- **HHH (Recency = High, Frequency = High, Monetary = High):**
    - Suggestion: These are the most valuable and active customers for the business.
    - Marketing Strategy: Show appreciation and reward their loyalty with exclusive offers, VIP access, or loyalty programs. Cross-sell complementary products to maximize their spending.

- **MLL (Recency = Medium, Frequency = Low, Monetary = Low):**
    - Suggestion: This segment needs to be re-engaged to increase their spending.
    - Marketing Strategy: Send targeted emails or offers to entice them back to the store. Offer discounts or promotions on popular or new products to pique their interest.

- **LLL (Recency = Low, Frequency = Low, Monetary = Low):**
    - Suggestion: This segment makes infrequent and small purchases.
    - Marketing Strategy: Create limited-time offers or bundle deals to encourage more frequent purchases. Implement a loyalty program to reward repeat purchases.

- **HMM (Recency = High, Frequency = Medium, Monetary = Medium):**
    - Suggestion: These customers are recent but not making frequent high-value purchases.
    - Marketing Strategy: Offer time-limited promotions or discounts to encourage them to make more purchases. Provide personalized product recommendations to increase their order value.

- **LMM (Recency = Low, Frequency = Medium, Monetary = Medium):**
    - Suggestion: These customers make moderate purchases but have low recency.
    - Marketing Strategy: Send personalized thank-you emails and offer incentives to encourage them to come back sooner. Provide exclusive deals on their favorite products.

- **MHH (Recency = Medium, Frequency = High, Monetary = High):**
    - Suggestion: These customers have moderate recency but make frequent high-value purchases.
    - Marketing Strategy: Provide personalized offers based on their preferences and purchase history. Offer loyalty rewards to encourage them to keep coming back.

- **MLM (Recency = Medium, Frequency = Low, Monetary = Medium):**
    - Suggestion: This segment has moderate recency and spending but makes infrequent purchases.
    - Marketing Strategy: Offer product recommendations based on their previous purchases. Implement a referral program to encourage them to refer friends and increase their engagement.

- **Who are your best customers?**
    - Customers with tags such as "HHH" (high recency, high frequency, high monetary value) are considered the best customers for the company. These customers have made recent, frequent, and high-value purchases, indicating that they are highly engaged and valuable to the business. They contribute significantly to the total sales revenue and are likely to be loyal and repeat customers

	Frequency	Recency	Monetary	CUSTOMERNAME	DAYS_SINCE_LASTORDER	ORDERNUMBER	SALES	RFM
32	H	H	H	Euro Shopping Channel	42	259	912294.11	HHH
53	H	H	H	Mini Gifts Distributors Ltd.	219	180	654858.06	HHH
6	H	H	H	Australian Collectors, Co.	229	55	200995.41	HHH
43	H	H	H	La Rochelle Gifts	139	53	180124.90	HHH
44	H	H	H	Land of Toys Inc.	216	49	164069.44	HHH

- **Which customers are on the verge of churning?**
    - Customers with the "LMM" tag (low recency, medium frequency, medium monetary value) and sorted in decreasing order of days since their last order might be on the verge of churning. The "LMM" segment represents customers who used to make medium-value purchases at a moderate frequency but have not visited the store or made a purchase recently.
    - The fact that they have a low recency value and have not made a purchase for a considerable period suggests a decline in their engagement with the company. This change in behavior could indicate that they are becoming less interested in the products or services offered by the company and may be at risk of churning.

	Frequency	Recency	Monetary	CUSTOMERNAME	DAYS_SINCE_LASTORDER	ORDERNUMBER	SALES	RFM
24	M	L	M	Cruz & Sons Co.	971	26	94015.73	LMM
56	M	L	M	Norway Gifts By Mail, Co.	825	24	79224.23	LMM
46	M	L	M	Marseille Mini Autos	757	25	74936.14	LMM
13	M	L	M	Blauer See Auto, Co.	705	22	85171.59	LMM
73	M	L	M	Stylish Desk Decors, Co.	702	26	88804.50	LMM

- **Who are your lost customers?**
    - Your lost customers are those with the "LLL" tag (low recency, low frequency, low monetary value). These customers have the lowest engagement with your company. They have not visited your store or made a purchase frequently, and when they did, their spending and number of orders were also low.
    - Given their minimal interaction and low value to the company, it may not be worthwhile to put significant effort into retaining these customers

	Frequency	Recency	Monetary	CUSTOMERNAME	DAYS_SINCE_LASTORDER	ORDERNUMBER	SALES	RFM
10	L	L	L	Auto-Moto Classics Inc.	1353	8	26479.26	LLL
35	L	L	L	Gift Ideas Corp.	947	19	57294.42	LLL
40	L	L	L	Iberia Gift Imports, Corp.	904	15	54723.62	LLL
70	L	L	L	Signal Collectibles Ltd.	836	15	50218.51	LLL
12	L	L	L	Bavarian Collectables Imports, Co.	801	14	34993.92	LLL

- **Who are your loyal customers?**
    - These customers have the highest frequency of orders, indicating that they visit your store frequently and make purchases regularly. They contribute significantly to your company's revenue and are likely to be satisfied with your products or services, leading to their continued loyalty.

	Frequency	Recency	Monetary	CUSTOMERNAME	DAYS_SINCE_LASTORDER	ORDERNUMBER	SALES	RFM
32	H	H	H	Euro Shopping Channel	42	259	912294.11	HHH
53	H	H	H	Mini Gifts Distributors Ltd.	219	180	654858.06	HHH
6	H	H	H	Australian Collectors, Co.	229	55	200995.41	HHH
43	H	H	H	La Rochelle Gifts	139	53	180124.90	HHH
0	H	M	H	AV Stores, Co.	421	51	157807.81	MHH