

Exploratory Data Analysis and RFM Analysis for Customer Insights

Problem Statement

A thriving automotive parts manufacturer struggles to transform its rich transaction data into strategic business insights due to limited internal analytics expertise.

Objective

- Uncover customer purchasing patterns
- Develop data-driven customer segmentation
- Generate actionable marketing insights
- Craft personalized marketing strategies to boost sales and retention

Summary Statistics

	count	mean	std	min	25%	50%	75%	max
QUANTITYORDERED	2747.00000	35.10302	9.76214	6.00000	27.00000	35.00000	43.00000	97.00000
PRICEEACH	2747.00000	101.09895	42.04255	26.88000	68.74500	95.55000	127.10000	252.87000
SALES	2747.00000	3553.04758	1838.95390	482.13000	2204.35000	3184.80000	4503.09500	14082.80000
NoOfDaysSincelastOrder	2747.00000	383.08591	230.23130	0.00000	196.00000	342.00000	570.00000	876.00000
MSRP	2747.00000	100.69166	40.11480	33.00000	68.00000	99.00000	124.00000	214.00000

Order Quantity

- Avg. 35 items per transaction
- 50% of orders: 27-43 items

Sales Performance

- Avg. order value: \$3,553.05
- Sales range: \$482.13 - \$14,082.80
- Median sales: \$2,204.35 - \$4,503.10

MSRP Overview

- Avg. MSRP: \$100.69
- MSRP range: \$33 - \$214
- Median MSRP: \$68 - \$124

Pricing Analysis

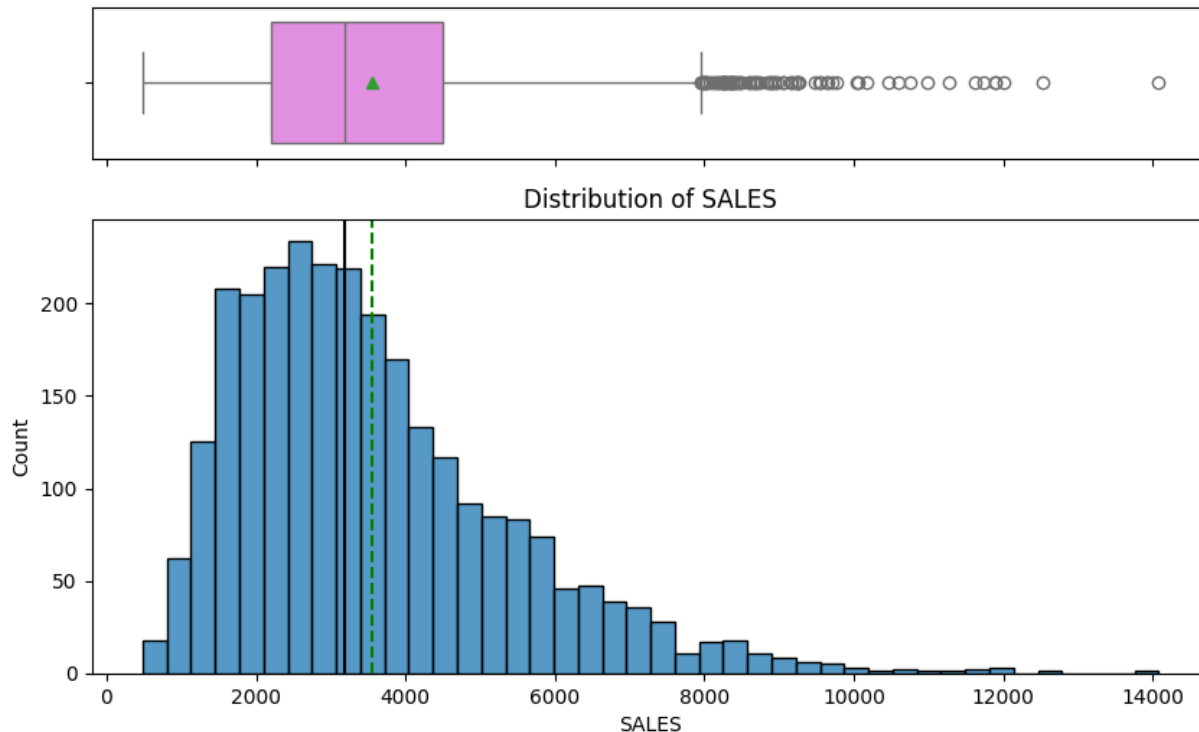
- Avg. unit price: \$101.10
- Price range: \$26.88 - \$252.87
- Median price: \$68.75 - \$127.10

Order Frequency

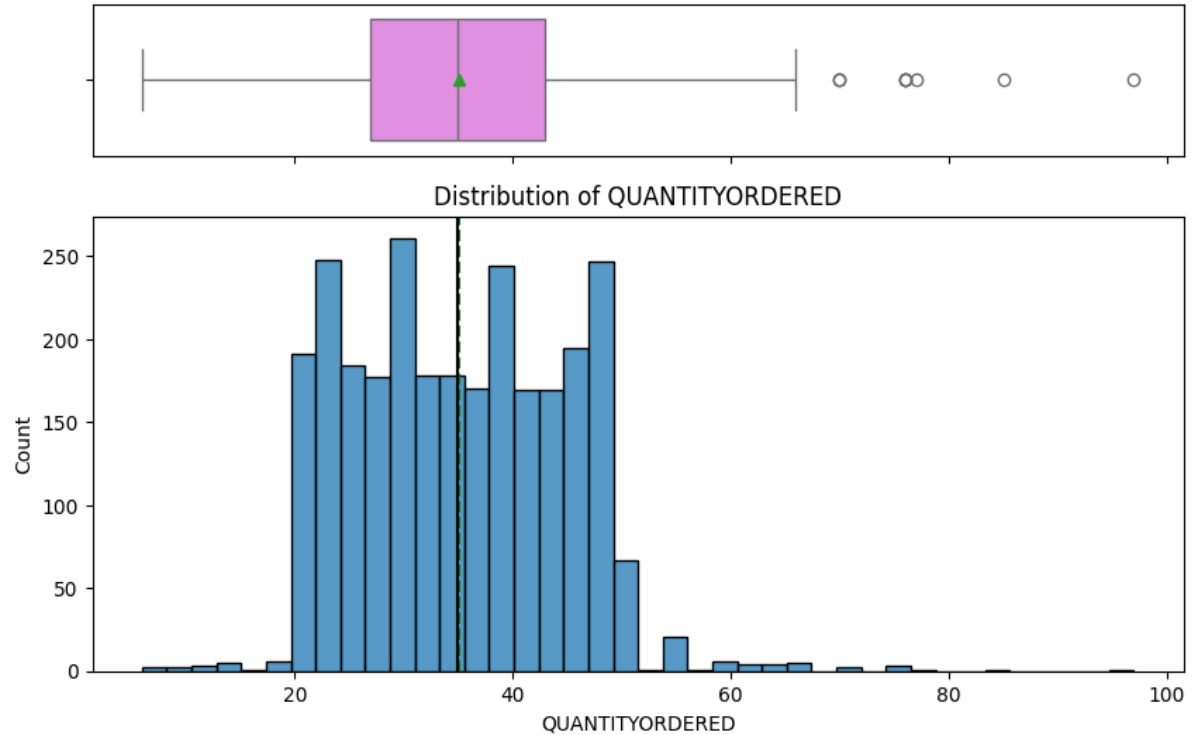
- Avg. days between orders: 383 days
- Interval range: 0-876 days
- Median interval: 196-570 days

Sales

- The sales distribution is positively skewed, indicating a concentration of moderate-sized sales with some high-volume outliers.
- The majority of sales fall within the 4,000-6,000 units range, with a peak around 5,000 units.
- There is a long tail of lower sales values, with a gradual decline in frequency as sales increase beyond 6,000 units.



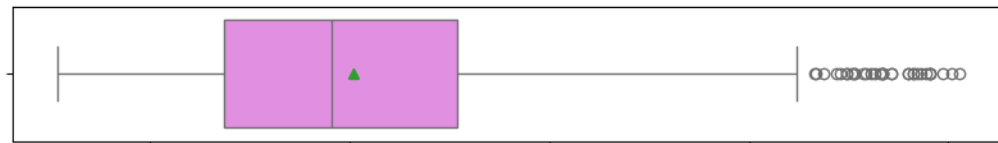
Quantity Ordered



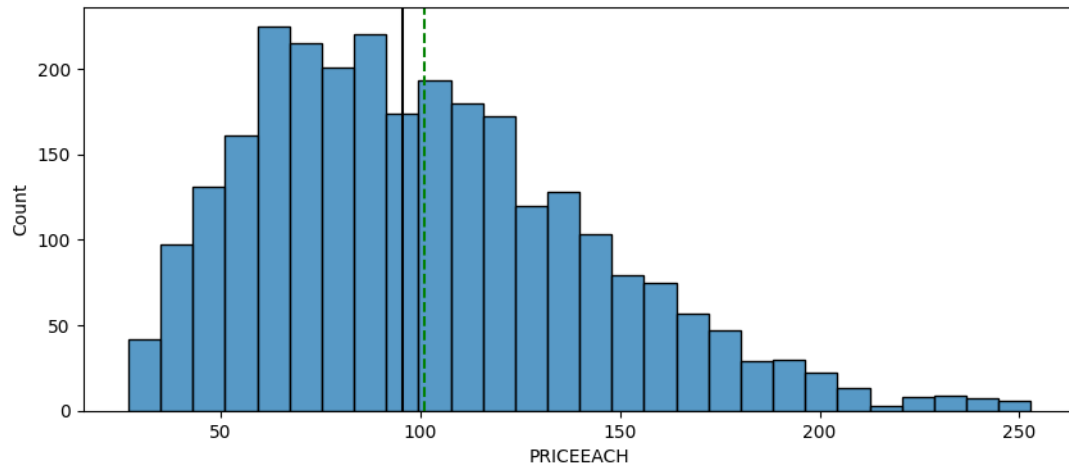
- The distribution of quantities ordered shows a peak around 30-40 items per order, with a substantial number of larger orders as well.
- There is a long right-tail in the distribution, indicating the presence of some very high-volume orders.
- 50% of orders fall between 27-43 items, suggesting a core group of customers making moderately-sized purchases.

Price Each

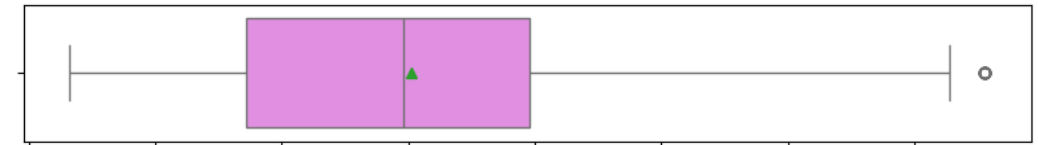
- Prices span a wide range, from around \$30 to over \$250 per unit.
- The distribution exhibits a long right tail, indicating some high-priced outliers.
- Approximately 50% of prices fall between \$68.75 and \$127.10, suggesting a concentrated core pricing range.
- The average price per unit is \$101.10.



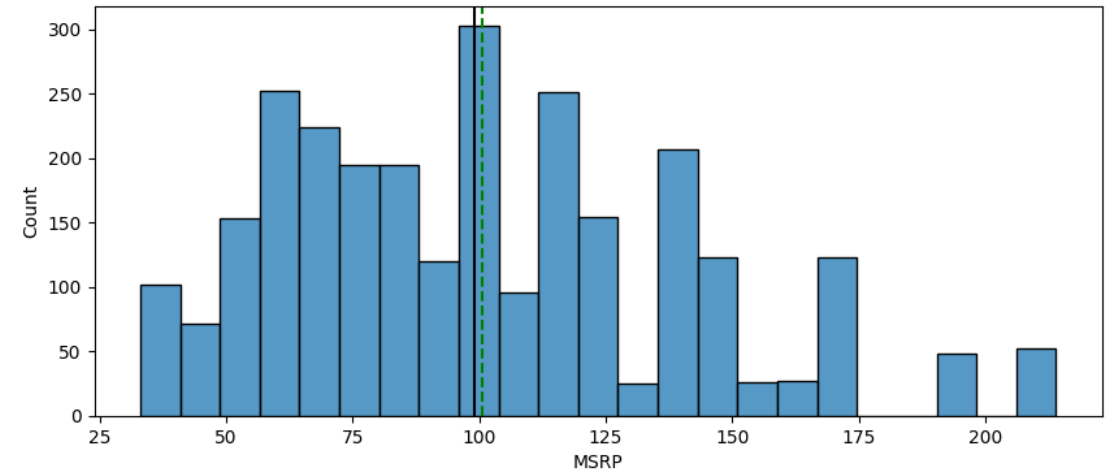
Distribution of PRICEEACH



MSRP

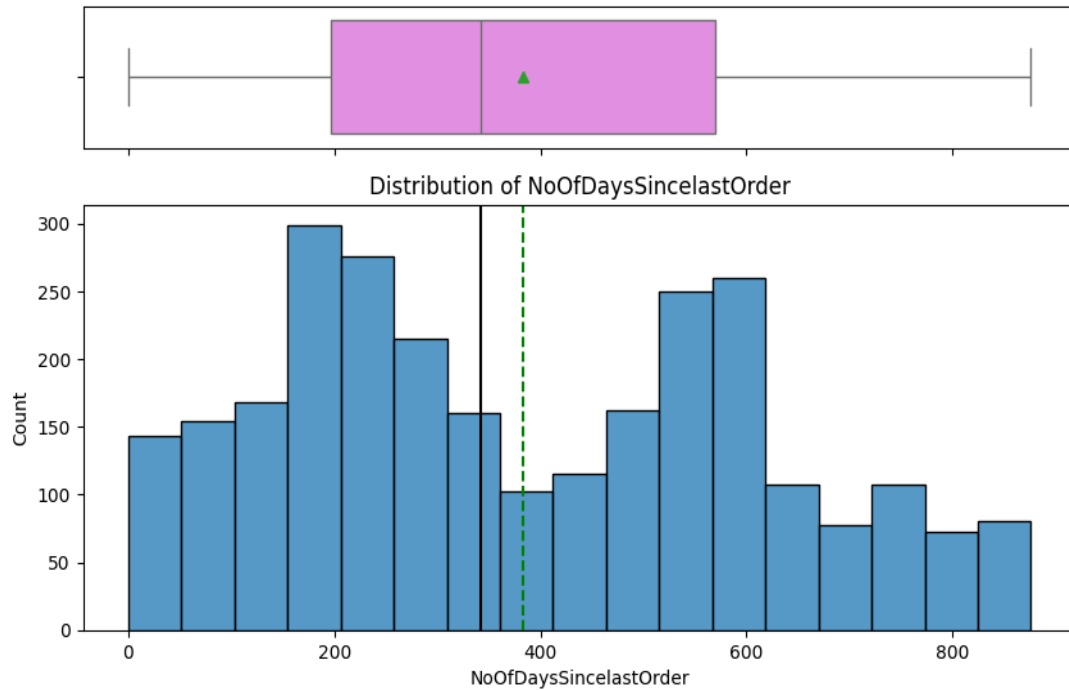


Distribution of MSRP



- MSRP ranges from \$33 to \$214, with an average of \$100.69.
- Approximately 50% of MSRP values fall between \$68 and \$124.
- The distribution shows a prominent peak around \$100, suggesting a core group of products priced in this mid-range.
- There are also some higher-priced outliers beyond \$150.

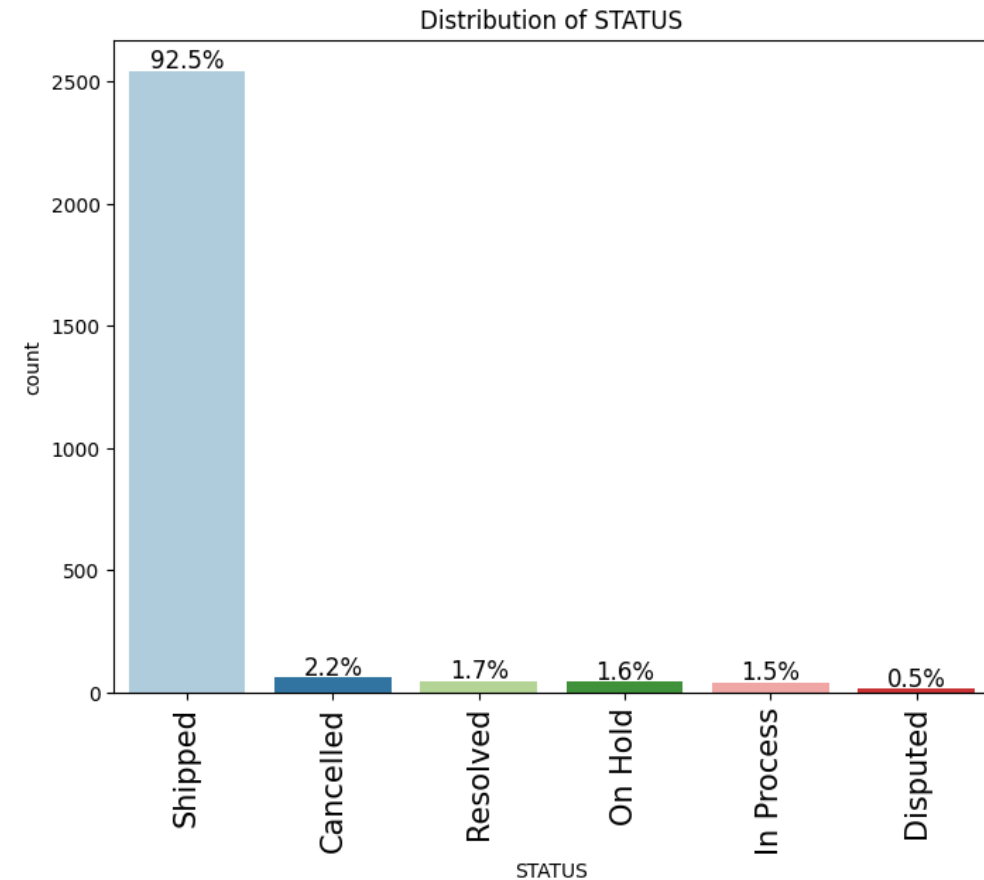
Days since last Order



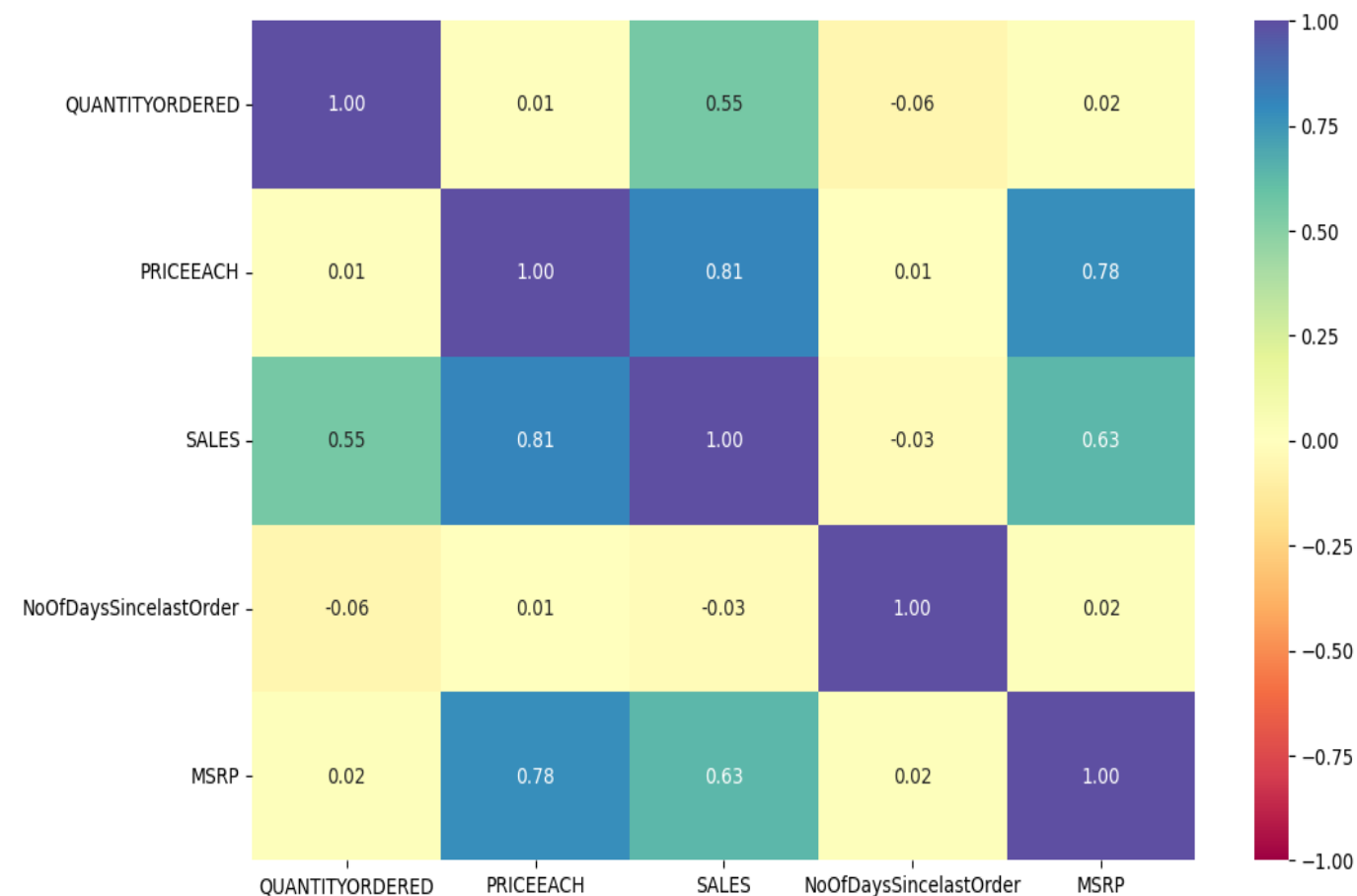
- The distribution shows a wide range in the number of days since a customer's last order, from 0 days up to nearly 900 days.
- The majority of orders fall within the 196-570 day range, suggesting a core group of customers making purchases every 6 months to 1.5 years on average.
- There are also a substantial number of customers with very recent orders, within the past 200 days.
- The long right tail indicates some customers have not placed an order in several years.

Order Status

- The vast majority (92.5%) of orders are in the "Shipped" status.
- Small percentages are in the "Cancelled" (2.2%), "Resolved" (1.7%), "On Hold" (1.6%), and "In Process" (1.5%) statuses.
- Only 0.5% of orders are in the "Disputed" status.



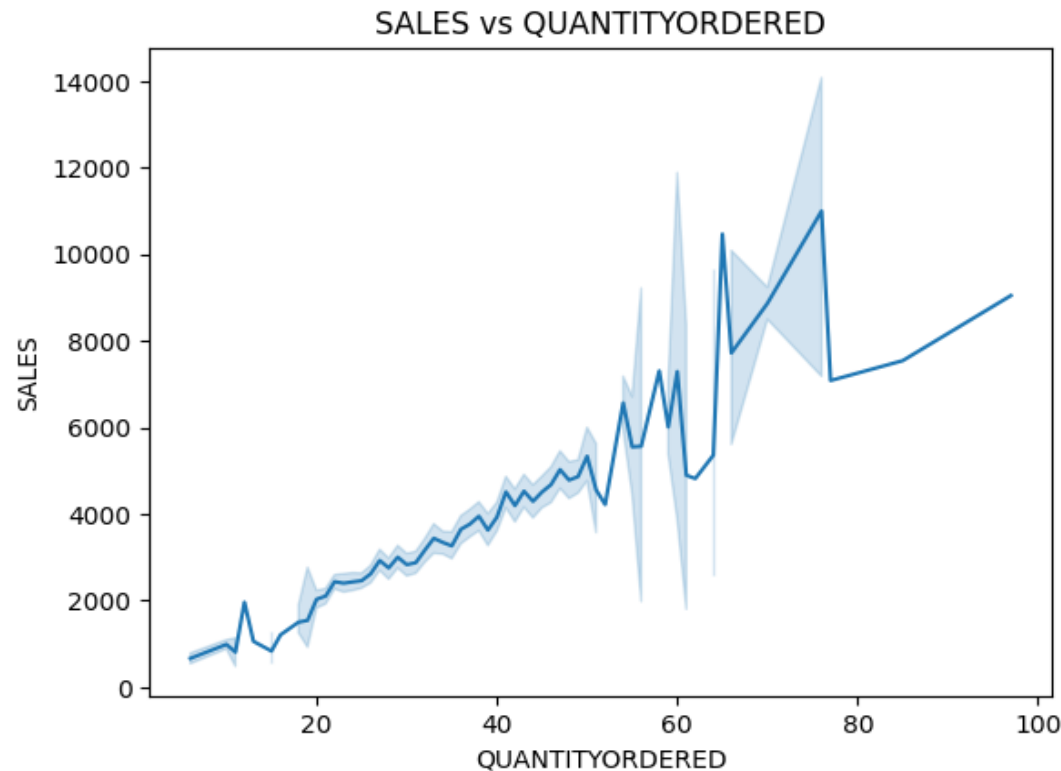
Bivariate Analysis



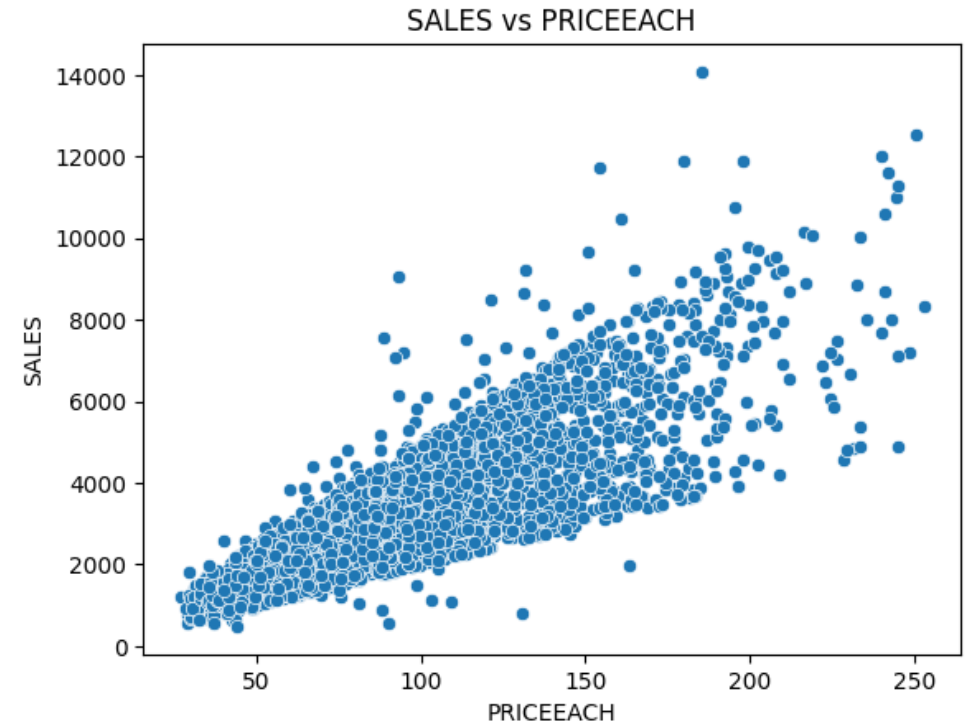
- Quantity Ordered strongly positively correlated (1.0)
- Price Each moderately positively correlated (0.81) with Sales
- Sales moderately positively correlated (0.63) with MSRP
- Time since last order strongly negatively correlated (-0.06) with Quantity Ordered
- Time since last order weakly negatively correlated (-0.03) with Sales

Sales vs Quantity Ordered

- Sales and Quantity Ordered exhibit a strong positive correlation, with sales increasing as order quantities rise.
- The relationship shows a steep, nonlinear trend, with a sharp spike in sales for larger order sizes.
- This suggests high-volume orders are a key driver of overall revenue for the company.



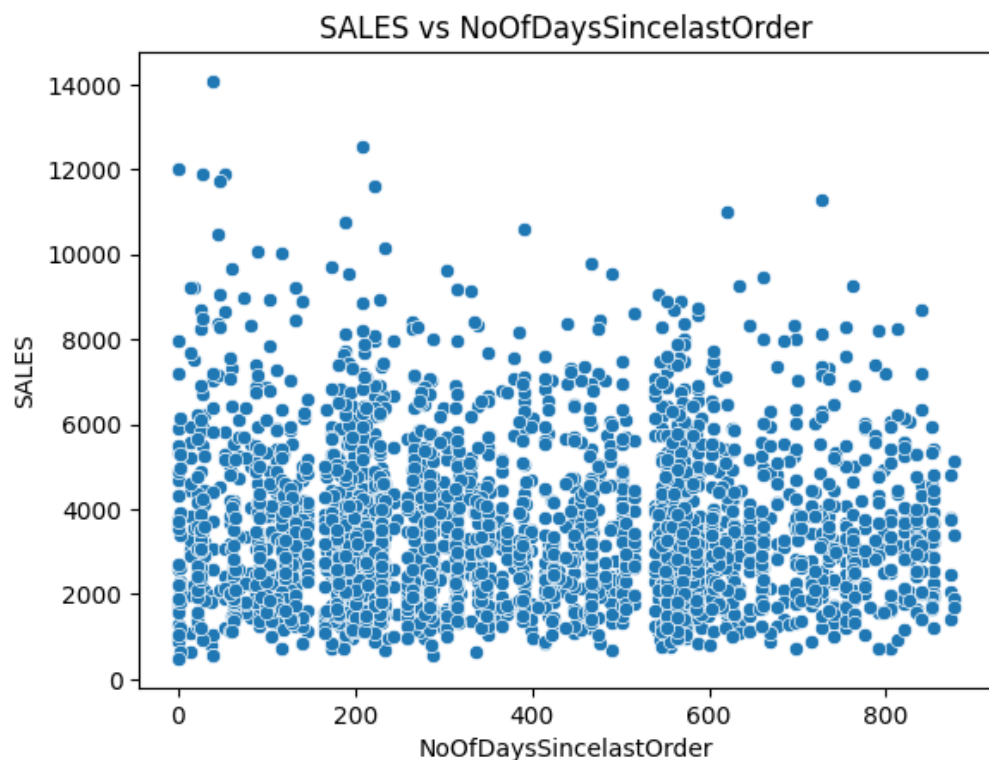
Sales vs Price Each



- Sales and Price Each exhibit a moderate positive correlation, with higher prices generally associated with higher sales.
- However, the relationship shows substantial scatter, indicating other factors beyond price also significantly influence sales.
- Some high-price products achieve low sales, while lower-priced products can generate high revenue.

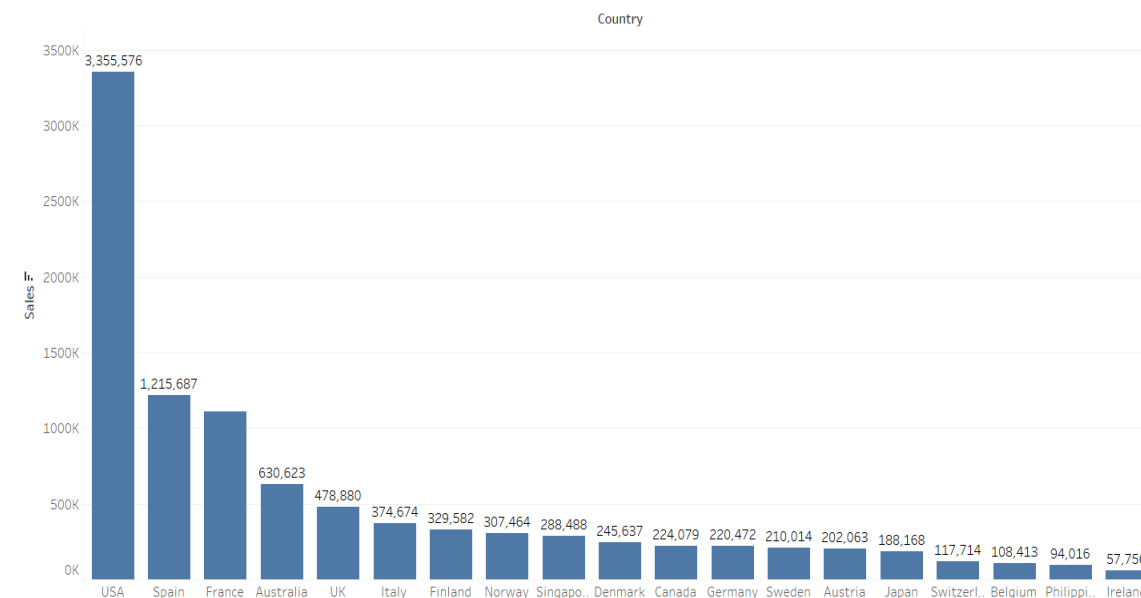
Sales vs Days Since Last Ordered

- Sales do not show a strong correlation with the number of days since the last order.
- Sales values are spread across all ranges of NoOfDaysSinceLastOrder, indicating no clear trend.
- A cluster of orders occurs within 200 days since the last order, which might signify a pattern of repeat purchases within this timeframe.



Sales across Country

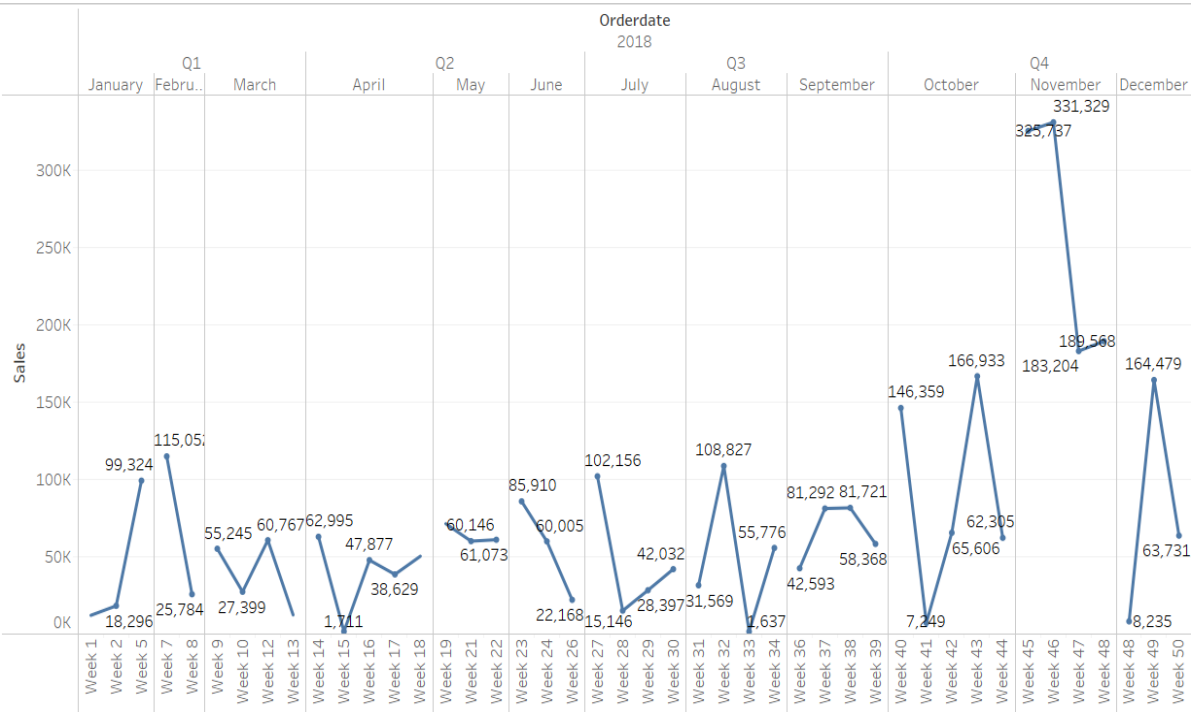
Sales Across Country



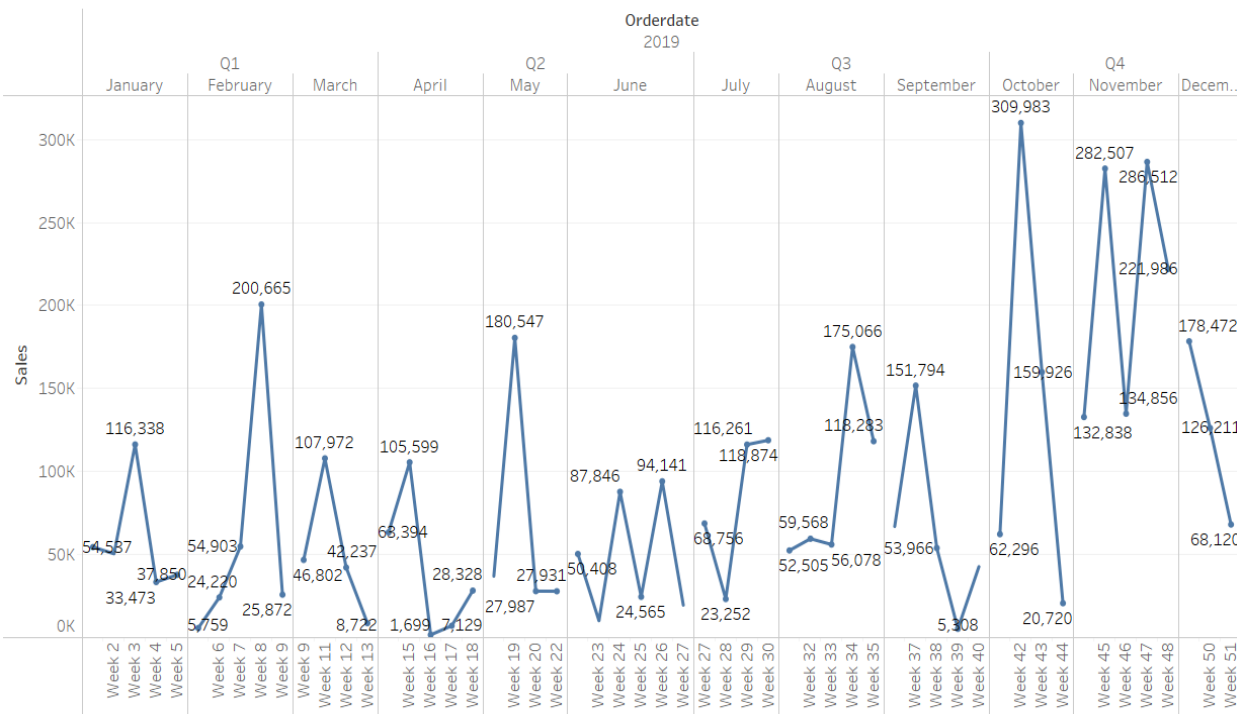
- USA dominates sales with \$3,355,576, accounting for a significant portion of the total revenue.
- Other top-performing countries include Spain (\$1,215,687) and France (\$630,623), contributing considerably but lagging behind the USA.
- The bottom-performing countries, such as Ireland (\$57,756) and Philippines (\$94,016), have much lower sales, highlighting potential growth opportunities.

Weekly, Monthly, Quarterly, and Yearly trends in Sales

Weekly, Monthly, Quarterly, and Yearly trends in Sales



Weekly, Monthly, Quarterly, and Yearly trends in Sales



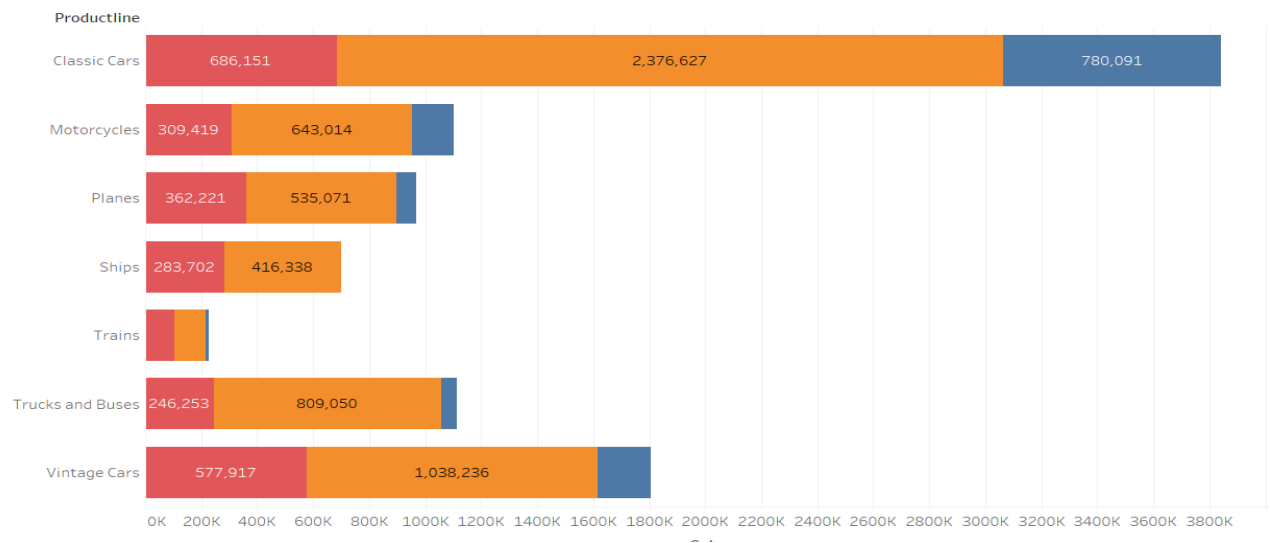
- High Volatility: Significant sales fluctuations indicate seasonal or irregular demand.
- Peaks and Droughts: Sudden demand surges are followed by inactivity.
- Quarterly Patterns: Peaks observed in Q3 and Q4, likely driven by seasonal factors.
- Yearly Growth: 2018 shows an upward sales trend, though not consistent.

- High Volatility: Sales fluctuate significantly across time frames, indicating seasonality or irregular demand.
- Sharp Peaks: Sudden surges in sales are followed by periods of low activity.
- Quarterly Patterns: Sales peak in Q3 and Q4, likely due to seasonal factors.
- Yearly Growth: Sales in 2019 show an upward trend despite inconsistencies.

Sales across Product line and Deal Size

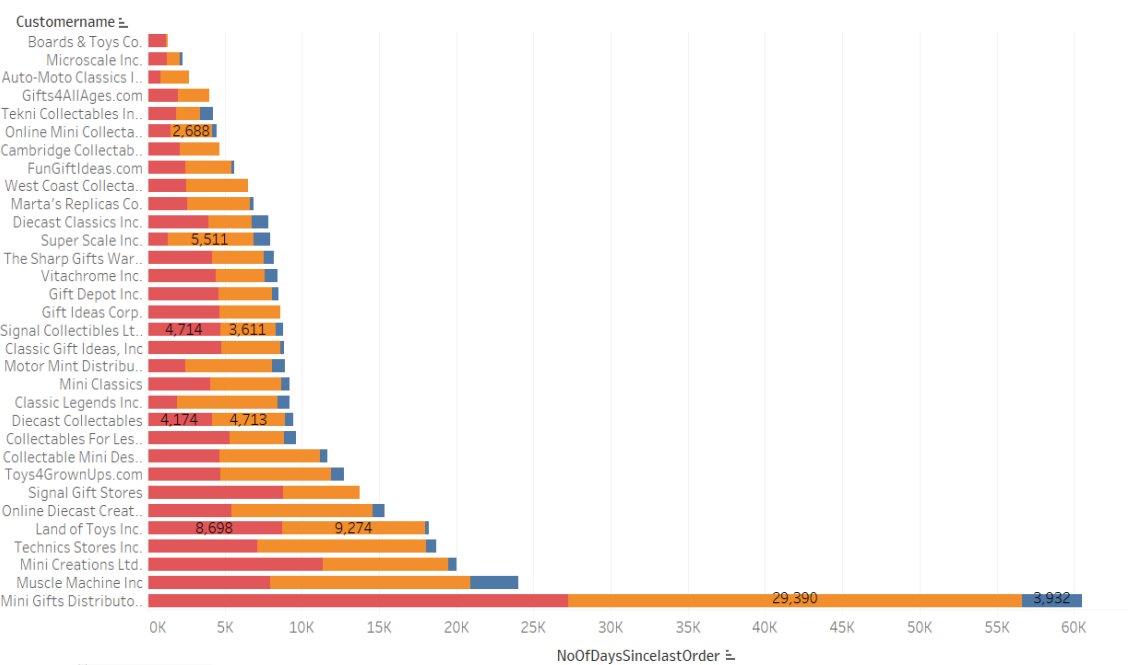
- Distribution Overview: Sales are distributed across small, medium, and large deal sizes for each product line.
- Product Line Comparison: Enables performance comparison based on deal size.
- Medium Deal Dominance: Medium-sized deals contribute the most sales, especially in Classic Cars, Motorcycles, and Vintage Cars.
- Product Line Variations: Classic Cars and Motorcycles show a higher share of large deals, unlike Planes and Ships.

Sales across Product line and Deal size



Customers vs Days since last order

Customers vs No of Days since last order



- Recency Distribution: Visualizes customer distribution based on days since their last order.
- Long-Tail Effect: Many customers haven't made recent purchases, indicating retention opportunities.
- Deal Size Insights: Color coding reveals the relationship between recency and deal size.
- Segmentation Potential: Identifies high-value customers with long inactivity for targeted retention strategies.

RFM Analysis: Customer Value Segmentation

What is RFM?

- **R**ecency: How recently a customer purchased
- **F**requency: How often they buy
- **M**onetary: Total customer spending

Key Benefits

- Identify high-value customers
- Personalize marketing strategies
- Optimize customer retention efforts

Segmentation Framework

- Score customers on RFM dimensions
- Create distinct customer segments
- Develop targeted engagement strategies

Customer Segmentation

Based on RFM (Recency, Frequency, Monetary) analysis, customers are segmented into three key segments:

- High Value (H): Customers with high recency, frequency, and monetary value.
- Medium Value (M): Customers with moderate values across RFM.
- Low Value (L): Customers with low engagement or contribution across RFM dimensions.

Assign Scores for Recency, Frequency, and Monetary

- H → 3 points
- M → 2 points
- L → 1 point

RFM Score: Recency Score + Frequency Score + Monetary Score

Criteria for segmentation

1. Gold Segment

- RFM Score: 7 or higher
- Customers who are highly recent, frequent, and have a high monetary value.

2. Silver Segment

- RFM Score: Between 4 and 6
- Customers with moderate recency, frequency, and spending.

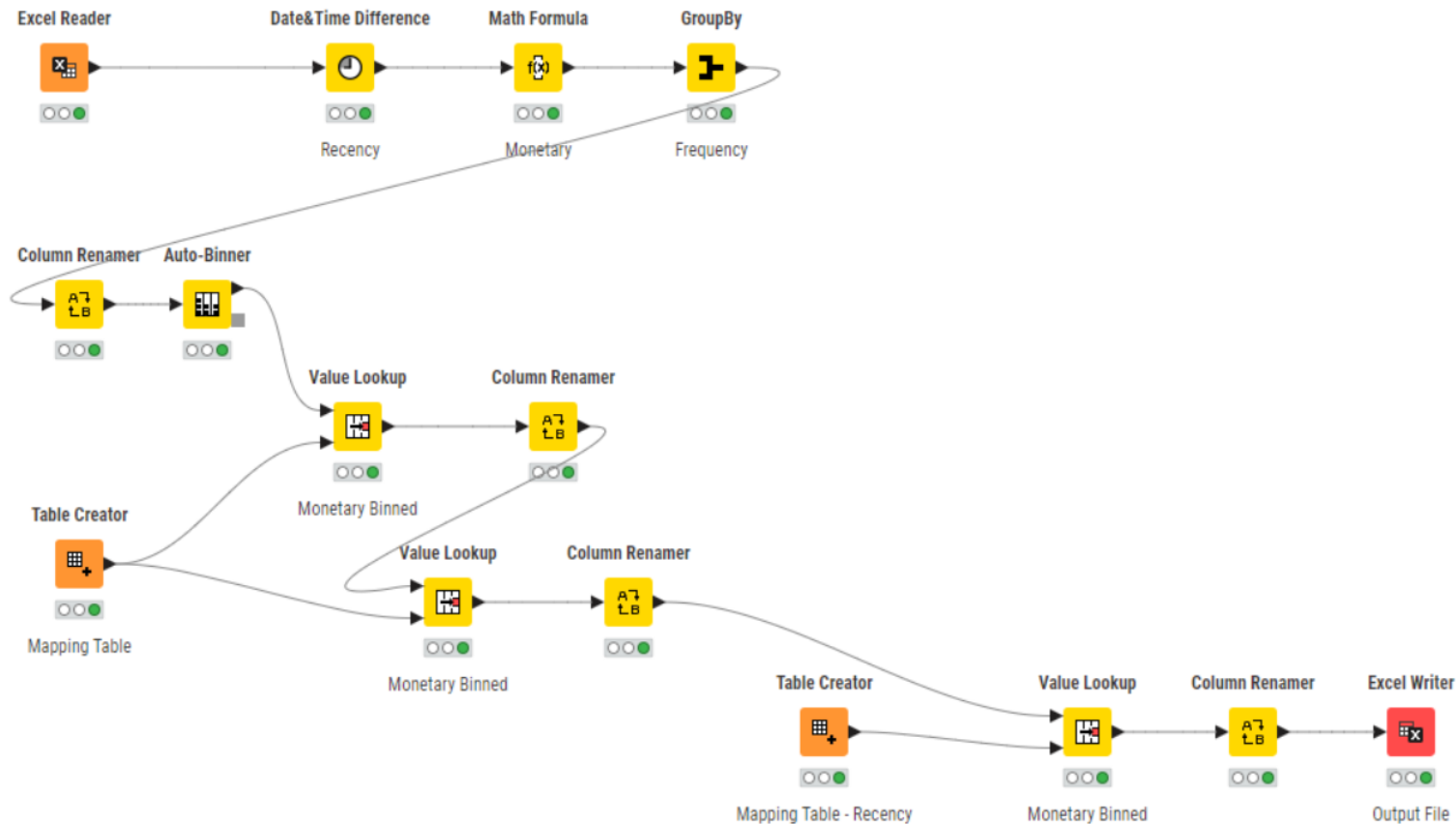
3. Bronze Segment

- RFM Score: Less than 4
- Customers with low recency, frequency, or spending.

Customer Segments and KNIME Workflow

		MONETARY		
RECENCY	FREQ	H	M	L
H	H	9	2	
	M	1	9	2
	L			
M	H	6	3	1
	M	5	16	5
	L		5	3
L	H			
	M	1	1	
	L		8	12

Segments	#Customers
Gold	35
Silver	42
Bronze	12



RFM Output

CUSTOMERNAME	Frequency	QUANTITY	PRICEEAC	ORDERLIN	SALES (M	MSRP (CITY (First	COUNTRY	Recency	Monetary	Frequency	Recency_I	Monetary	Monetary	Frequency	Recency_I	Monetary	Frequency	Recency_I	RFM_Scor	Segment
AV Stores, Co.	3	34.86275	91.08451	18	3094.271	92.84	Manchest	UK	195	157807.8	Bin 2	Bin 2	Bin 3	H	M	M	3	2	2	7	Gold
Alpha Cognac	3	34.35	101.16	12	3524.422	97.15	Toulouse	France	63	70488.44	Bin 2	Bin 1	Bin 1	L	M	H	1	2	3	6	Silver
Amica Models & Co.	2	32.42308	110.8527	17	3619.895	107.7	Torino	Italy	264	94117.26	Bin 1	Bin 3	Bin 2	M	L	L	2	1	1	4	Silver
Anna's Decorations, Ltd	4	31.93478	106.4241	14	3347.742	104.7	North Syd	Australia	82	153996.1	Bin 3	Bin 2	Bin 3	H	H	M	3	3	2	8	Gold
Atelier graphique	3	38.57143	92.23857	4	3454.28	95.57	Nantes	France	187	24179.96	Bin 2	Bin 2	Bin 1	L	M	M	1	2	2	5	Silver
Australian Collectables, Ltd	3	30.65217	90.04174	16	2808.324	88.13	Glen Wave	Australia	21	64591.46	Bin 2	Bin 1	Bin 1	L	M	H	1	2	3	6	Silver
Australian Collectors, Co.	5	35.01818	104.5902	15	3654.462	103.5	Melbourne	Australia	183	200995.4	Bin 3	Bin 2	Bin 3	H	H	M	3	3	2	8	Gold
Australian Gift Network, Co	3	36.33333	110.554	6	3964.608	111.5	South Bris	Australia	118	59469.12	Bin 2	Bin 2	Bin 1	L	M	M	1	2	2	5	Silver
Auto Assoc. & Cie.	2	35.38889	99.4878	17	3601.907	100.4	Versailles	France	232	64834.32	Bin 1	Bin 3	Bin 1	L	L	L	1	1	1	3	Bronze
Auto Canal Petit	3	37.07407	94.25519	15	3450.765	94.85	Paris	France	53	93170.66	Bin 2	Bin 1	Bin 2	M	M	H	2	2	3	7	Gold

Top 5 Customers

Best Customers: Customers with high Recency (R), Frequency (F), and Monetary (M) scores

Best Customers
Danish Wholesale Imports
Diecast Classics Inc.
Euro Shopping Channel
La Rochelle Gifts
Mini Gifts Distributors Ltd.

Customers on the Verge of Churning: Customers with Medium or Low Recency, and Low or Medium Frequency but moderate spending (Monetary).

Customers on the Verge of Churning
Classic Legends Inc.
Collectables For Less Inc.
Cruz & Sons Co.
Enaco Distributors
FunGiftIdeas.com

Top 5 Customers

Lost Customers: Customers with Low Recency, Low Frequency, and Low Monetary scores.

Lost Customers
Auto Assoc. & Cie.
Bavarian Collectables Imports, Co.
CAF Imports
Cambridge Collectables Co.
Clover Collections, Co.

Loyal Customers: Customers with High Frequency, regardless of Recency or Monetary scores.

Loyal Customers
Danish Wholesale Imports
Diecast Classics Inc.
Euro Shopping Channel
La Rochelle Gifts
Mini Gifts Distributors Ltd.