



? Sd V Y
S` VDV S[^3` S`ke de
Bda \Vf #

By: Sangram Keshari Patro



Contents:

- Agenda & Executive Summary of the data
 - Problem statement
 - Executive Summary
 - Data dictionary
 - Summary Stats
 - Assumptions about data
- Exploratory Analysis & Insights
 - Univariate, Bivariate, and multivariate analysis using data visualization
 - Weekly, Monthly, Quarterly, Yearly Trends in Sales
 - Sales Across different Categories of different features in the given data
 - Summarize the inferences from the above analysis
- Customer Segmentation using RFM analysis
 - What is RFM and which tool used
 - What all parameters used and assumptions made
 - Output table head
 - Workflow image to be put when KNIME used
- Inferences from RFM Analysis and identified segments
 - Who are your best customers?
 - Which customers are on the verge of churning?
 - Who are your lost customers?
 - Who are your loyal customers?



Agenda :

01

Agenda &
Executive Summary
of the data

02

Exploratory
analysis and
insights

03

RFM analysis for
customer
segmentation

04

Identification of
customers based
on different
parameters

05

Recommendations

01 Agenda & Executive Summary of the data

- Problem statement
- About Data (Info, Shape, Summary Stats, your assumptions about data)

Problem Statement:

The automobile parts manufacturing company wants to analyze three years of sales transactions to:

1. Discover patterns in customer purchase behavior.
2. Segment customers based on their behavior.
3. Improve targeted marketing strategies.
4. Drive personalized experiences to increase sales and customer retention.



Executive Summary:

- **Data:** past 3 years.
- **Objective:** identify the underlying buying patterns of the customers and recommend customized marketing strategies for different segments of customers.
- **Dataset:** 20 columns and 2747 rows,
- **Missing values and Duplicate values:** None
- **Outliers:** some columns has few outliers
- The exploratory analysis and insights provide a clear understanding of the data and highlight the key trends and patterns in sales.
- RFM analysis has been performed to segment the customers into four categories based on their buying behavior, and customized marketing strategies have been recommended for each segment.
- The presentation concludes with recommendations for the company to enhance its customer relationships and drive business growth.

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

Numeric columns: ORDERNUMBER, QUANTITYORDERED, PRICEEACH, ORDERLINENUMBER SALES, DAYS_SINCE_LASTORDER, MSRP.

Statistical Summary of Numerical Columns

	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

Inference :

- The average number of items ordered per sales order is 35, with a standard deviation of 9.76.
- The average price of each item is 101.09, with a standard deviation of 42.04.
- The average sales amount per order is 3553.05, with a standard deviation of 1838.95.
- The average time since the last order is 1757.09 days, with a standard deviation of 819.28.
- The summary statistics do not indicate any red flags or abnormalities that could potentially indicate issues with the data.

Statistical Summary of Categorical Columns

	count	unique	top	freq
STATUS	2747	6	Shipped	2541
PRODUCTLINE	2747	7	Classic Cars	949
PRODUCTCODE	2747	109	S18_3232	51
CUSTOMERNAME	2747	89	Euro Shopping Channel	259
PHONE	2747	88	(91) 555 94 44	259
ADDRESSLINE1	2747	89	C/ Moralarzal, 86	259
CITY	2747	71	Madrid	304
POSTALCODE	2747	73	28034	259
COUNTRY	2747	19	USA	928
CONTACTLASTNAME	2747	76	Freyre	259
CONTACTFIRSTNAME	2747	72	Diego	259
DEALSIZE	2747	3	Medium	1349

Inference :

- Most orders are shipped from the USA, dominated by medium-sized deals and purchases of Classic Cars, with frequent orders from Euro Shopping Channel.

02 Exploratory Analysis & Insights

Univariate, Bivariate, and multivariate analysis using data visualization

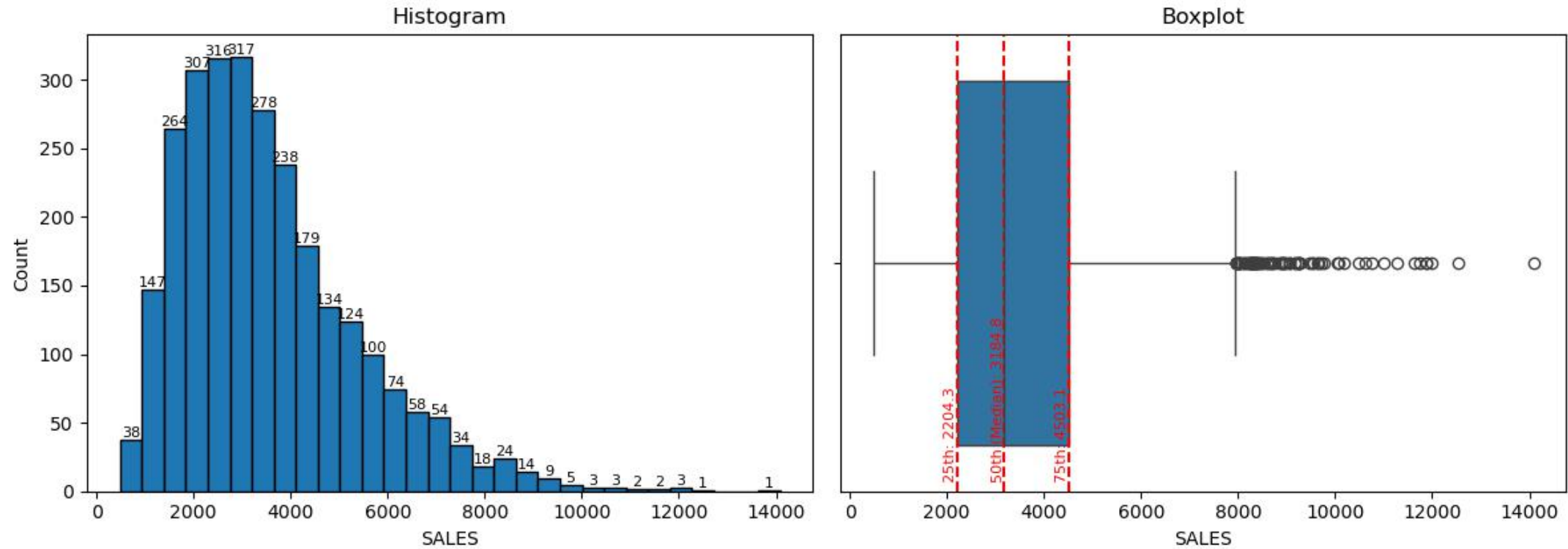
- Weekly, Monthly, Quarterly, Yearly Trends in Sales
- Sales Across different Categories of different features in the given data
- Summarize the inferences from the above analysis

Outliers has not been treated.



Univariate Analysis

Visualization for SALES



Sales

Skewed Right: Sales mostly range between 2,000–5,000 with a right tail.

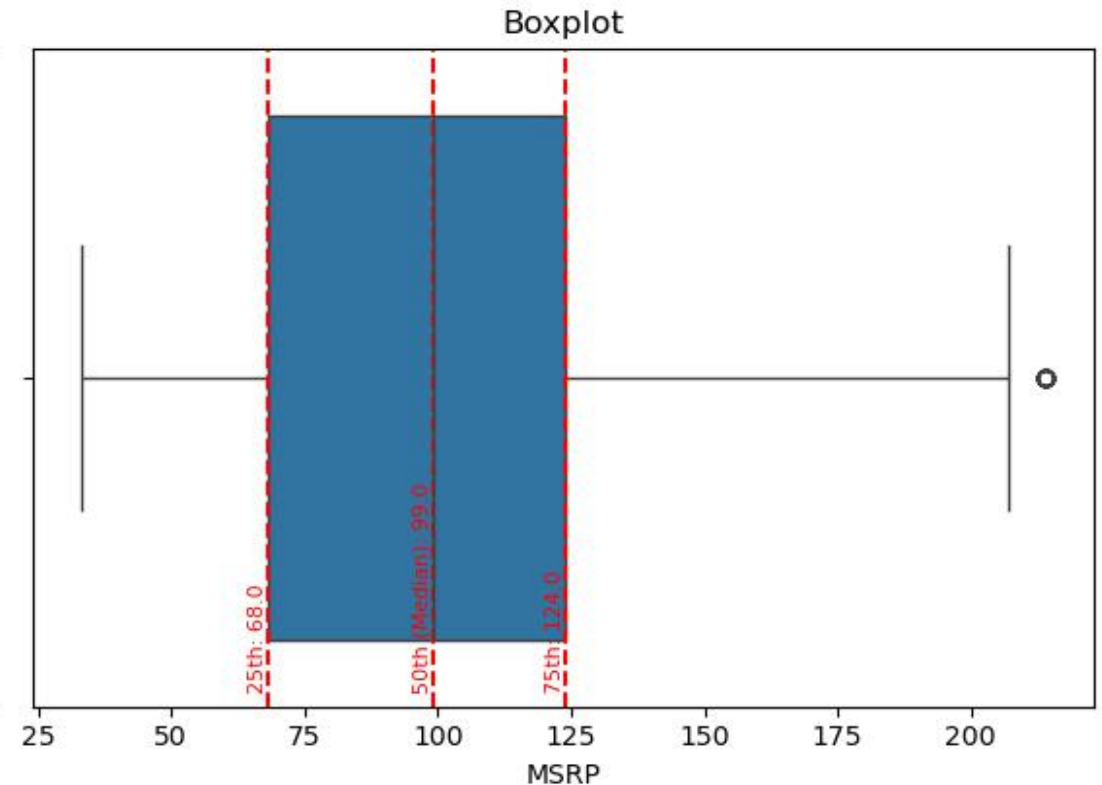
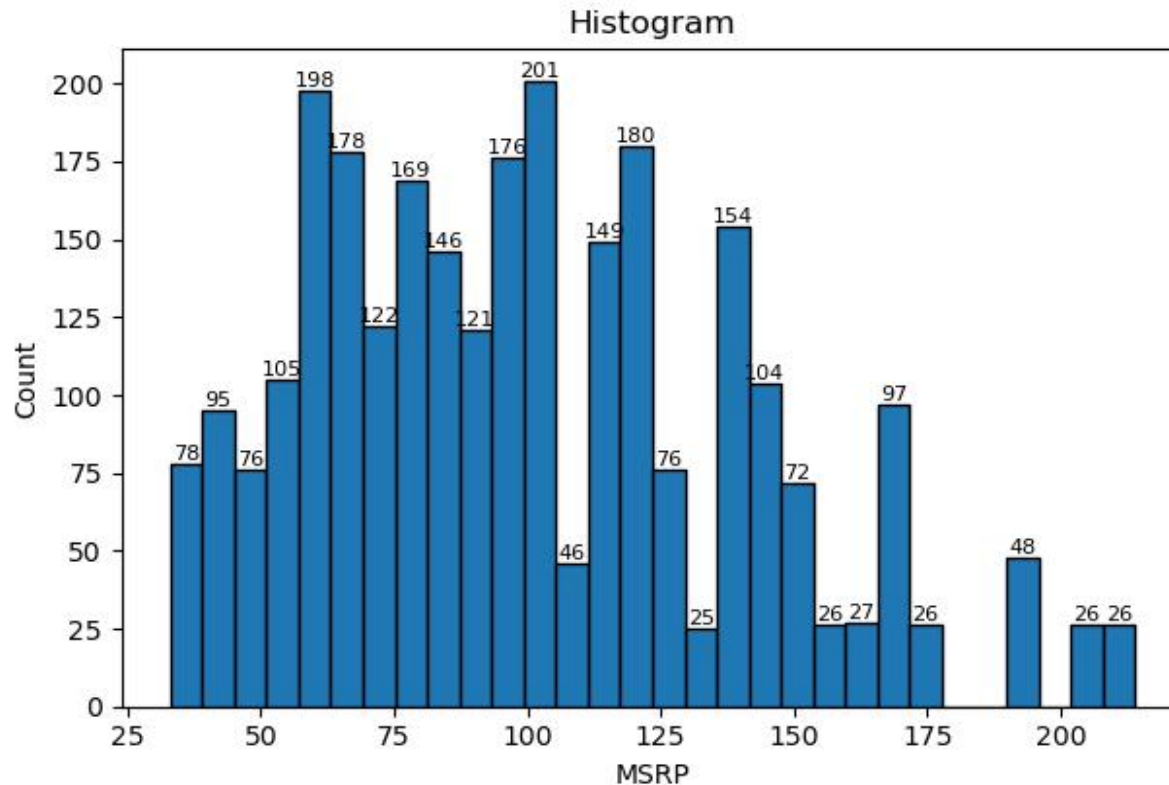
Outliers: Multiple high-value sales outliers are present.

Mid-Range Dominance: Most sales are under \$6,000.

MRSP

- Right Skew: MSRP values are mostly between 50–125, with fewer high-priced items.
- Outliers: A few high MSRP outliers are clearly visible beyond the upper whisker.

Visualization for MSRP

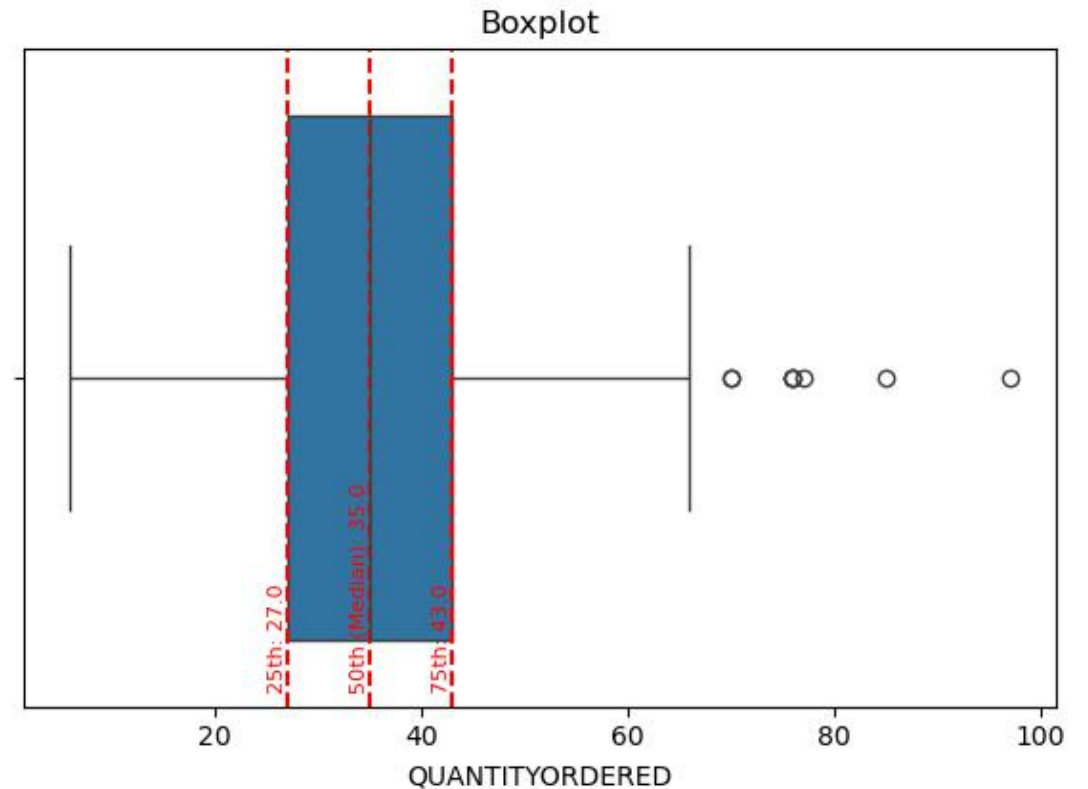
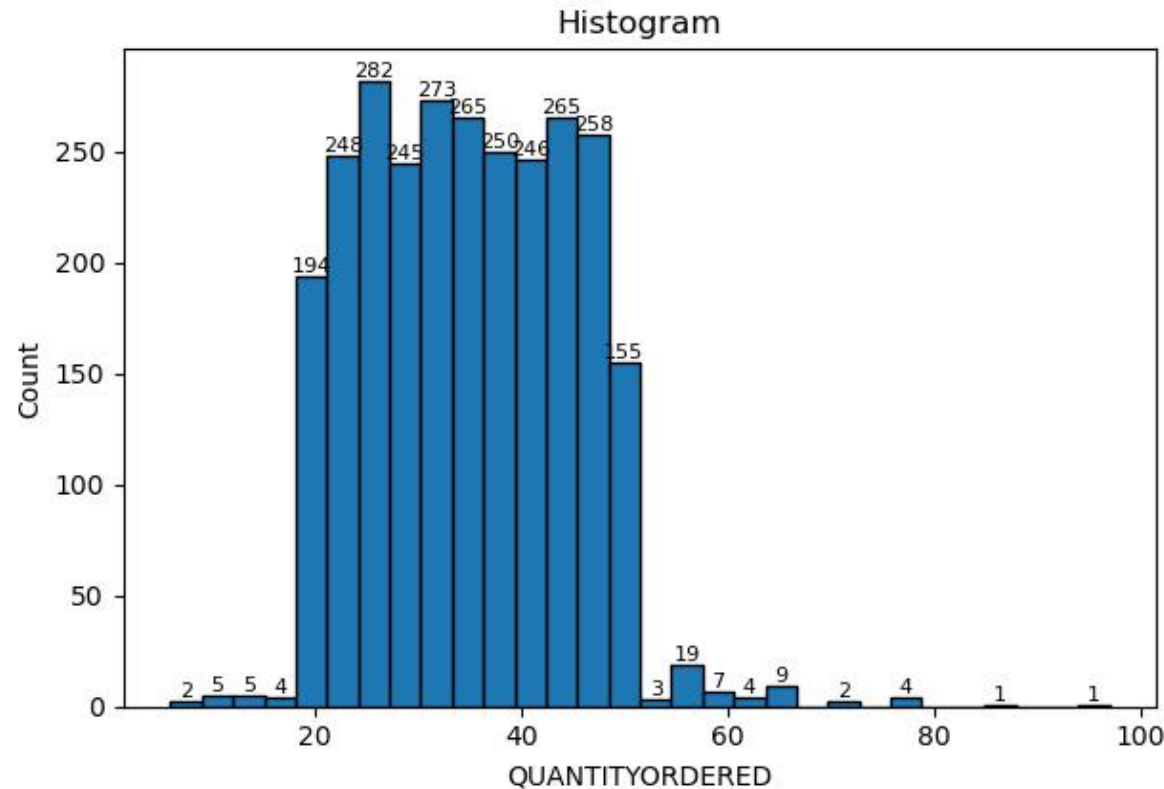


QUANTITY ORDERED

- Slight Right Skew: Most orders range between 20–45 units, with fewer larger orders.

Outliers: A few unusually high quantity orders (above 70) are visible.

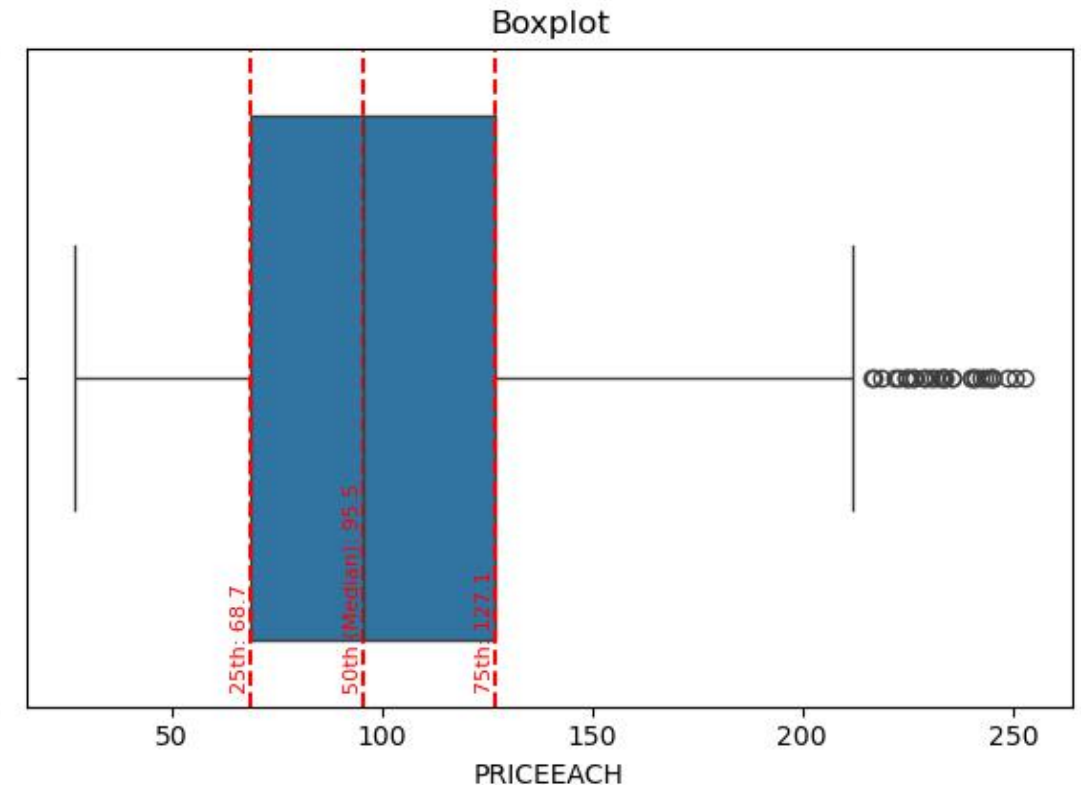
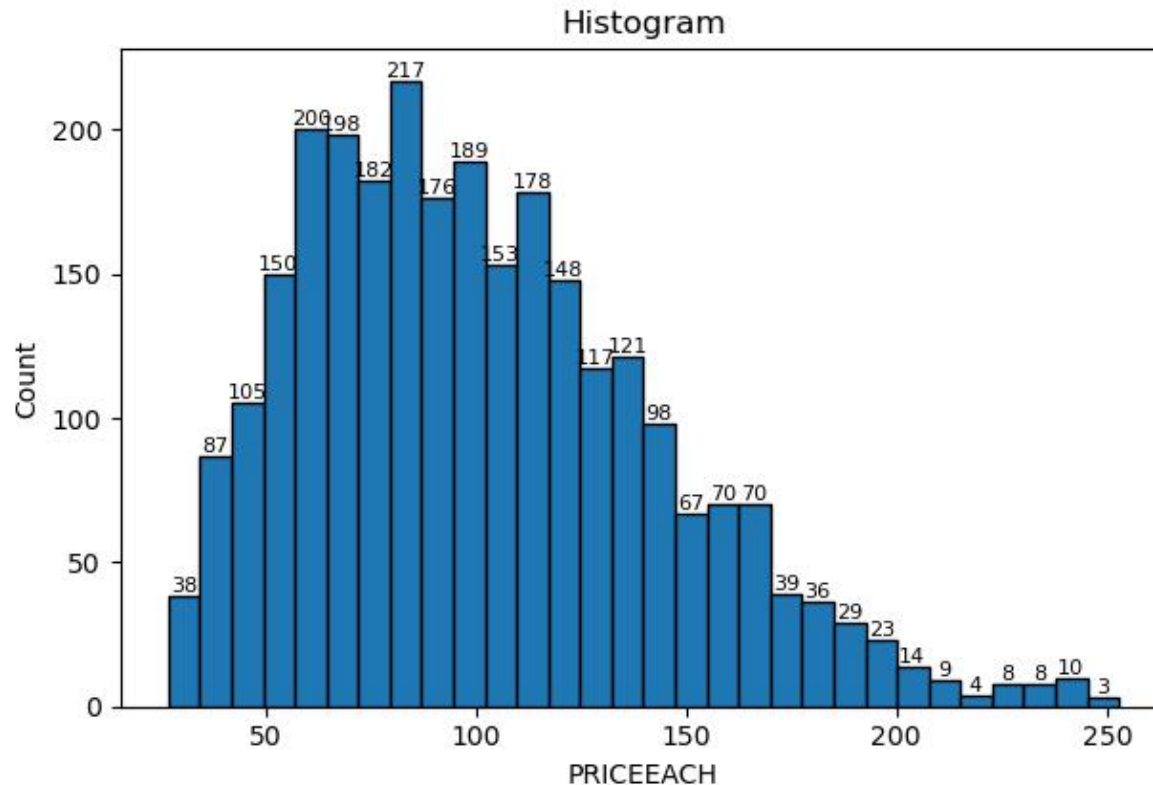
Visualization for QUANTITYORDERED



PRICE OF EACH

- Right Skewed: Most product prices fall between 40–120, tapering off at higher prices.
- Outliers: Prices above 200 appear as distinct outliers.
- Price Range Cluster: The central 50% of prices lie between approximately 68.7 and 127.1.

Visualization for PRICEEACH



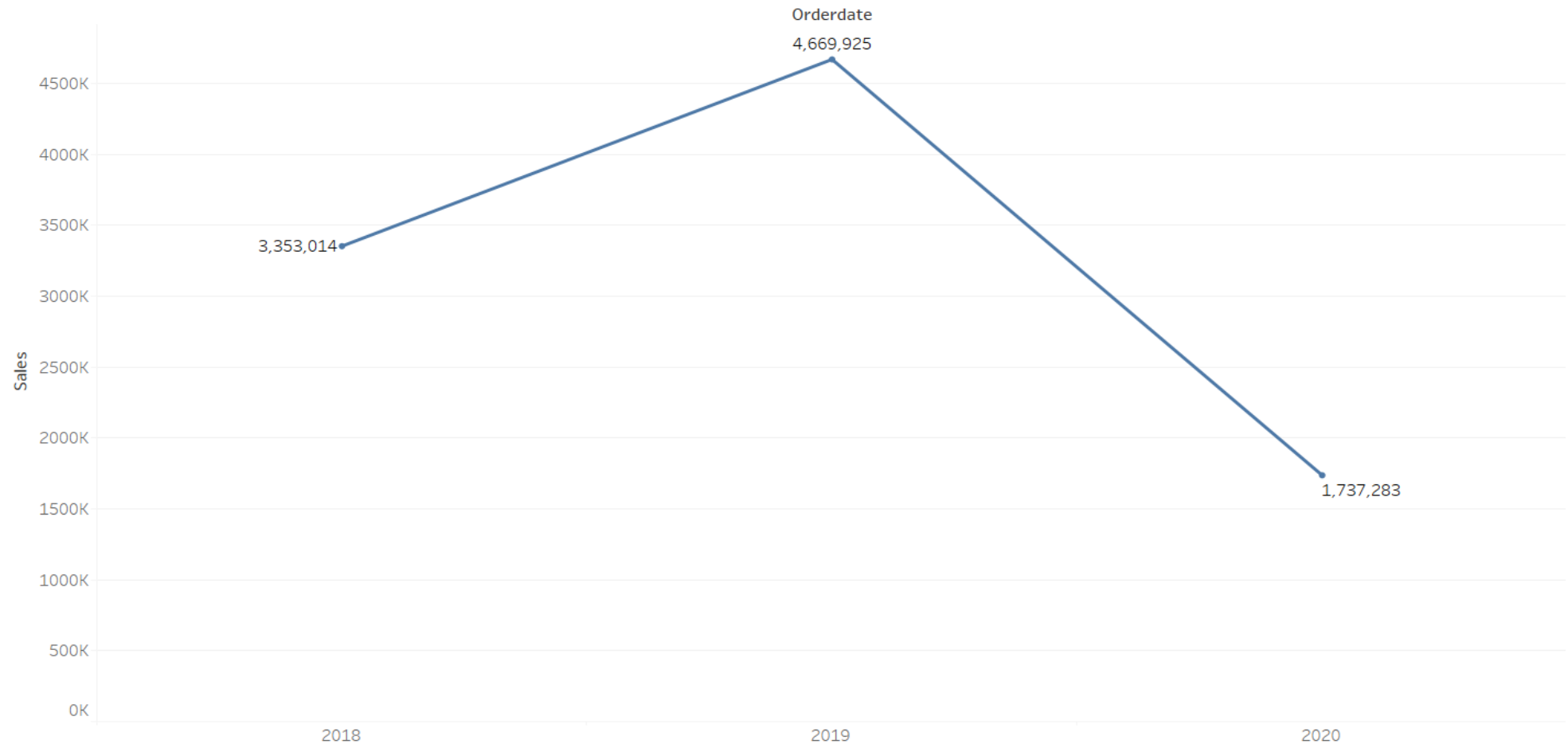


Bivariate Analysis



Yearly Sales

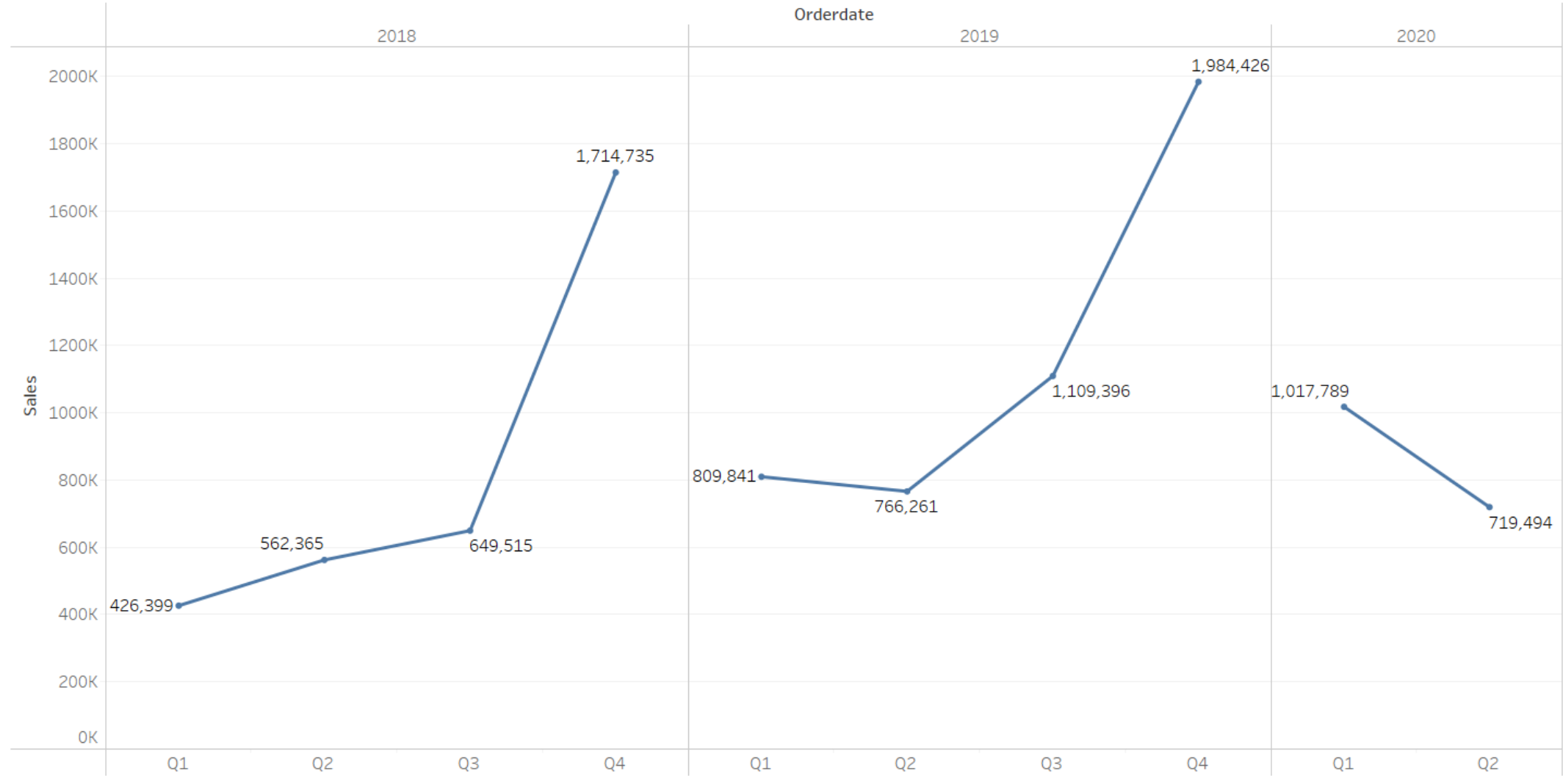
- Peak in 2019: Sales peaked at 4.67M in 2019, showing strong year-over-year growth from 2018 (3.35M).
- Sharp Decline: A significant drop occurred in 2020, with sales falling to 1.74M.



Quarterly Sales

We can see that in the quarter 4 there is high sales as compared to other.

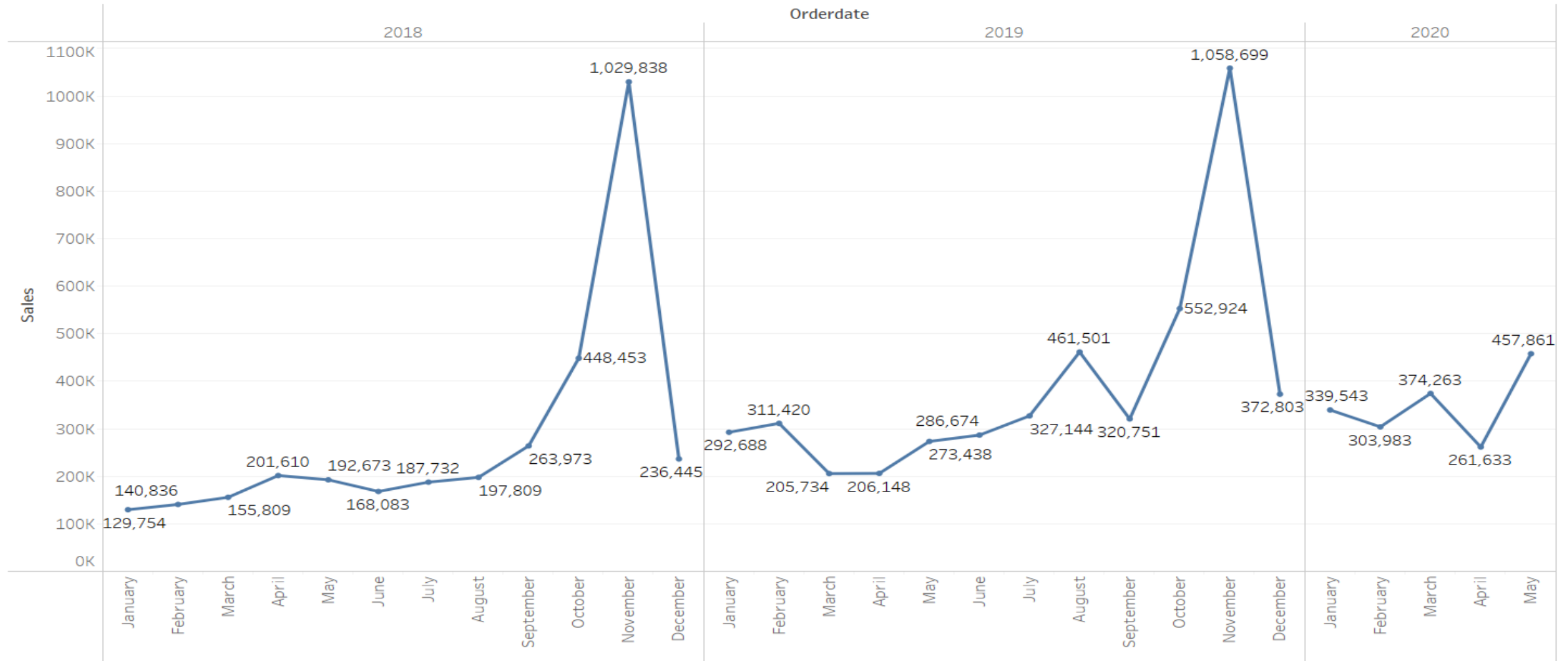
Yearly Sales



Monthly Sales

- Seasonal Growth: Sales steadily rise from mid-year towards November, suggesting seasonal trends.
- Recovery Signs in 2020: Despite lower overall sales, May 2020 shows a rebound with 458K,

Monthly Sales



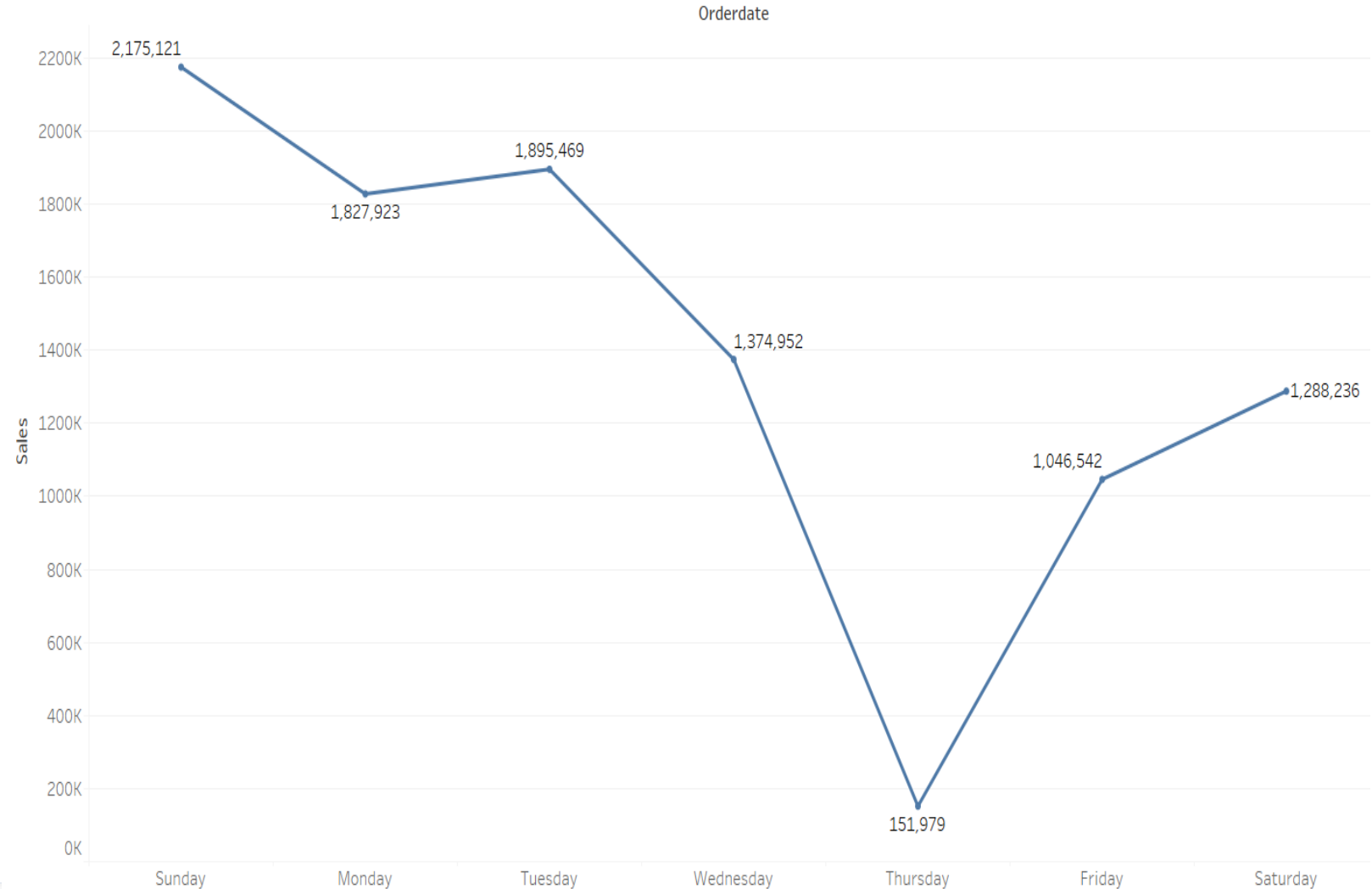
Weekday Sales

Peak on Sunday: Sales are highest on Sunday (2.18M).

Midweek Drop: Lowest sales on Thursday (152K).

Weekend Rise: Sales rebound Friday (1.05M) and Saturday (1.29M).

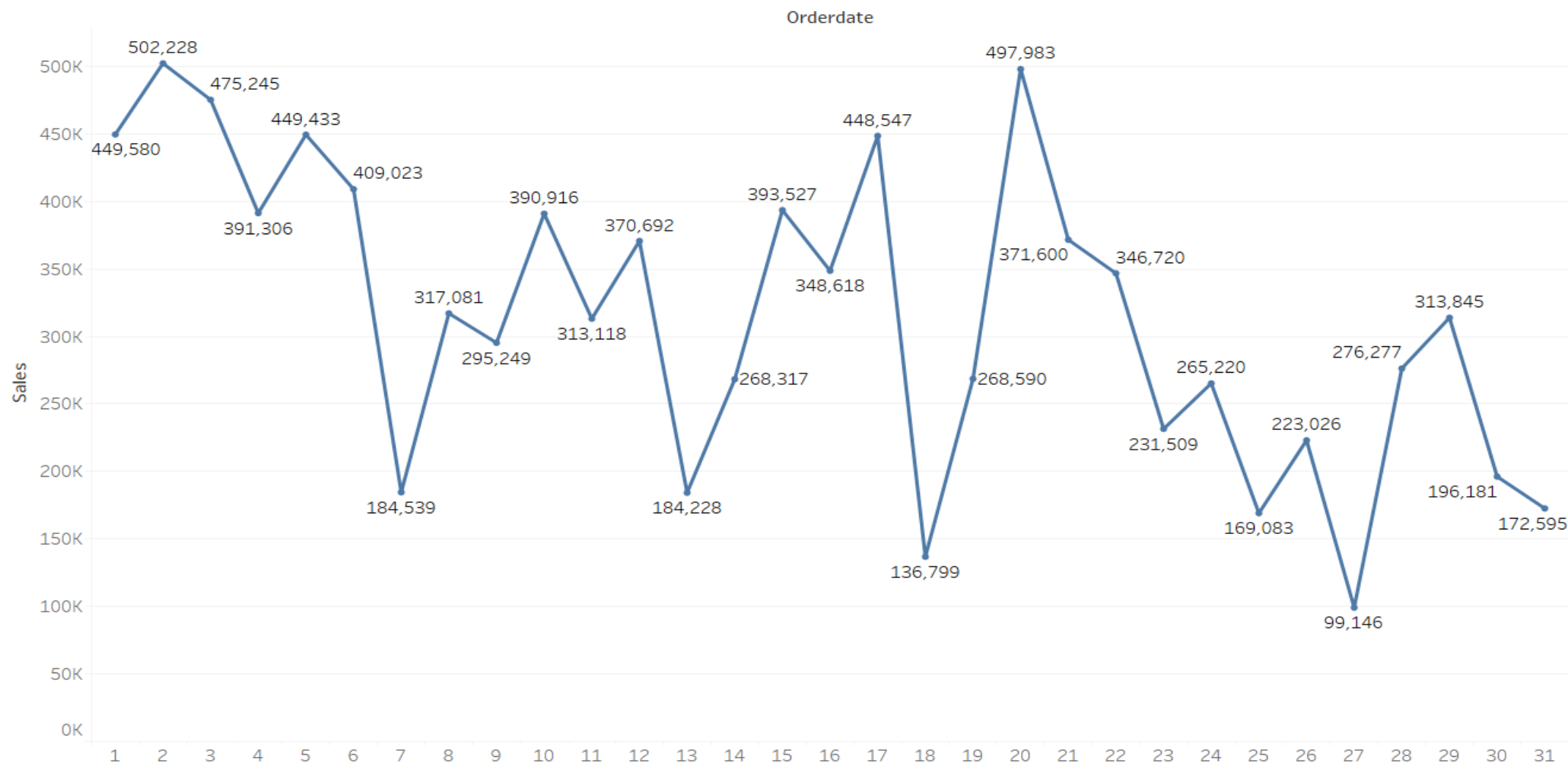
Weekday Sales



Day Sales

In the stating days of month there is more sales as compared to the end

Day Sales

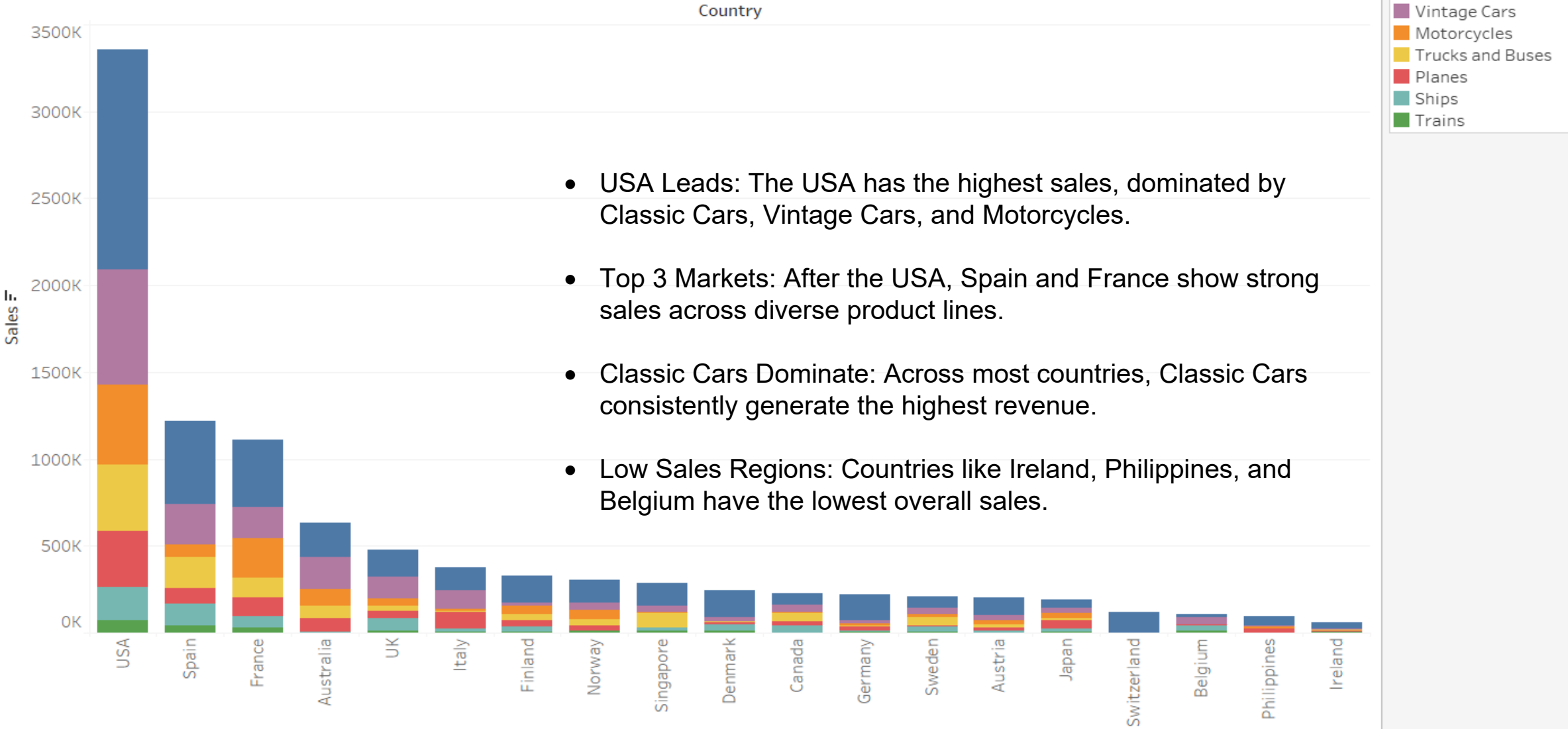




Multi-Variate Analysis

Productline, Country & Sales

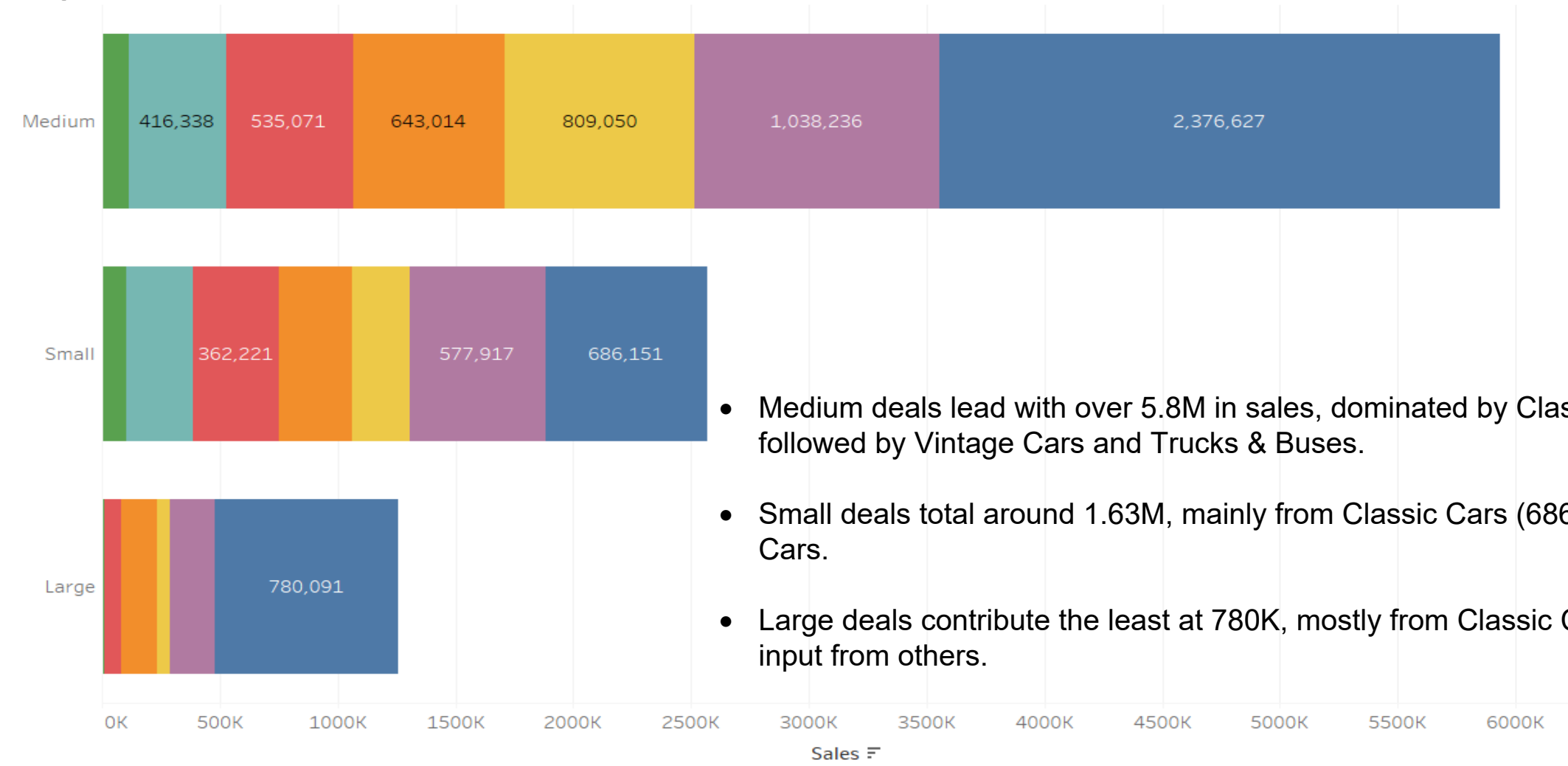
Productline, Country & Sales



DealSize, Country & Sales

DealSize, Country & Sales

Deal..



- Productline
- Classic Cars
 - Vintage Cars
 - Trucks and Buses
 - Motorcycles
 - Planes
 - Ships
 - Trains

- Medium deals lead with over 5.8M in sales, dominated by Classic Cars (2.38M), followed by Vintage Cars and Trucks & Buses.
- Small deals total around 1.63M, mainly from Classic Cars (686K) and Vintage Cars.
- Large deals contribute the least at 780K, mostly from Classic Cars, with minimal input from others.

Productline vs status

Classic Cars dominate Shipped and In Process orders. Motorcycles have high dispute rate (44%).

Vintage Cars show steady sales across all statuses. Ships and Planes are more involved in Resolved and Cancelled.

Trains have very low sales share overall. Most On Hold orders belong to Classic Cars.

Running Sum of SUM(S...

25,799

9,019,094

iii Columns

Productline

Rows

Status

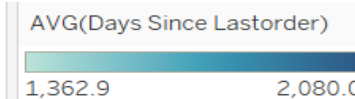
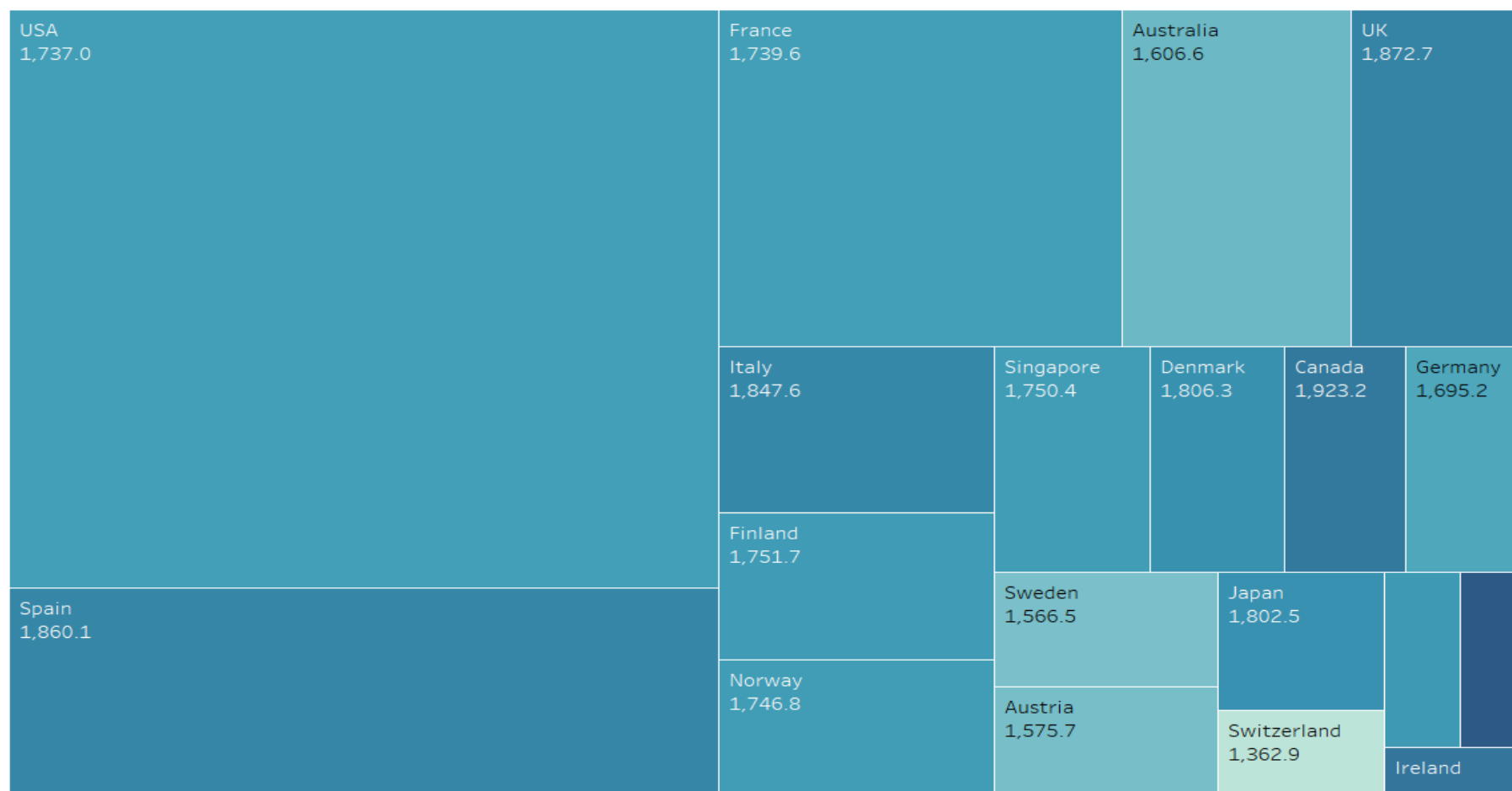
Productline vs status

Productline							
Status	Classic Ca..	Motorcyc..	Planes	Ships	Trains	Trucks an..	Vintage C..
Cancelled	30.46%		18.22%	29.14%	2.61%		19.57%
Disputed	36.02%	44.07%	5.32%	4.25%			10.34%
In Process	39.91%					29.73%	30.36%
On Hold	27.40%	2.79%	19.40%	13.22%	3.25%	11.28%	22.66%
Resolved	17.12%		22.91%	26.45%		13.58%	19.94%
Shipped	40.19%	11.83%	9.54%	6.40%	2.39%	11.40%	18.26%

Avg. days since last order, Sales vs Country

- Philippines has the longest gap (2,080 days) but low sales.
- USA records the highest sales with a lower gap (1,737 days).
- Switzerland shows the shortest gap (1,362.9 days), indicating frequent purchases. Countries like Ireland and Canada have high gaps, suggesting declining engagement

Avg. days since last order, Sales vs Country



Status, Sales and Deal Size

Status, Sales and Deal Size

Dealsize	Status						Shipped
	Cancelled	Disputed	In Process	On Hold	Resolved		
Large		42,747	27,152	42,941	8,885		1,137,231
Medium	137,575	22,165	83,787	106,342	103,967		5,477,396
Small	56,912	7,301	33,791	29,697	37,867		2,404,467

SUM(Sales)

7,301

5,477,396

- Large deals have no cancelled orders — a positive sign.
- Most Disputed and On Hold deals are Large and Medium.
- Medium deals dominate Shipped sales (5.48M).
- Small deals also contribute heavily to Shipped (2.40M).

INFERENCES

Peak Performance in 2019: Sales reached their highest at 4.67M, growing significantly from 2018 (3.35M).

Seasonality Observed: Sales consistently peak from mid-year to November.

USA Market Dominance: The USA leads in sales, largely driven by Classic Cars, Vintage Cars, and Motorcycles.

Thursday has the lowest sales, and **Sunday** has the highest sales.

Classic Cars Lead Globally: This product line consistently generates the highest revenue.

Underperforming Regions: Countries like Ireland, Philippines, and Belgium show minimal engagement.

Medium deals are the most profitable
Small deals contribute significantly.

Motorcycles have the highest dispute rate
Ships & Planes are mostly in Resolved and Cancelled categories.

USA has the shortest time gaps (~1,737 days).
Switzerland shows the most frequent purchases.

Classic Cars dominate both Shipped and In Process statuses.

Recommendation

Boost Seasonal Promotions: Launch targeted marketing campaigns from June to November to leverage peak demand.

Investigate 2020 Decline: Analyze operational and market changes in 2020 to mitigate future downturns.

Target Midweek Sales Dip: Run discounts or flash sales on Thursdays to balance weekly revenue.

Enhance Sunday Inventory: Ensure high inventory on Sundays to capitalize on maximum sales activity.

Expand Classic Cars Line: Given its consistent performance, invest further in product variety and inventory.

Revitalize Low-Sales Regions: Run awareness and promotional campaigns in Ireland, Belgium, and the Philippines.

Monitor On Hold Orders: Investigate and resolve delays in Classic Cars to improve customer satisfaction.

Increase Engagement in High-Gap Countries: Use email marketing and loyalty programs to engage customers in Canada and Ireland.

Upsell from Small to Medium Deals: Design incentive programs that encourage small-deal customers to scale up.

Capitalize on Large Deal Reliability: Promote Large deals to clients who value consistency and reliability (no cancellations).

03 Customer Segmentation using RFM analysis

- What is RFM and which tool used
- What all parameters used and assumptions made
- Output table head
- Workflow image to be put when KNIME used

What all parameters used and assumptions made

What is RFM ?

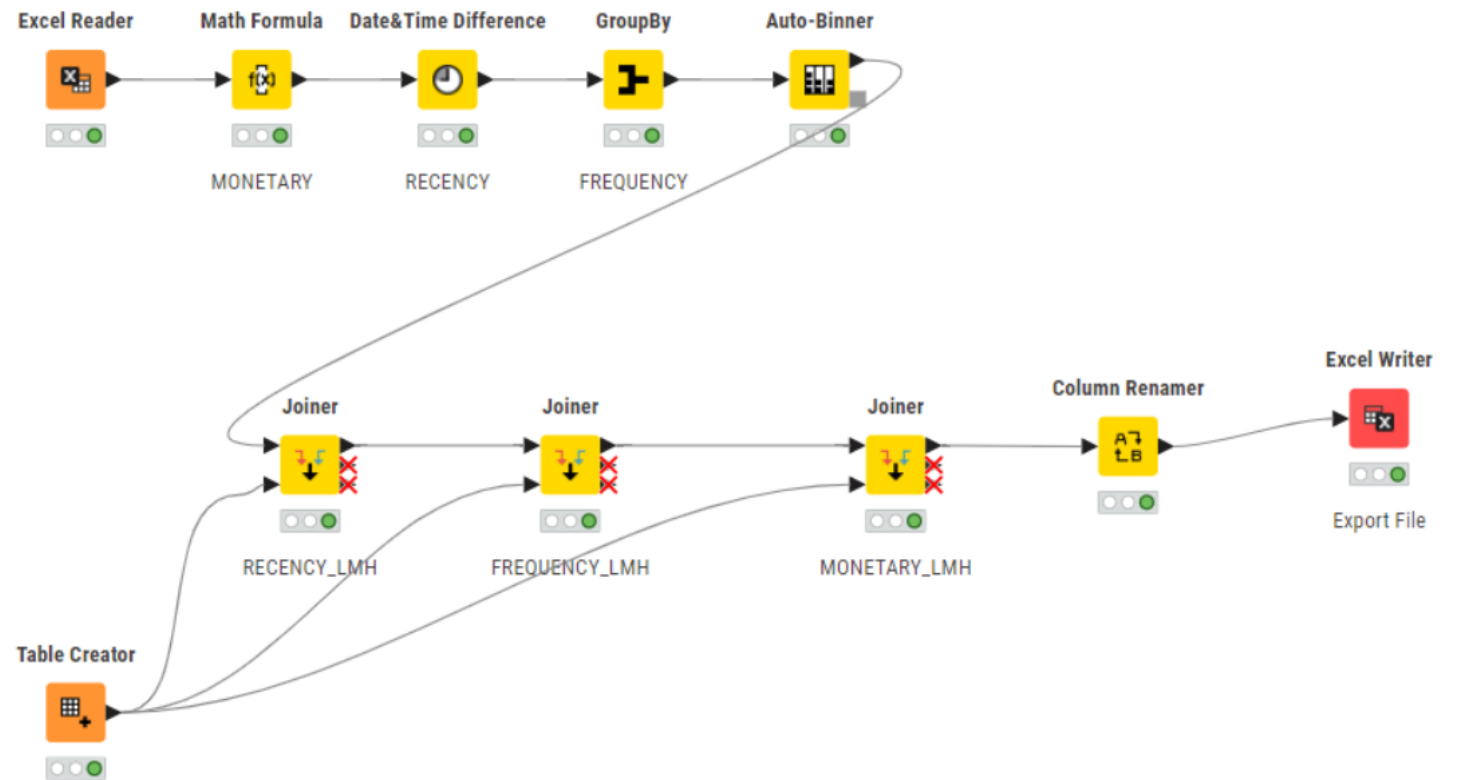
- Recency, frequency, monetary value (RFM) is a marketing analysis tool used to identify a firm's best clients based on the nature of their spending habits.
- An RFM analysis evaluates clients and customers by scoring them in three categories: how recently they've made a purchase, how often they buy, and the size of their purchases.

- As per instructions the column 'Days since last order' is ignored and new column **Recency** as '[Max(order date)- reference order date customerwise]'
- We have assumed '01-06-2020' as a reference date (as Max(order date)+ 1 day) and created recency column.
- The calculated formula for:
 - Recency :- [min(Recency) customer wise].
 - Frequency:- [count(customer name) customer wise]. We can also take order quantity
 - Monetary:- [sum(unit price * qty ordered) customer wise]. We can also take sales
 - Based on above we have made 3 bins : high , medium , low

Tool Used



KNIME WORKFLOW



Few rows of output

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC
1	CUSTOMER	ORDERID	QUANTITY	PRICE	UNITPRICE	ORDERDATE	SALSALES	ORDERDATE	DAYS_S	STATUS	PRODUCTID	MSRP	PRODUCTNAME	PHONE	ADDRESS	CITY	POSTAL	COUNTY	CONTACTNAME	CONTACTPHONE	DEALSIZE	MONETARY	RECENT	ORDERID	MONETARY	RECENT	Recency	FREQUENCY	MONETARY_LMH
2	AV Store	51	34.8627	91.0845	9.01961	3094.27	#####	1803.8	Shipped	Classic C	92.8431	S12_1108	(171) 555- Fauntleroy Manches	EC2 5NT	51	Ashworth	Victoria	Medium	157808	197	Bin 3	Bin 3	Bin 2	M	H	H			
3	Alpha Co	20	34.35	101.16	4.95	3524.42	#####	2236.2	Shipped	Classic C	97.15	S10_4757	61.77.655 1 rue Alsé Toulouse	31000	20	Roulet	Annette	Medium	70488.4	65	Bin 1	Bin 1	Bin 1	H	L	L			
4	Amica M	26	32.4231	110.853	7.61538	3619.89	#####	1318.62	Shipped	Classic C	107.654	S10_1949	011-49885 Via Mont Torino	10100	26	Accorti	Paolo	Large	94117.3	266	Bin 2	Bin 2	Bin 3	L	M	M			
5	Anna's C	46	31.9348	106.424	6.43478	3347.74	#####	1463.59	Shipped	Classic C	104.717	S10_1949	02 9936 E 201 Miller North Sy	2060	46	O'Hara	Anna	Small	153996	84	Bin 3	Bin 3	Bin 2	M	H	H			
6	Atelier gr	7	38.5714	92.2386	2	3454.28	#####	1424.43	Shipped	Motorcyc	95.5714	S10_2016	40.32.255 54, rue R Nantes	44000	7	Schmitt	Carine	Medium	24180	189	Bin 1	Bin 1	Bin 2	M	L	L			
7	Australia	23	30.6522	90.0417	6.69565	2808.32	#####	1689.7	Shipped	Vintage C	88.1304	S18_1342	61-9-3844 7 Allen S Glen Wa	3150	23	Connery	Sean	Medium	64591.5	23	Bin 2	Bin 1	Bin 1	H	M	L			
8	Australia	55	35.0182	104.59	7.03636	3654.46	#####	1698.78	Shipped	Motorcyc	103.527	S10_1678	03 9520 4 636 St Ki Melbourn	3004	55	Fergusson	Peter	Medium	200995	185	Bin 3	Bin 3	Bin 2	M	H	H			
9	Australia	15	36.3333	110.554	3.06667	3964.61	#####	1349.13	Shipped	Classic C	111.533	S10_1949	61-7-3844 31 Dunca South Br	4101	15	Calagher	Tony	Large	59469.1	120	Bin 1	Bin 1	Bin 2	M	L	L			
10	Auto Ass	18	35.3889	99.4878	8.55556	3601.91	#####	1266.28	Shipped	Classic C	100.389	S10_1949	30.59.855 67, avenue Versailles	78000	18	Tonini	Daniel	Large	64834.3	234	Bin 1	Bin 1	Bin 3	L	L	L			
11	Auto Car	27	37.0741	94.2552	6.33333	3450.77	#####	1763.44	Shipped	Motorcyc	94.8519	S10_1678	(1) 47.55.125, rue L Paris	75016	27	Perrier	Dominique	Medium	93170.7	55	Bin 2	Bin 2	Bin 1	H	M	M			
12	Auto-Mot	8	35.875	92.8	2	3309.91	#####	2281.75	Shipped	Ships	87.375	S18_3025	61755584 16780 Poi Brickhav	58339	8	Taylor	Leslie	Medium	26479.3	181	Bin 1	Bin 1	Bin 2	M	L	L			
13	Baane M	32	33.8125	108.574	6.34375	3643.72	#####	1805.75	Shipped	Motorcyc	107.469	S10_1678	07-98 95 Erling Sk Stavem	4110	32	Bergulfs	Jonas	Medium	116599	209	Bin 2	Bin 2	Bin 2	M	M	M			
14	Bavarian	14	28.6429	84.2893	7.5	2499.57	#####	2239.36	Shipped	Planes	82.7143	S18_1662	+49 89 61 Hansastr Munich	80686	14	Donnerr	Michael	Medium	34993.9	260	Bin 1	Bin 1	Bin 3	L	L	L			
15	Blauer Si	22	36.8636	108.031	3.86364	3871.44	#####	1697.32	Shipped	Classic C	105.818	S12_1099	+49 69 66 Lyonerst Frankfurt	60528	22	Keitel	Roland	Medium	85171.6	209	Bin 2	Bin 2	Bin 2	M	M	M			
16	Boards &	3	34	89.8067	1.33333	3043.12	#####	1968.67	Shipped	Classic C	92.3333	S12_3380	31055523 4097 Dou Glendale	92561	3	Young	Leslie	Medium	9129.35	114	Bin 1	Bin 1	Bin 2	M	L	L			

Inferences from RFM Analysis and identified segments

Who are your best customers?

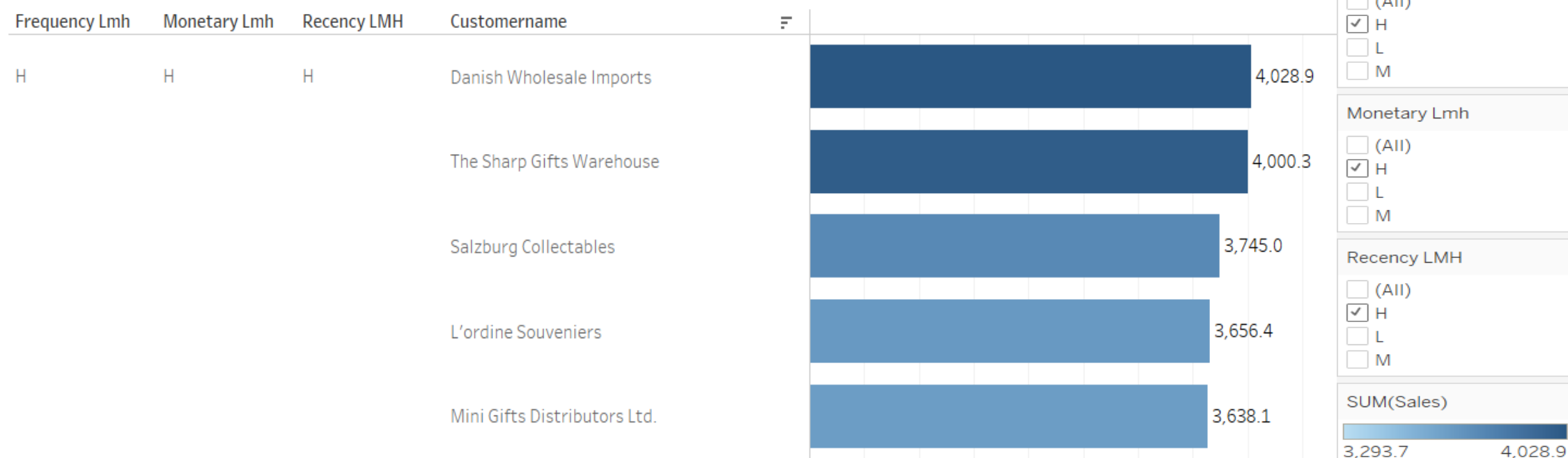
Which customers are on the verge of churning?

Who are your lost customers?

Who are your loyal customers?

Top 5 best customers

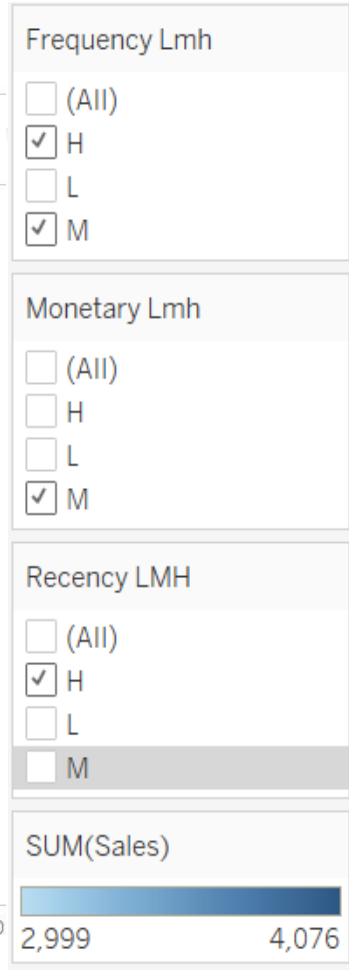
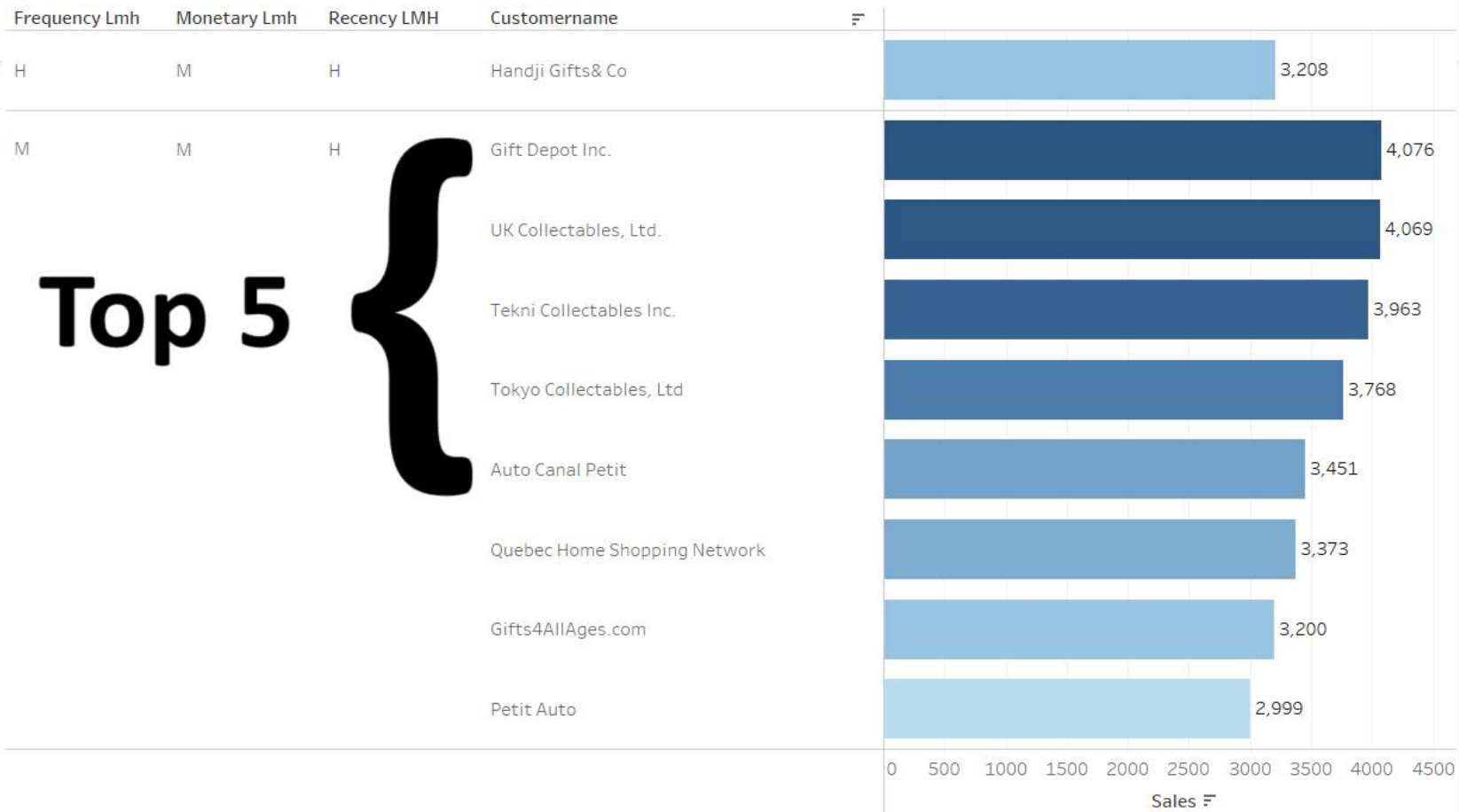
Best Customers



- According to RFM score we have grouped the top customers and arranged them by descending order of sales.
- We have given importance to recency more and ordered the customers accordingly.
- These customers are very active and profitable. We should focus on retention and exclusive offers.

Top loyal customers

Top Loyal Customers



- These customers buy often and regularly, but may not spend as much as top customers i.e HHM or HHL. But there was only 1 customer in these two categories so, I have included HMM.
- These are habitual customers who might not spend a lot per transaction, but they come back.
- Focus on loyalty programs and upselling.

Customers on the Verge of Churning

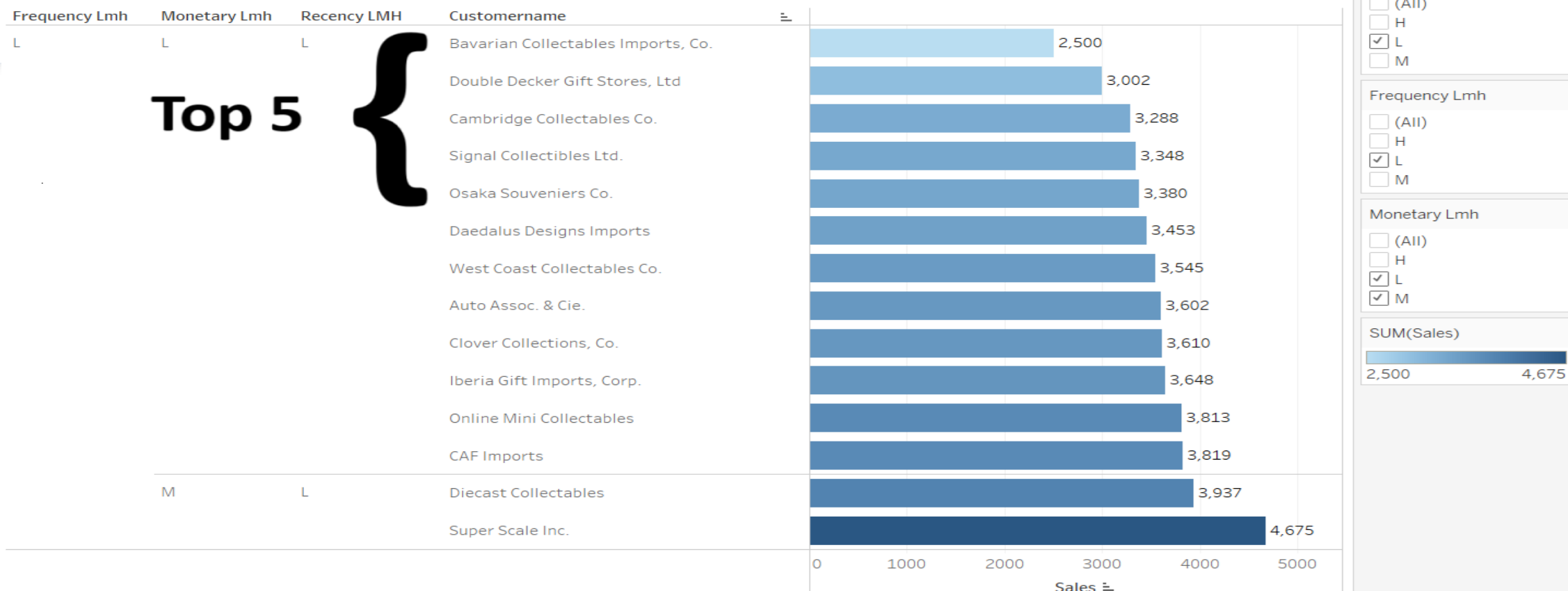
Verge of Churning Customers



- These are customers who used to be active but haven't purchased in a while (LHH,LHM,LMH,LMM).
- These customers were once valuable but are now disengaged. Focus on re-engagement campaigns like win-back emails or special incentives.

Top Lost Customers

Lost Customers



1. Very low recency: No purchases in a long time.
2. Rare purchase frequency: Bought only once or twice.
3. Low monetary value: Spent little overall.
4. Low engagement: Unresponsive to past campaigns.
5. High churn risk: Unlikely to return without a strong incentive or reason. These customers are ideal for low-cost reactivation efforts or might be removed from premium targeting lists.

05 Recommendations



Recommendations

A targeted strategy for each group is crucial to maximize retention and improve overall customer experience.

Best Customers

- Provide personalized recognition to reinforce their value to the company.
- Offer exclusive deals, VIP perks, and early access to new products or services.
- Maintain consistent engagement to strengthen long-term loyalty.

Loyal Customers

- Keep them engaged through regular promotions, loyalty points, and occasional surprises.
- Encourage referrals and reviews to turn them into brand advocates.
- Aim to elevate them to the 'Best Customer' category through personalized incentives.

Customers on the Verge of Churn

- Take proactive measures such as feedback surveys to understand their concerns.
- Offer customized discounts or support to re-engage them.
- Use personalized communication to rebuild trust and interest.

Lost Customers

- Perform churn analysis to identify why they stopped purchasing.
- Develop win-back campaigns based on past preferences and behaviors.
- Use these insights to enhance retention strategies for similar customer profiles in the future.