

MRA PROJECT

ISSAC ABRAHAM

CONTENTS

- Problem Statement
- About Data
- Exploratory Analysis
- RFM Analysis
- Inferences and Recommendations

EXECUTIVE SUMMARY

- The automobile parts manufacturing company lacks an in-house data science team and has hired us as consultants.
- Our task is to analyse three years' worth of transaction data to identify buying patterns, provide insights about customers, and recommend customized marketing strategies.
- This presentation will cover the problem statement, an overview of the data, exploratory analysis findings, RFM analysis results, and final recommendations.

SHAPE AND INFO OF THE DATASET

- The dataset contains 2747 entries with 20 columns, including information such as order number, quantity ordered, price each, sales amount, order date, and customer details. Data types range from integers and floats to objects, indicating categorical variables. It has 2747 rows and 20 columns.

Shape of the dataset:
(2747, 20)

```
Information about the dataset:
<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   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
dtypes: float64(2), int64(6), object(12)
memory usage: 429.3+ KB
None
```


SUMMARY STATISTICS

- The summary statistics reveal valuable insights about the numeric columns in the dataset. For instance, the average sales amount per transaction is approximately \$3,553, with a standard deviation of \$1,838, indicating considerable variability in transaction values. Additionally, the dataset spans a time period from ORDERDATE 43106 to 43982, suggesting continuous transaction records over several years.

Summary statistics of numeric columns:

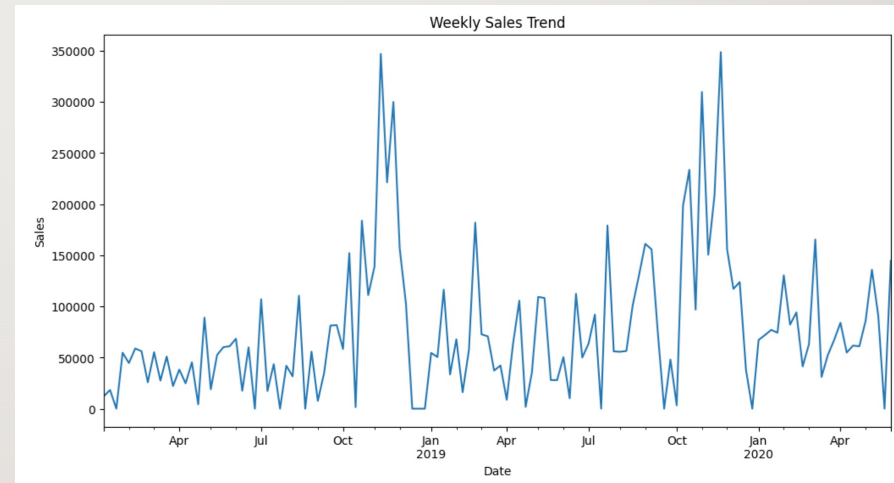
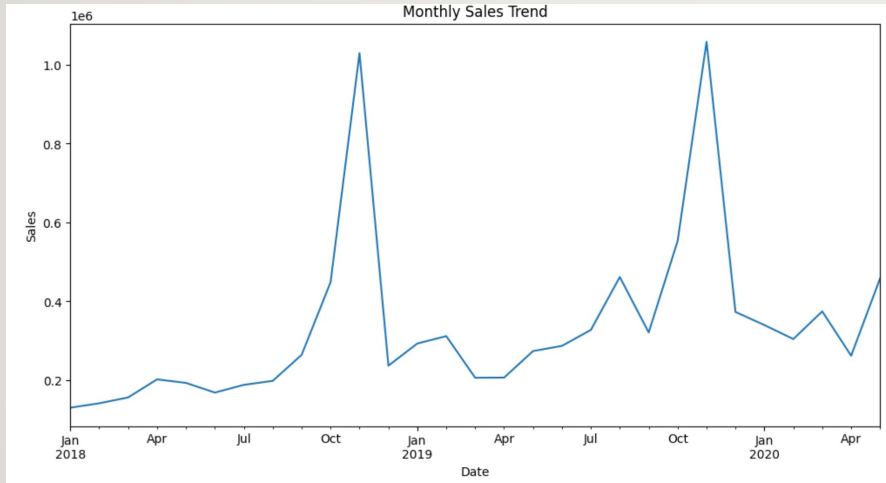
	ORDERNUMBER	QUANTITYORDERED	PRICEEACH	ORDERLINENUMBER \
count	2747.000000	2747.000000	2747.000000	2747.000000
mean	10259.761558	35.103021	101.098951	6.491081
std	91.877521	9.762135	42.042548	4.230544
min	10100.000000	6.000000	26.880000	1.000000
25%	10181.000000	27.000000	68.745000	3.000000
50%	10264.000000	35.000000	95.550000	9.000000
75%	10334.500000	43.000000	127.100000	18.000000
max	10425.000000	97.000000	252.870000	18.000000

click to scroll output

	SALES	ORDERDATE	DAYS_SINCE_LASTORDER	MSRP
count	2747.000000	2747.000000	2747.000000	2747.000000
mean	3553.047583	43598.914088	1757.085912	100.691664
std	1838.953901	230.231295	819.280576	40.114802
min	482.130000	43106.000000	42.000000	33.000000
25%	2204.350000	43412.000000	1077.000000	68.000000
50%	3184.800000	43640.000000	1761.000000	99.000000
75%	4503.095000	43786.000000	2436.500000	124.000000
max	14082.800000	43982.000000	3562.000000	214.000000

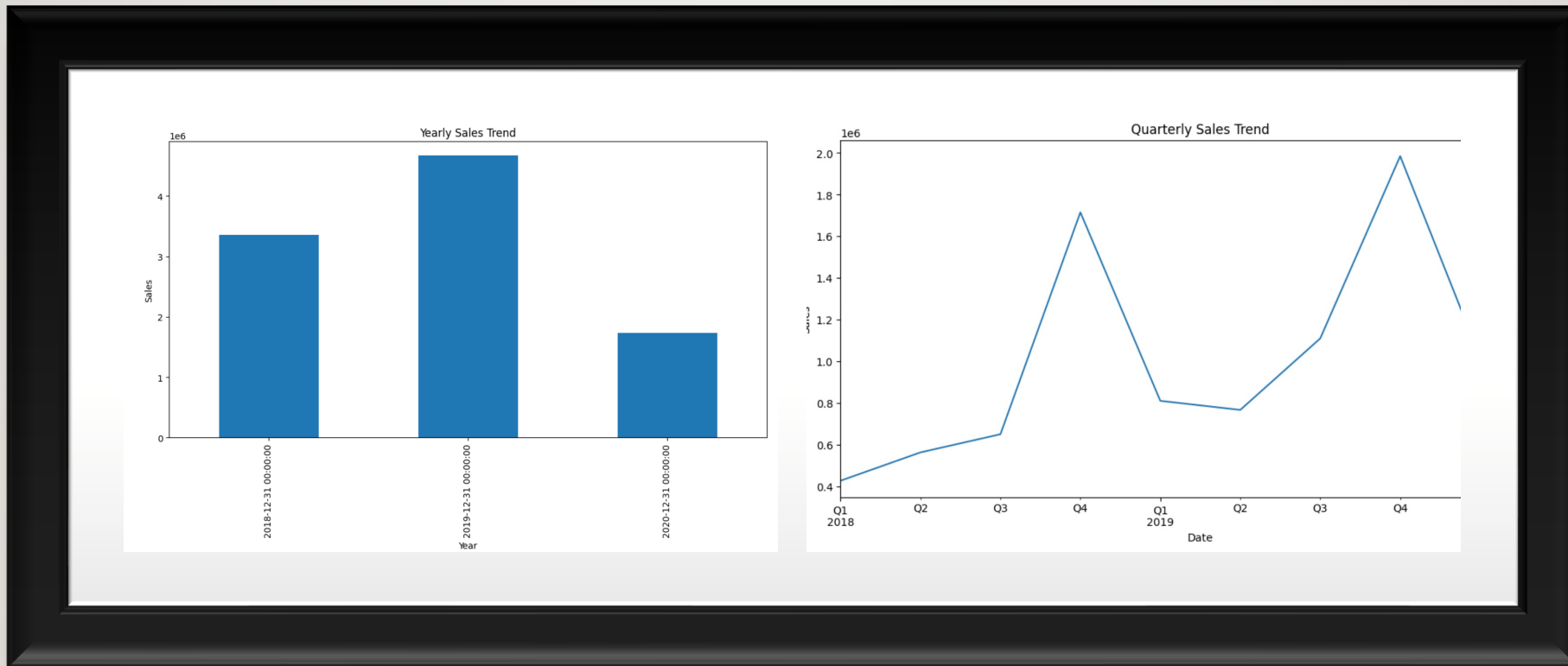
ASSUMPTIONS OF THE DATA

1. **Distribution Pattern:** Numeric variables like quantity ordered, price each, and sales amount may exhibit a typical distribution, given their proximity of mean and median.
2. **Stable Order Frequency:** The order date's low standard deviation implies consistent transaction volumes over time, indicating stable business operations.
3. **Product Range Diversity:** Varied MSRP values suggest a diverse array of products spanning different price categories.
4. **Customer Engagement Levels:** The range of values for DAYS_SINCE_LASTORDER reflects varying degrees of customer activity, with some making frequent purchases and others showing longer intervals between orders.



WEEKLY & MONTHLY TRENDS IN SALES

WE CAN SEE THAT THE OVER SALES IN THE YEAR PLUMMETED IN NOVEMBER.

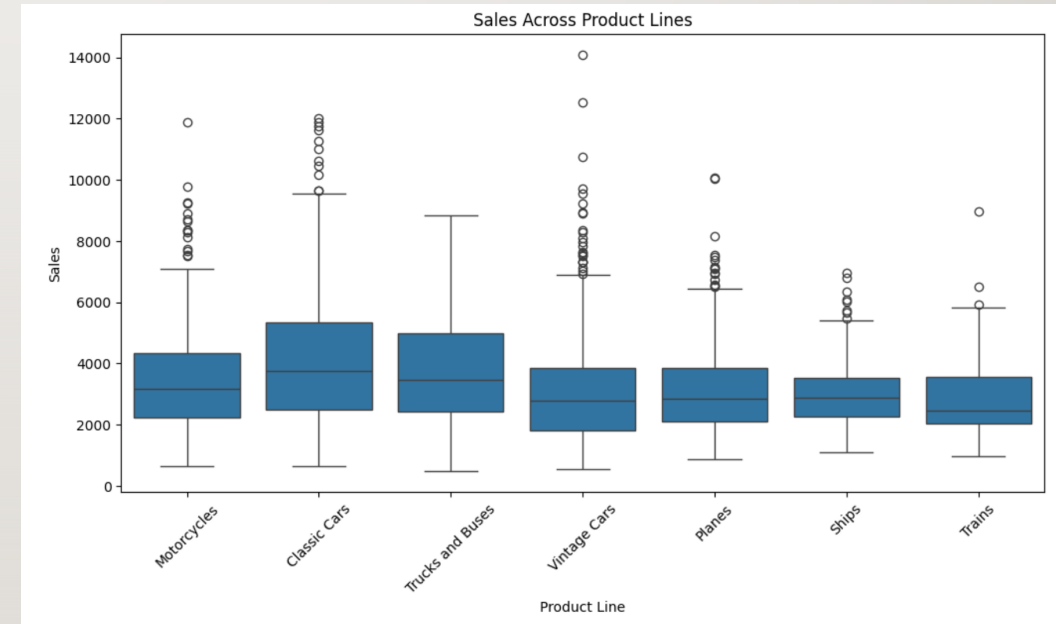


QUARTERLY AND YEARLY SALES TREND

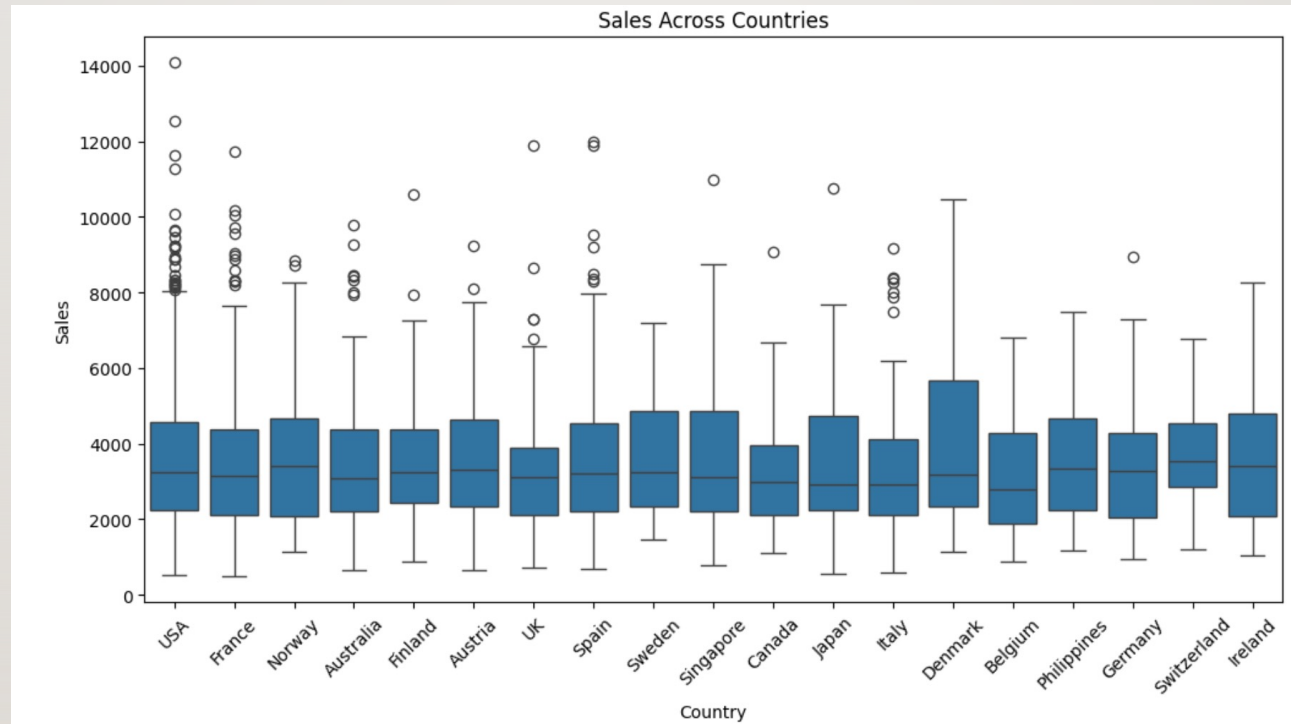
Year 2019 showed the maximum sales after 2018 and there is declining sales in 2020, Also throughout the years the fourth quarter showed a steep in sales.

BIVARIATE ANALYSIS: SALES ACROSS DIFFERENT CATEGORIES

- We can see that the median sales across classic cars category is the highest followed by Trucks and buses and motorcycles.



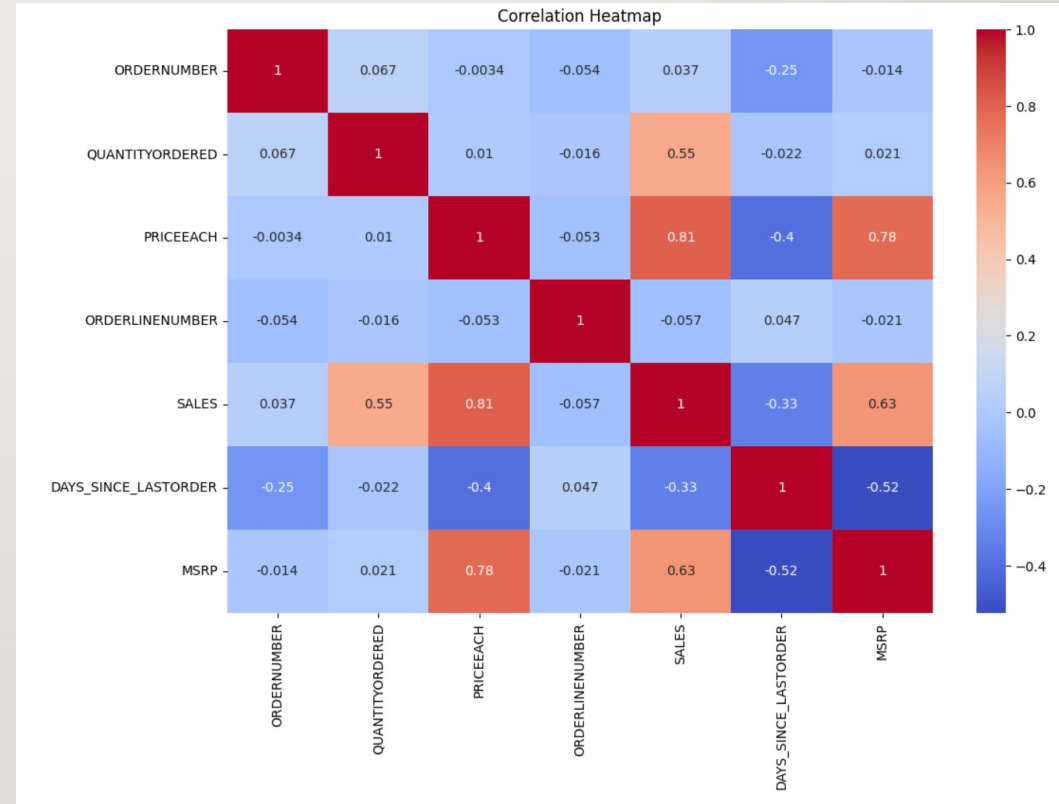
BIVARIATE ANALYSIS: SALES ACROSS COUNTRIES



The median sales across most of the countries remain in the range of 2000-5000

CORRELATION HEAT MAP

WE CAN SEE THAT THE VARIABLES PRICE
OF EACH ITEM AND MSRP ARE HIGHLY
CORRELATED TO OVERALL SALES



INFERENCES

- November experienced a notable decline in overall sales, signaling a potential seasonal trend.
- Classic cars emerged as the top-selling category, followed closely by trucks, buses, and motorcycles.
- Median sales remained consistent across most countries, typically falling within the range of 2000 to 5000.
- The prices of each item and Manufacturer's Suggested Retail Price (MSRP) demonstrated a significant correlation with overall sales, indicating their influential role in driving revenue.

WHAT IS RFM?

PARAMETERS & TOOLS USED

- RFM stands for Recency, Frequency, and Monetary Value.
- It is a method used for analyzing customer value based on their transaction history.
- Customer name, Quantity ordered, order date, Price each, Sales and order quantity are the parameters in the dataset.
- KNIME was used for RFM scores

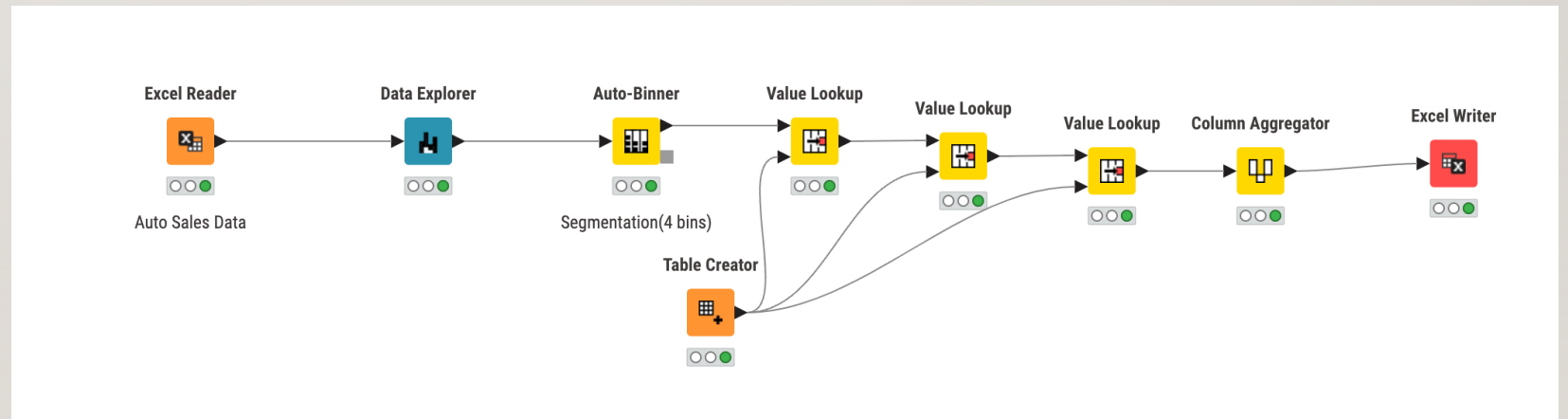


ASSUMPTIONS MADE

- Recent customers are more likely to purchase again.
- Frequent customers are more engaged and loyal.
- Customers who spend more are likely to be higher-value customers.

KNIME WORKFLOW

- We have done the data preprocessing. Data segmentation is done by auto binner node converting to 4 bins



POSTALC... String	COUNTRY String	CONTAC... String	CONTAC... String	DEALSIZE String	ORDERN... String	SALES [Bi... String	DAYS_SI... String	Recency String	Frequency String	Monetary String	Concaten... String	
10022	USA	Yu	Kwai	Small	Bin 1	Bin 2	Bin 1	?	?	?	?, ?, ?	
51100	France	Henriot	Paul	Small	Bin 1	Bin 2	Bin 1	?	?	?	?, ?, ?	
75508	France	Da Cunha	Daniel	Medium	Bin 1	Bin 3	Bin 1	?	?	3	?, ?, 3	
90003	USA	Young	Julie	Medium	Bin 1	Bin 3	Bin 1	?	?	3	?, ?, 3	
94217	USA	Hirano	Juri	Medium	Bin 1	Bin 3	Bin 1	?	?	3	?, ?, 3	
59000	France	Rance	Martine	Small	Bin 1	Bin 2	Bin 1	?	?	?	?, ?, ?	
N 5804	Norway	Oeztan	Veysel	Medium	Bin 2	Bin 4	Bin 1	?	?	4	?, ?, 4	
75016	France	Perrier	Dominique	Medium	Bin 2	Bin 4	Bin 1	?	?	4	?, ?, 4	
3004	Australia	Ferguson	Peter	Medium	Bin 2	Bin 3	Bin 1	?	?	3	?, ?, 3	
10022	USA	Frick	Michael	Small	Bin 2	Bin 2	Bin 1	?	?	?	?, ?, ?	
94019	USA	Brown	William	Medium	Bin 2	Bin 3	Bin 1	?	?	3	?, ?, 3	

OUTPUT TABLE HEAD

TOP 5 BEST CUSTOMERS

- Top customers like Order Number 10348 and 10358 made high-value purchases, demonstrating frequent buying behavior and recent activity with RFM scores of 2, 4, 4.
- Customers such as Order Number 10371 and 10383 contributed substantially to revenue with significant spending and high-value transactions, reflecting RFM scores of 2, 4, 4.

Best Customers:

	ORDERNUMBER	QUANTITYORDERED	PRICEEACH	ORDERLINENUMBER	SALES	\
1605	10348	37	161.66	1	5981.42	
1606	10358	41	138.65	7	5684.65	
1681	10371	45	123.24	8	5545.80	
1682	10383	40	152.24	3	6089.60	
1704	10339	55	112.99	12	6214.45	

	ORDERDATE	DAYS_SINCE_LASTORDER	STATUS	PRODUCTLINE	MSRP	...	\
1605	NaN		1818	Shipped	Trucks and Buses	127	...
1606	NaN		1780	Shipped	Trucks and Buses	127	...
1681	NaN		1811	Shipped	Classic Cars	35	...
1682	NaN		1782	Shipped	Classic Cars	35	...
1704	NaN		1895	Shipped	Planes	68	...

	CONTACTLASTNAME	CONTACTFIRSTNAME	DEALSIZE	ORDERNUMBER [Binned]	\
1605	Sommer	Martin	Medium	Bin 4	
1606	Freyre	Diego	Medium	Bin 4	
1681	Nelson	Valarie	Medium	Bin 4	
1682	Freyre	Diego	Medium	Bin 4	
1704	Shimamura	Akiko	Medium	Bin 4	

	SALES [Binned]	DAYS_SINCE_LASTORDER [Binned]	Recency	Frequency	Monetary	\
1605	Bin 4		Bin 3	2.0	4.0	4.0
1606	Bin 4		Bin 3	2.0	4.0	4.0
1681	Bin 4		Bin 3	2.0	4.0	4.0
1682	Bin 4		Bin 3	2.0	4.0	4.0
1704	Bin 4		Bin 3	2.0	4.0	4.0

	RFM
1605	2, 4, 4
1606	2, 4, 4
1681	2, 4, 4
1682	2, 4, 4
1704	2, 4, 4

TOP 5 CHURNING CUSTOMERS

- Churning customers like Order Number 10301 and 10284 exhibited declining engagement, with infrequent purchases and diminishing monetary value, reflected in RFM scores of 1, 3, 4 and 1, 3, 3 respectively.
- Customers such as Order Number 10276 and 10299 showed signs of decreased activity and potential disengagement, indicating a need for targeted retention strategies, as indicated by RFM scores of 1, 3, 3.

Churning Customers:

	ORDERNUMBER	QUANTITYORDERED	PRICEEACH	ORDERLINENUMBER	SALES	\
1921	10301	50	154.47	11	7723.50	
2243	10284	50	81.86	4	4093.00	
2268	10276	38	113.28	13	4304.64	
2270	10299	38	115.32	7	4382.16	
2271	10309	50	84.70	3	4235.00	

	ORDERDATE	DAYS_SINCE_LASTORDER	STATUS	PRODUCTLINE	MSRP	...	\
1921	NaN	2526	Shipped	Classic Cars	140	...	
2243	NaN	2528	Shipped	Vintage Cars	68	...	
2268	NaN	2572	Shipped	Motorcycles	102	...	
2270	NaN	2515	Shipped	Motorcycles	102	...	
2271	NaN	2501	Shipped	Motorcycles	102	...	

	CONTACTLASTNAME	CONTACTFIRSTNAME	DEALSIZE	ORDERNUMBER [Binned]	\
1921	Klaeboe	Jan	Large	Bin 3	
2243	Klaeboe	Jan	Medium	Bin 3	
2268	Barajas	Miguel	Medium	Bin 3	
2270	Karttunen	Matti	Medium	Bin 3	
2271	Bergulfsen	Jonas	Medium	Bin 3	

	SALES [Binned]	DAYS_SINCE_LASTORDER [Binned]	Recency	Frequency	Monetary	\
1921	Bin 4		Bin 4	1.0	3.0	4.0
2243	Bin 3		Bin 4	1.0	3.0	3.0
2268	Bin 3		Bin 4	1.0	3.0	3.0
2270	Bin 3		Bin 4	1.0	3.0	3.0
2271	Bin 3		Bin 4	1.0	3.0	3.0

	RFM
1921	1, 3, 4
2243	1, 3, 3
2268	1, 3, 3
2270	1, 3, 3
2271	1, 3, 3

5 LOST CUSTOMERS

1. Lost customers such as Order Number 10302 and 10306 exhibited declining engagement, with infrequent purchases and diminishing monetary value, reflected in RFM scores of 2, 3, 3.
2. Customers like Order Number 10315 and 10322 showed signs of decreased activity and potential disengagement, indicating a need for targeted re-engagement strategies, as indicated by RFM scores of 2, 3, 3.

Lost Customers:						
	ORDERNUMBER	QUANTITYORDERED	PRICEEACH	ORDERLINENUMBER	SALES	\
1232	10302	38	89.27	2	3392.26	
1552	10306	29	110.60	7	3207.40	
1553	10315	35	120.43	6	4215.05	
1576	10322	30	116.67	4	3500.10	
1601	10305	24	132.90	10	3189.60	
	ORDERDATE	DAYS_SINCE_LASTORDER	STATUS	PRODUCTLINE	MSRP	... \
1232	NaN	1836	Shipped	Vintage Cars	92	...
1552	NaN	1783	Shipped	Ships	122	...
1553	NaN	1769	Shipped	Ships	122	...
1576	NaN	1786	Shipped	Vintage Cars	44	...
1601	NaN	1833	Shipped	Trucks and Buses	127	...
	CONTACTLASTNAME	CONTACTFIRSTNAME	DEALSIZE	ORDERNUMBER	[Binned]	\
1232	Devon	Elizabeth	Medium		Bin 3	
1552	Ashworth	Victoria	Medium		Bin 3	
1553	Labrune	Janine	Medium		Bin 3	
1576	Young	Valarie	Medium		Bin 3	
1601	Hernandez	Marta	Medium		Bin 3	
	SALES	[Binned]	DAYS_SINCE_LASTORDER	[Binned]	Recency	Frequency Monetary \
1232		Bin 3		Bin 3	2.0	3.0 3.0
1552		Bin 3		Bin 3	2.0	3.0 3.0
1553		Bin 3		Bin 3	2.0	3.0 3.0
1576		Bin 3		Bin 3	2.0	3.0 3.0
1601		Bin 3		Bin 3	2.0	3.0 3.0
	RFM					
1232	2,	3,	3			
1552	2,	3,	3			
1553	2,	3,	3			
1576	2,	3,	3			
1601	2,	3,	3			

TOP 5 LOYAL CUSTOMERS

- Loyal customers like Order Numbers 10348 and 10371 exhibit consistent high-value purchases with frequent transactions, reflected in RFM scores of 2, 4, 4.
- Customers such as Order Number 10339 demonstrate strong engagement and consistent buying behavior, indicating a valuable long-term relationship, as indicated by RFM scores of 2, 4, 4.

Loyal Customers:						
	ORDERNUMBER	QUANTITYORDERED	PRICEEACH	ORDERLINENUMBER	SALES	\
1605	10348	37	161.66	1	5981.42	
1606	10358	41	138.65	7	5684.65	
1681	10371	45	123.24	8	5545.80	
1682	10383	40	152.24	3	6089.60	
1704	10339	55	112.99	12	6214.45	
	ORDERDATE	DAYS_SINCE_LASTORDER	STATUS	PRODUCTLINE	MSRP	...
1605	NaN	1818	Shipped	Trucks and Buses	127	...
1606	NaN	1780	Shipped	Trucks and Buses	127	...
1681	NaN	1811	Shipped	Classic Cars	35	...
1682	NaN	1782	Shipped	Classic Cars	35	...
1704	NaN	1895	Shipped	Planes	68	...
	CONTACTLASTNAME	CONTACTFIRSTNAME	DEALSIZE	ORDERNUMBER	[Binned]	\
1605	Sommer	Martin	Medium		Bin 4	
1606	Freyre	Diego	Medium		Bin 4	
1681	Nelson	Valarie	Medium		Bin 4	
1682	Freyre	Diego	Medium		Bin 4	
1704	Shimamura	Akiko	Medium		Bin 4	
	SALES	[Binned]	DAYS_SINCE_LASTORDER	[Binned]	Recency	Frequency
1605	Bin 4		Bin 3		2.0	4.0
1606	Bin 4		Bin 3		2.0	4.0
1681	Bin 4		Bin 3		2.0	4.0
1682	Bin 4		Bin 3		2.0	4.0
1704	Bin 4		Bin 3		2.0	4.0
	RFM					
1605	2,	4,	4			
1606	2,	4,	4			
1681	2,	4,	4			
1682	2,	4,	4			
1704	2,	4,	4			