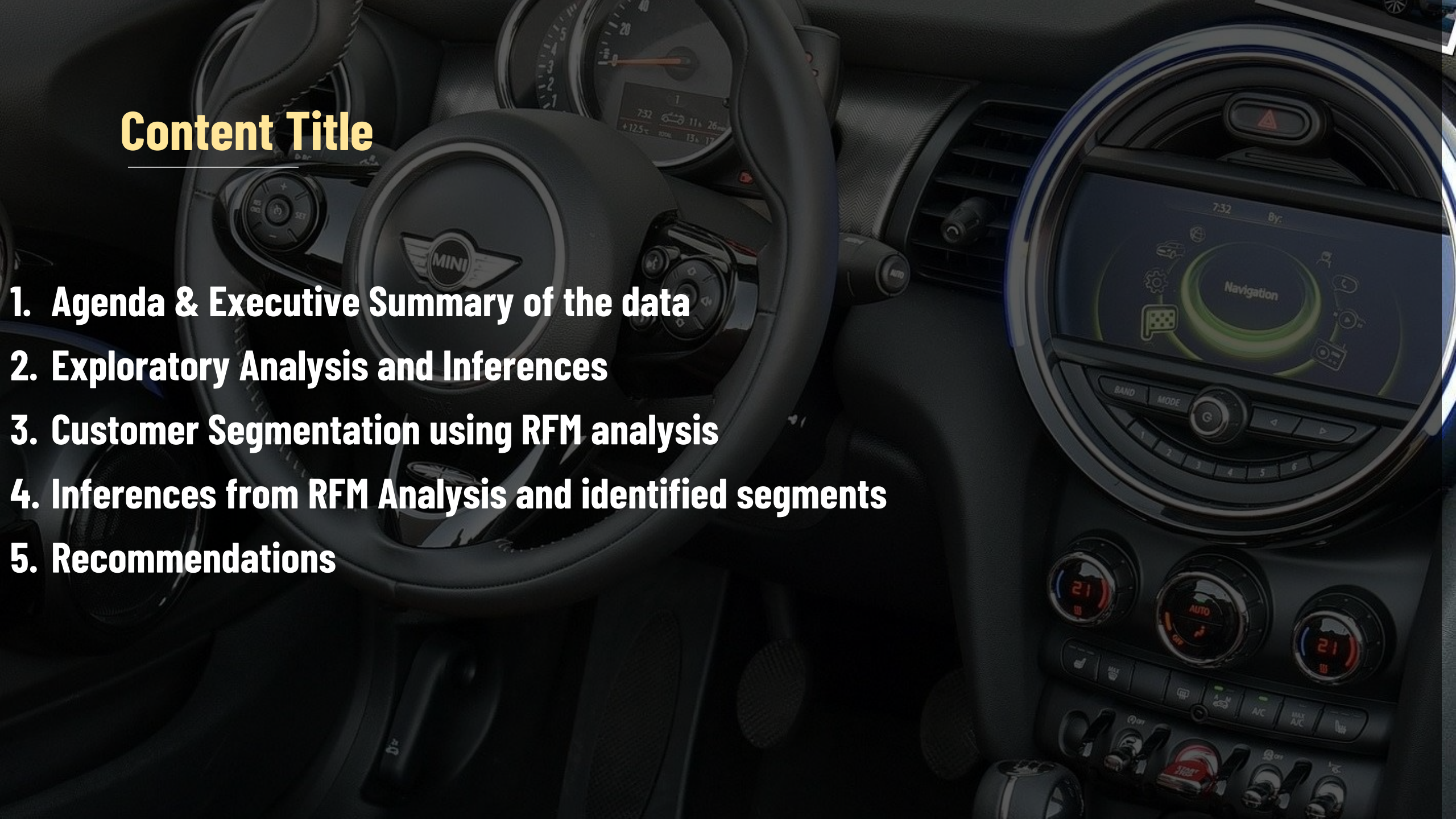
Marketing and Retail Analysis – Project 1

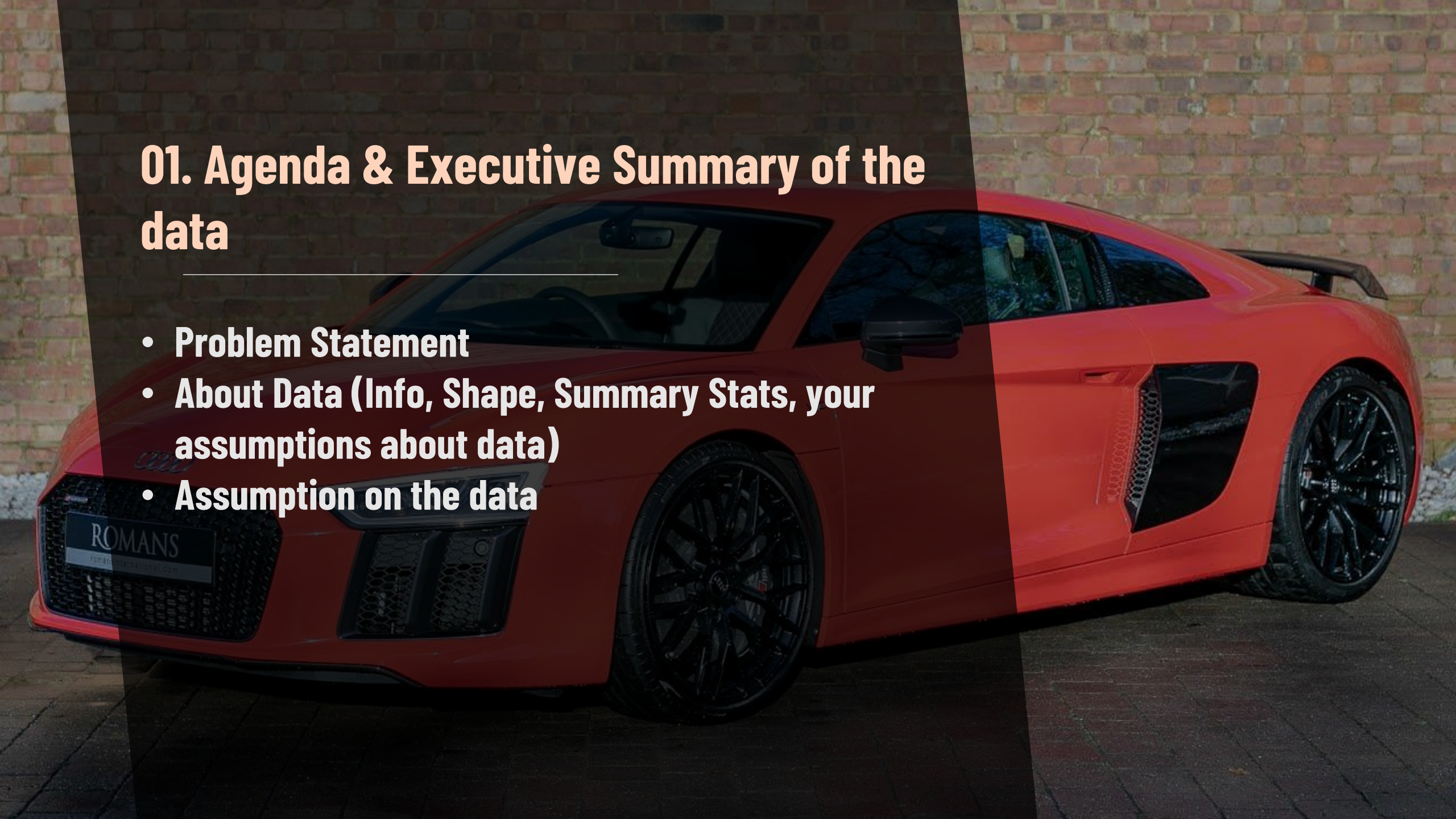
Automobile parts manufacturing company

Soumalya Sen
BATCH: PGPDSBA.O.JULY23.A



Content Title

1. Agenda & Executive Summary of the data
2. Exploratory Analysis and Inferences
3. Customer Segmentation using RFM analysis
4. Inferences from RFM Analysis and identified segments
5. Recommendations

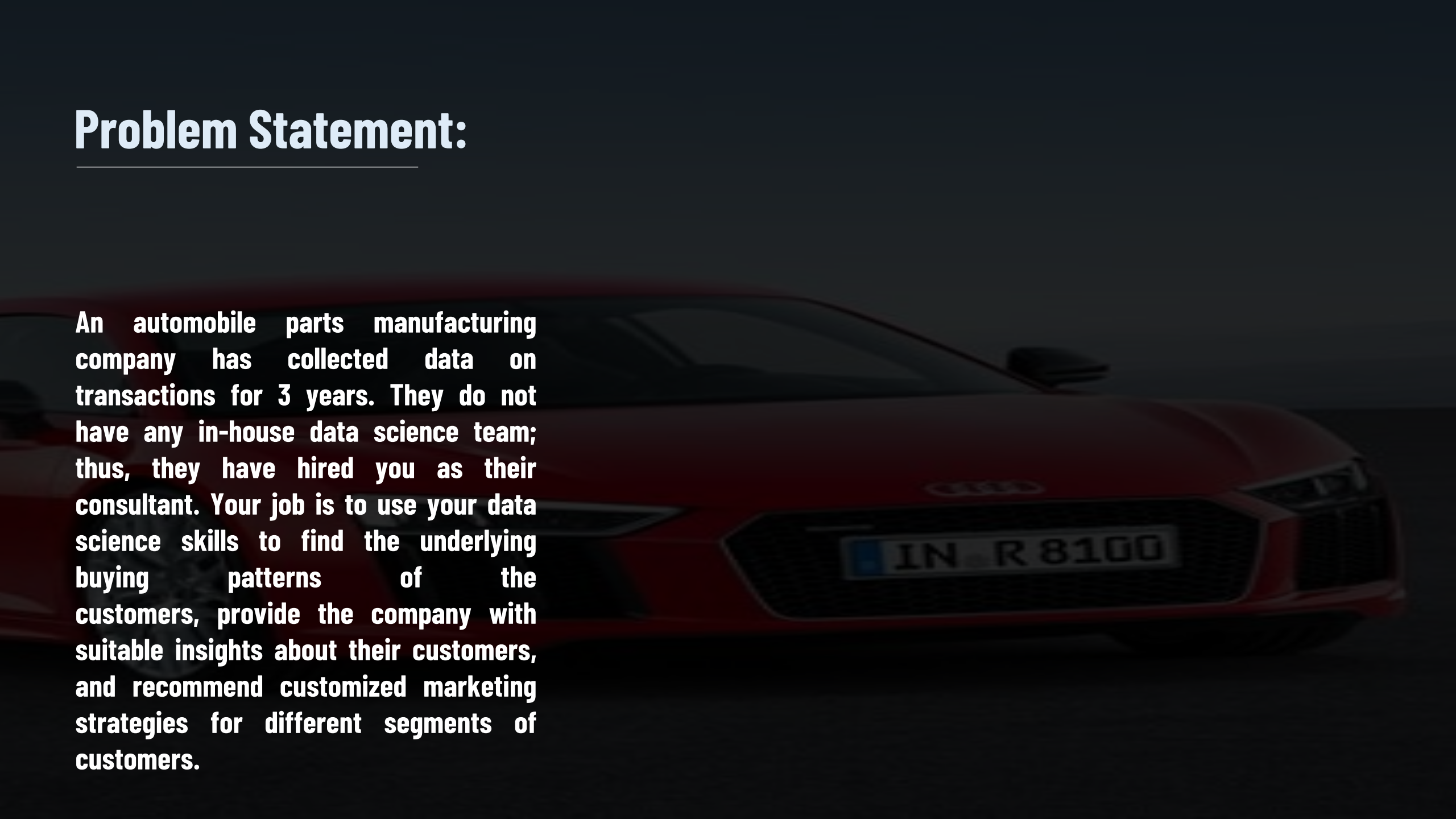


01. Agenda & Executive Summary of the data

- Problem Statement
- About Data (Info, Shape, Summary Stats, your assumptions about data)
- Assumption on the data

Problem Statement:

An automobile parts manufacturing company has collected data on 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 data science skills to find the underlying buying patterns of the customers, provide the company with suitable insights about their customers, and recommend customized marketing strategies for different segments of customers.

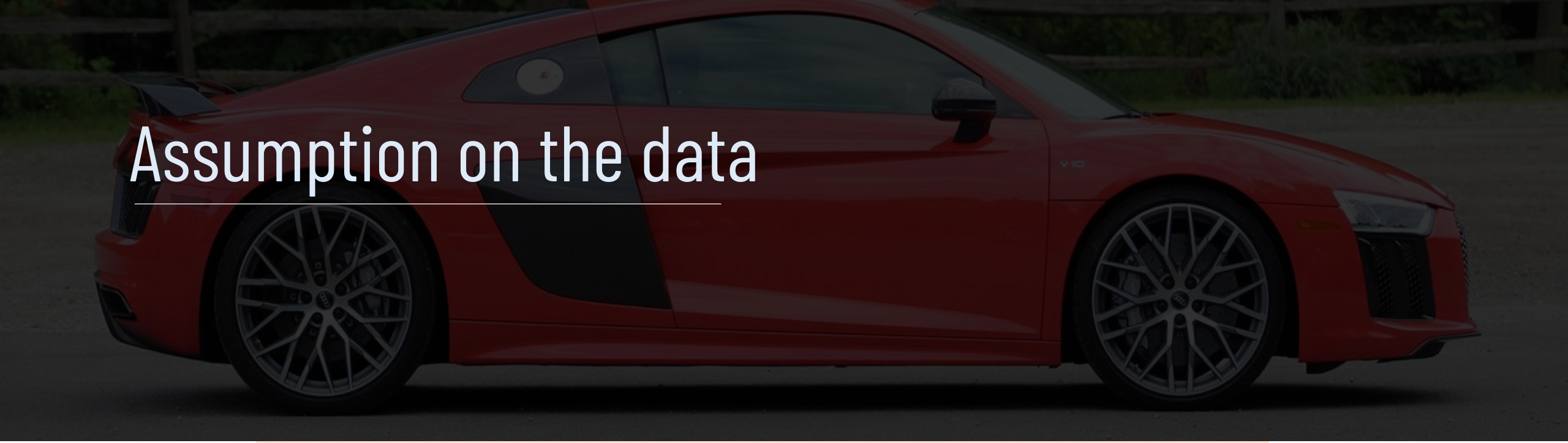


Executive Summary of the data:

Objective: The objective of this analysis was to identify the underlying buying patterns of customers, provide insightful information about customer behaviour, and recommend customized marketing strategies for different customer segments.

- The dataset consists of 2,747 transactions made by customers of an automobile parts manufacturing company.
- The dataset consists of 20 columns, including 2 float, 6 integer, and 12 object data types, providing a diverse range of information about the transactions and customers.
- The dataset is free from duplicate entries.
- Customers on average ordered 35 items per transaction, with a minimum of 6 and a maximum of 97.
- The average number of days since the last order is 1,757.09, with a minimum of 42 days and a maximum of 3,562 days.
- The average Manufacturer's Suggested Retail Price (MSRP) of the products is $100.69, with a standard deviation of $40.11.

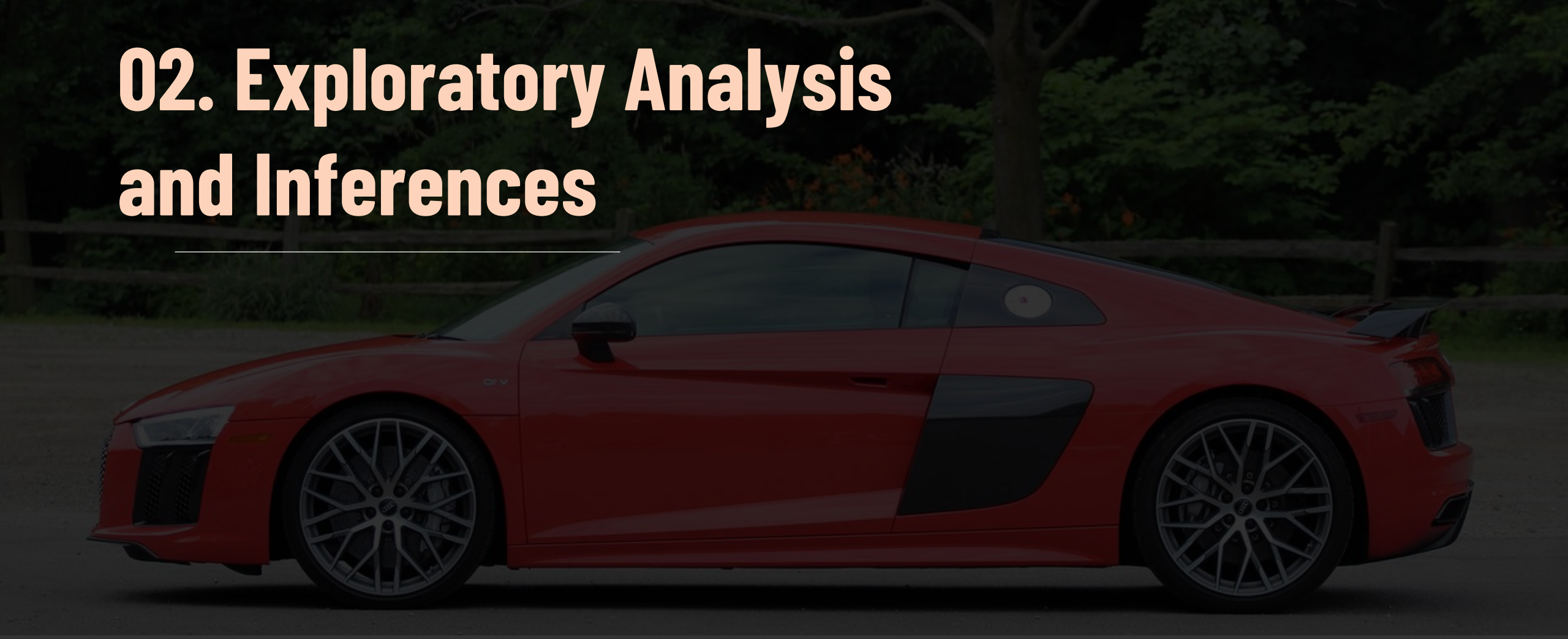
RFM (Recency, Frequency, Monetary) analysis is a powerful technique used to analyse customer behaviour based on their transaction history. By evaluating the recency of their last purchase, the frequency of their transactions, and the monetary value of their purchases, RFM analysis helps businesses identify and segment their customers into different groups to tailor marketing strategies and optimize customer retention efforts.

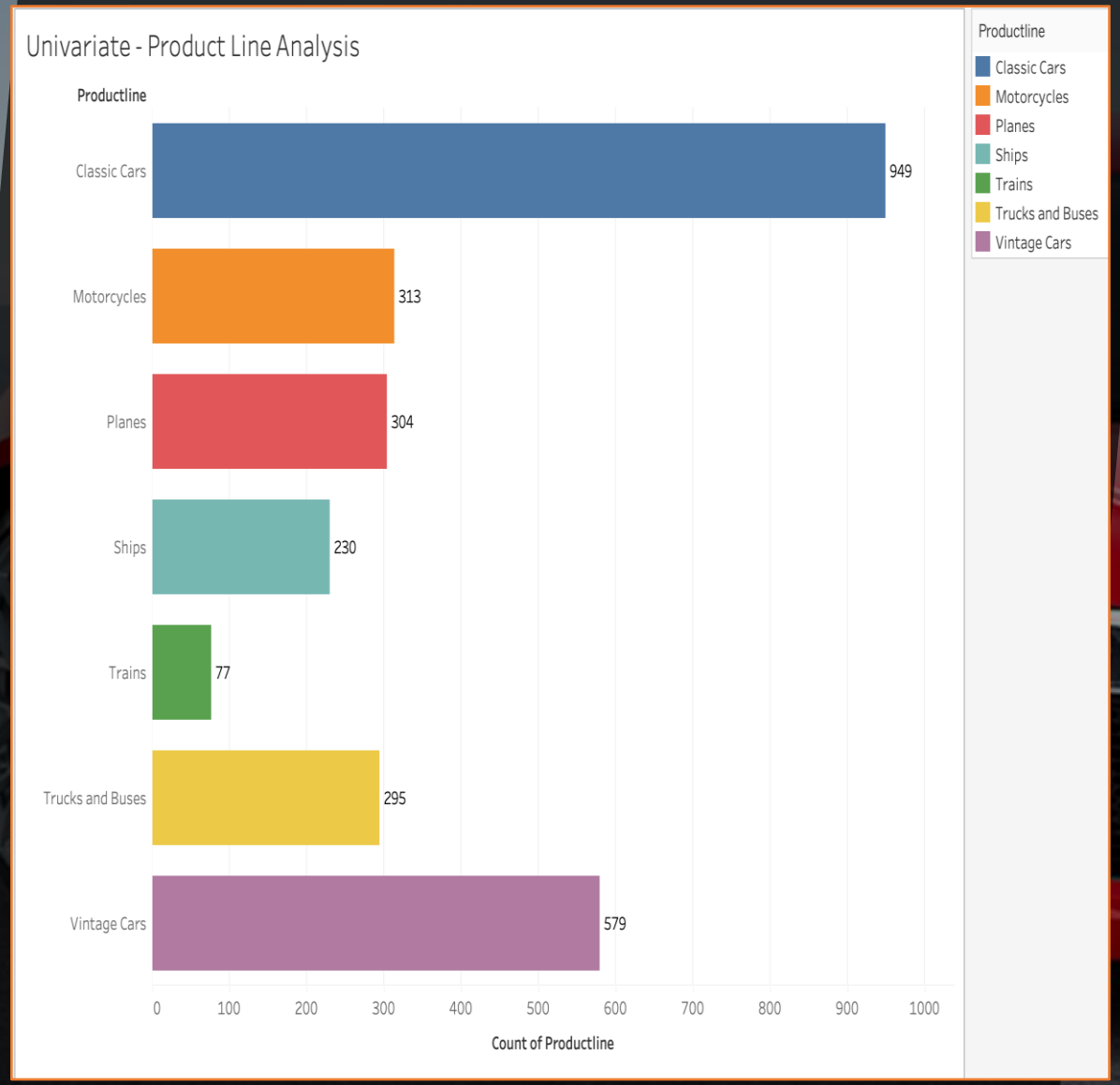


Assumption on the data

The dataset contains information about orders placed by customers. Each order is identified by a unique ORDERNUMBER.
The dataset contains information about customers' transactions and their associated details, such as the number of items purchased, days since the last order, and the MSRP of the products. This implies that the dataset includes transactional data related to customer purchases.
The dataset does not contain duplicate entries. This means that each transaction entry is unique, and there are no repeated records for the same transaction or customer.

02. Exploratory Analysis and Inferences





Product Line Analysis

Inferences :

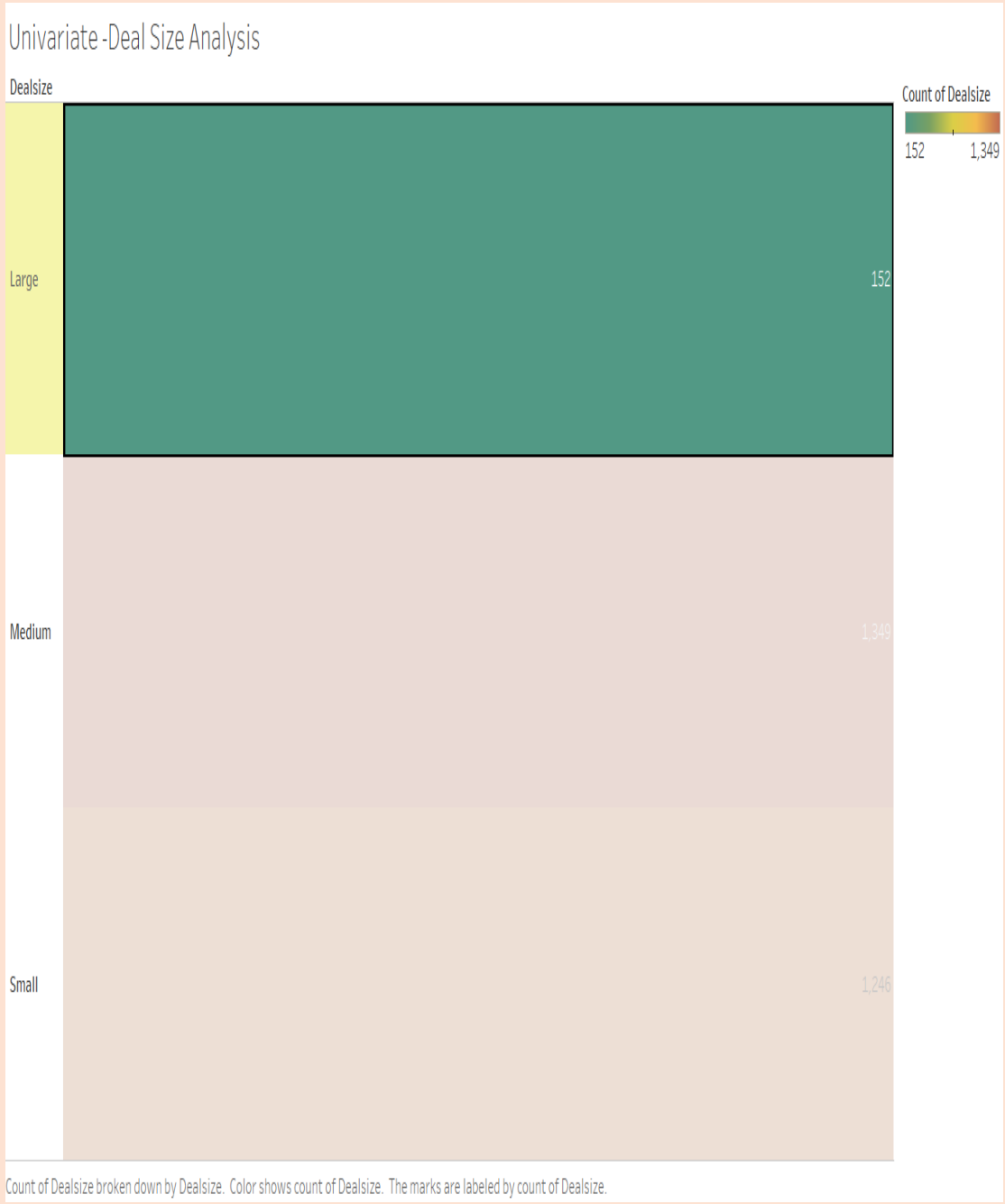
- Classic cars and vintage cars are the most popular product lines, indicating a strong demand for these vehicle models.

- Ships, trucks, buses, and trains have relatively lower counts, suggesting a niche market or specialized customer base for these product lines.

Deal Size Analysis

Inferences :

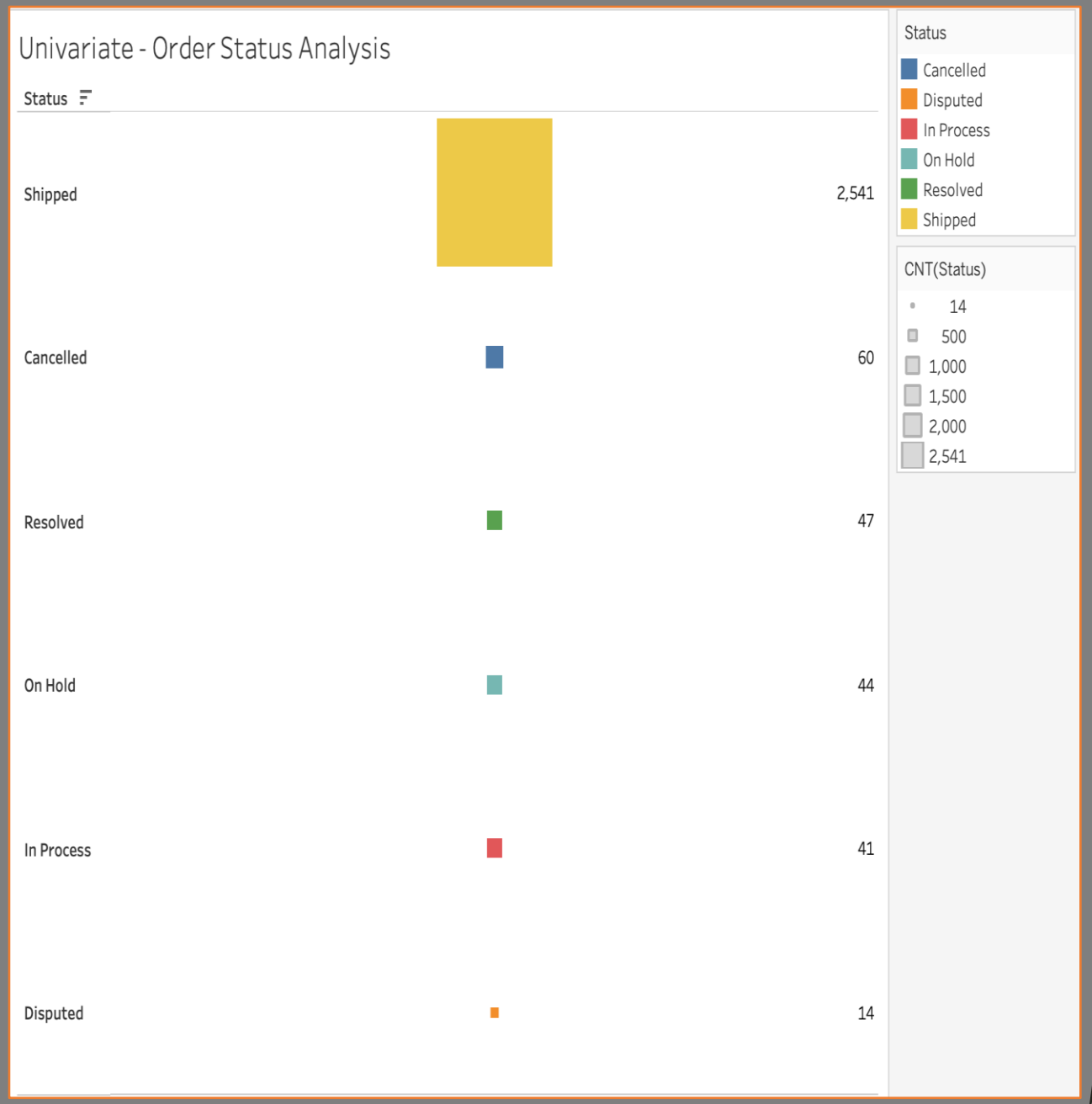
- The majority of the deals fall into the small and medium categories, with counts of 1,246 and 1,349, respectively. This indicates that the company primarily engages in transactions of moderate to smaller deal sizes.



Order Status Analysis

Inferences :

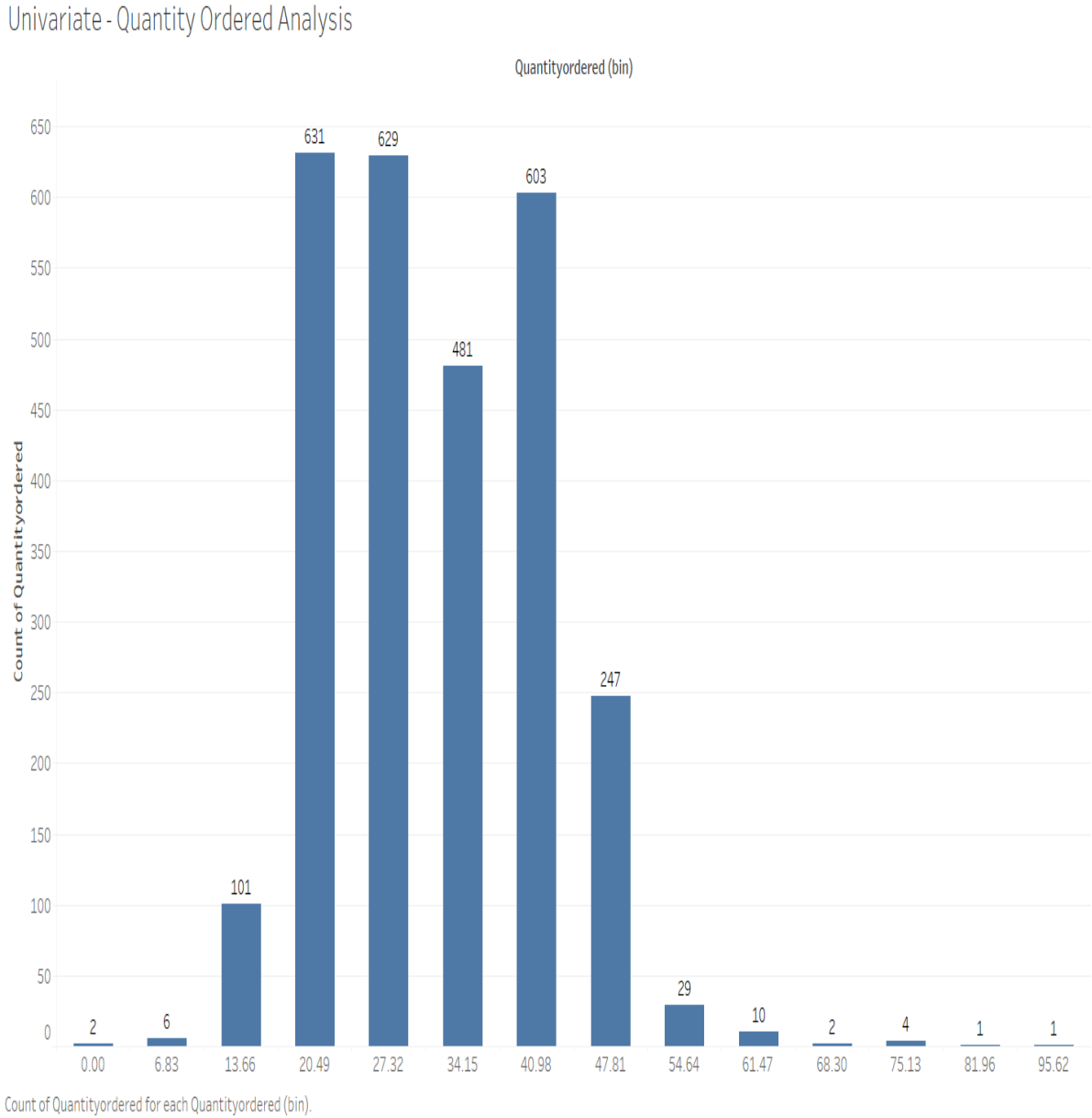
The significant count of 2,541 successfully shipped orders reflects the company's outstanding order fulfilment capabilities and commitment to delivering orders promptly.



Quantity Ordered Analysis

Inferences :

The bins with quality orders of 21, 28, 35, and 42 have the highest counts, indicating their popularity among customers. With counts ranging from 576 to 631, these quantities meet the demands of a significant portion of customers

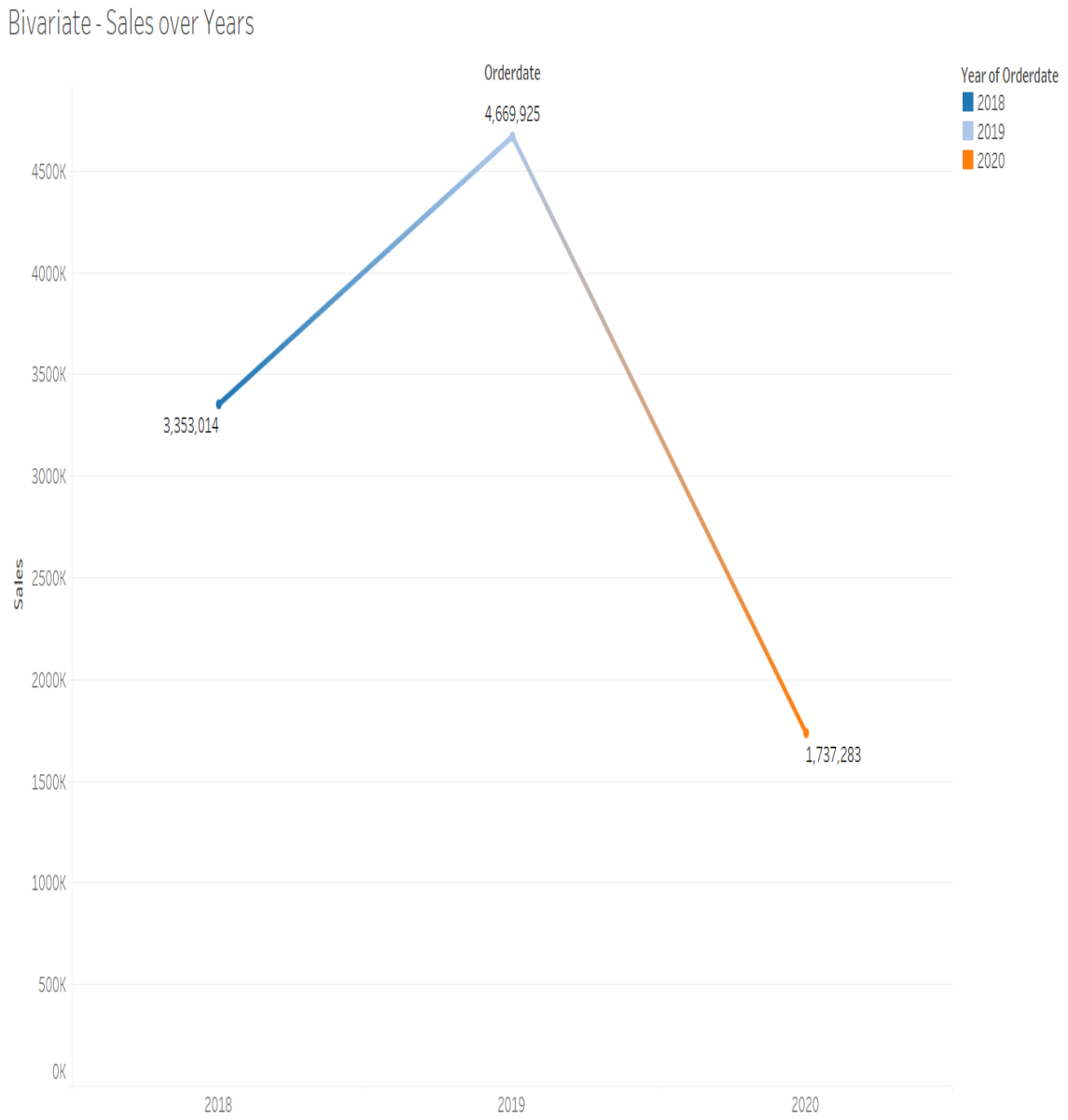


Sales over Years

Inference:

* Sales showed a positive trend and grew significantly from 2018 (3,353,014) to 2019 (4,669,925), suggesting that the business performed well and experienced an increase in customer demand.

* However, since we only have sales data until May for 2020 (1,737,283), it is difficult to make conclusive statements about the entire year.

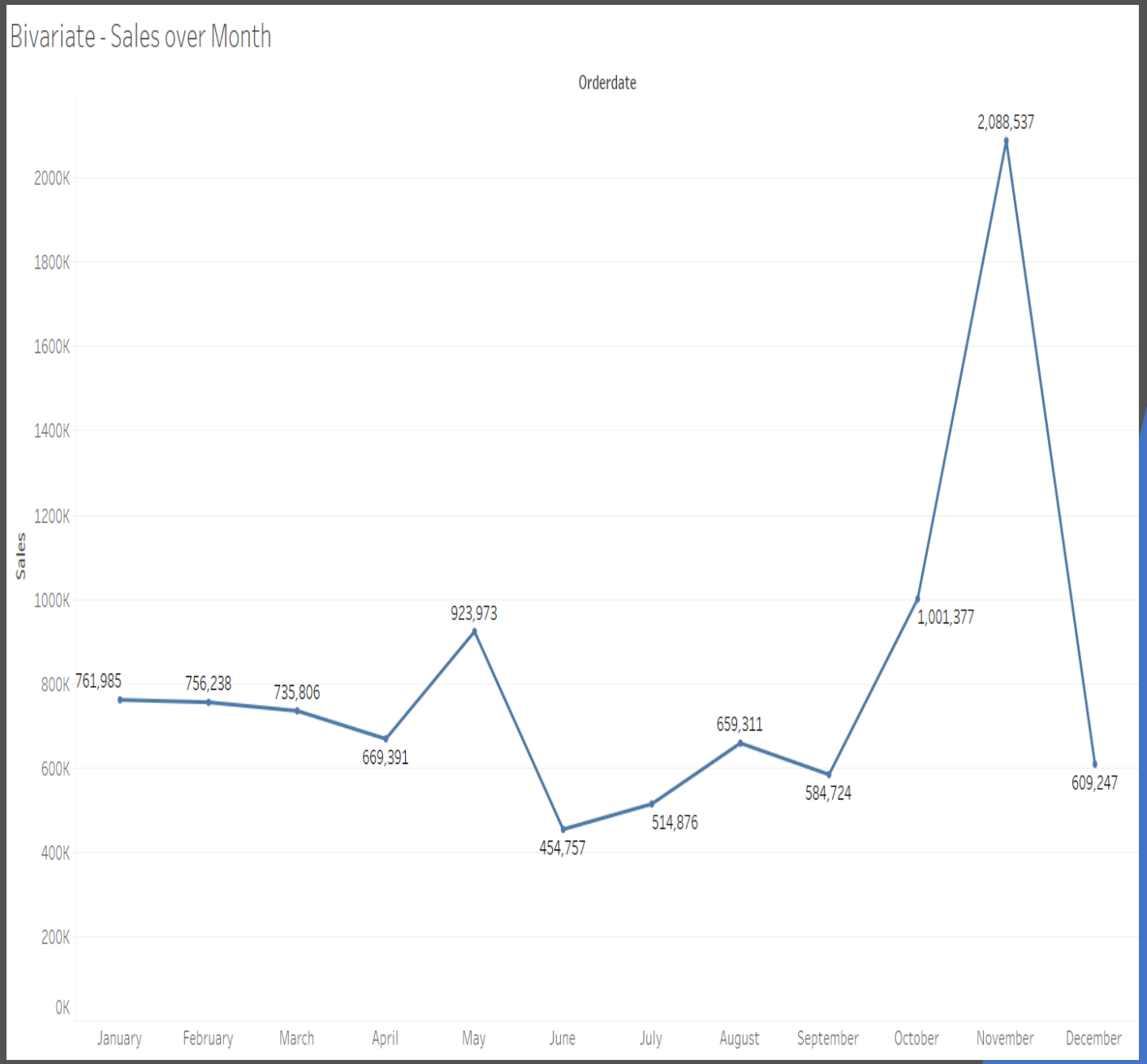


Sales over Month

Inference:

* The sales data exhibits fluctuations throughout the year, with peak sales recorded in November (2,088,537) and October (1,001,377), indicating heightened customer activity likely influenced by seasonal factor.

* In contrast, June records the lowest sales (454,757), suggesting a potential decrease in customer demand during that period.

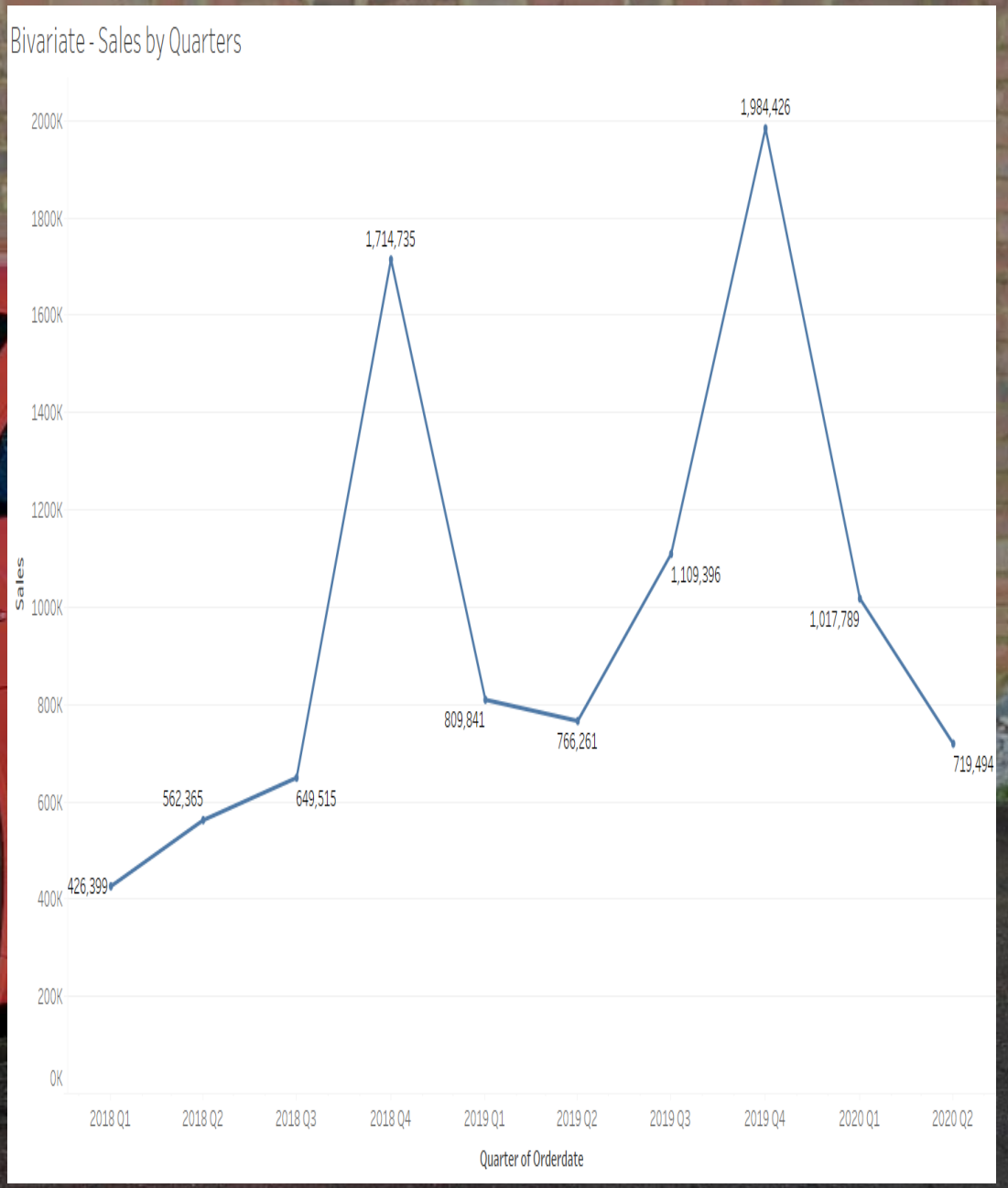


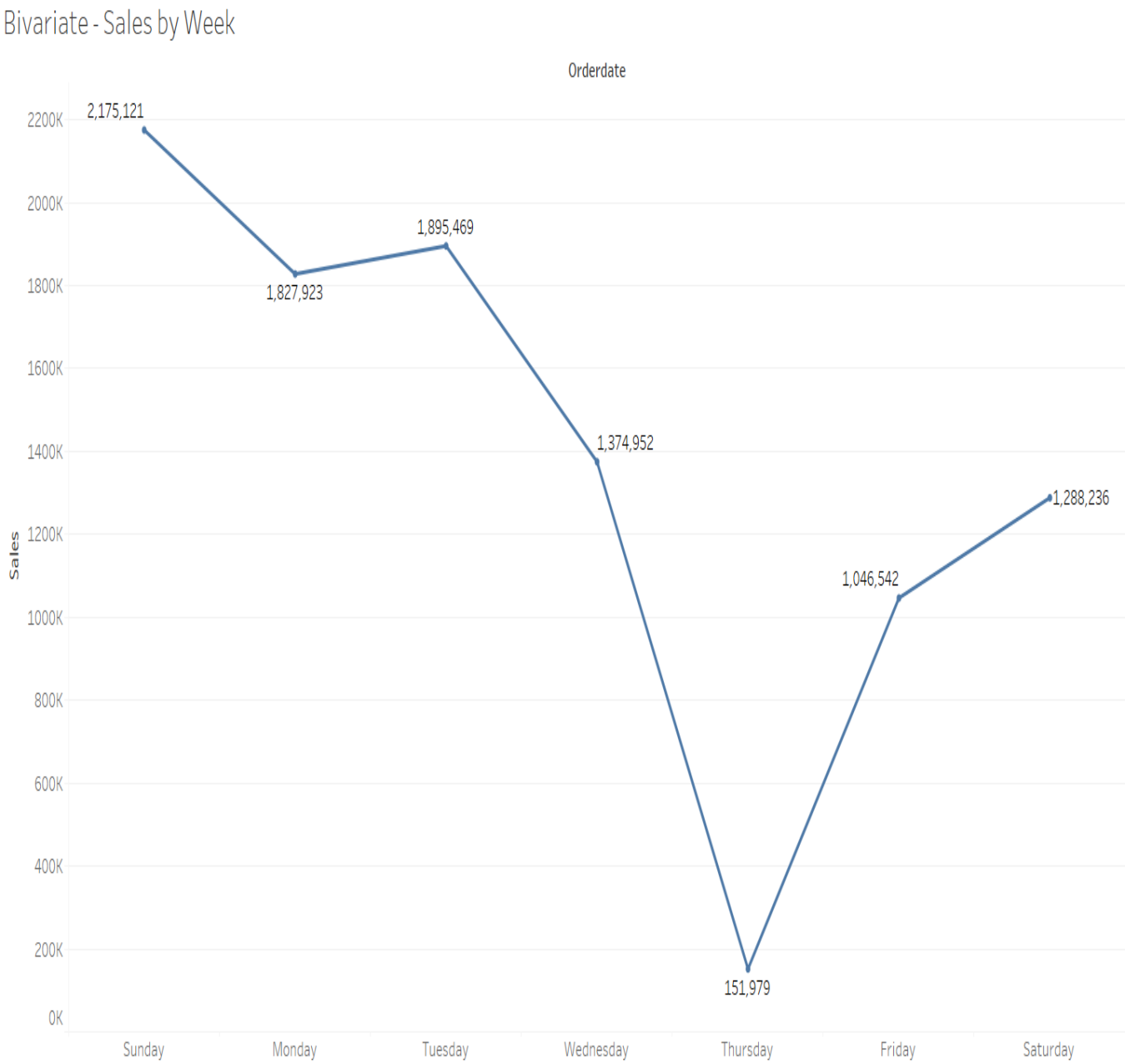
Sales by Quarters

Inference:

* The quarterly sales data over the three-year period reveals a consistent pattern of higher sales in Q4 (3,699,161) compared to other quarters. This indicates a potential seasonal trend where the end of the year witnesses increased customer spending.

* Q1 (2,254,029) and Q2 (2,048,120) show relatively stable sales figures, while Q3 (1,758,911) records the lowest sales among the quarters.





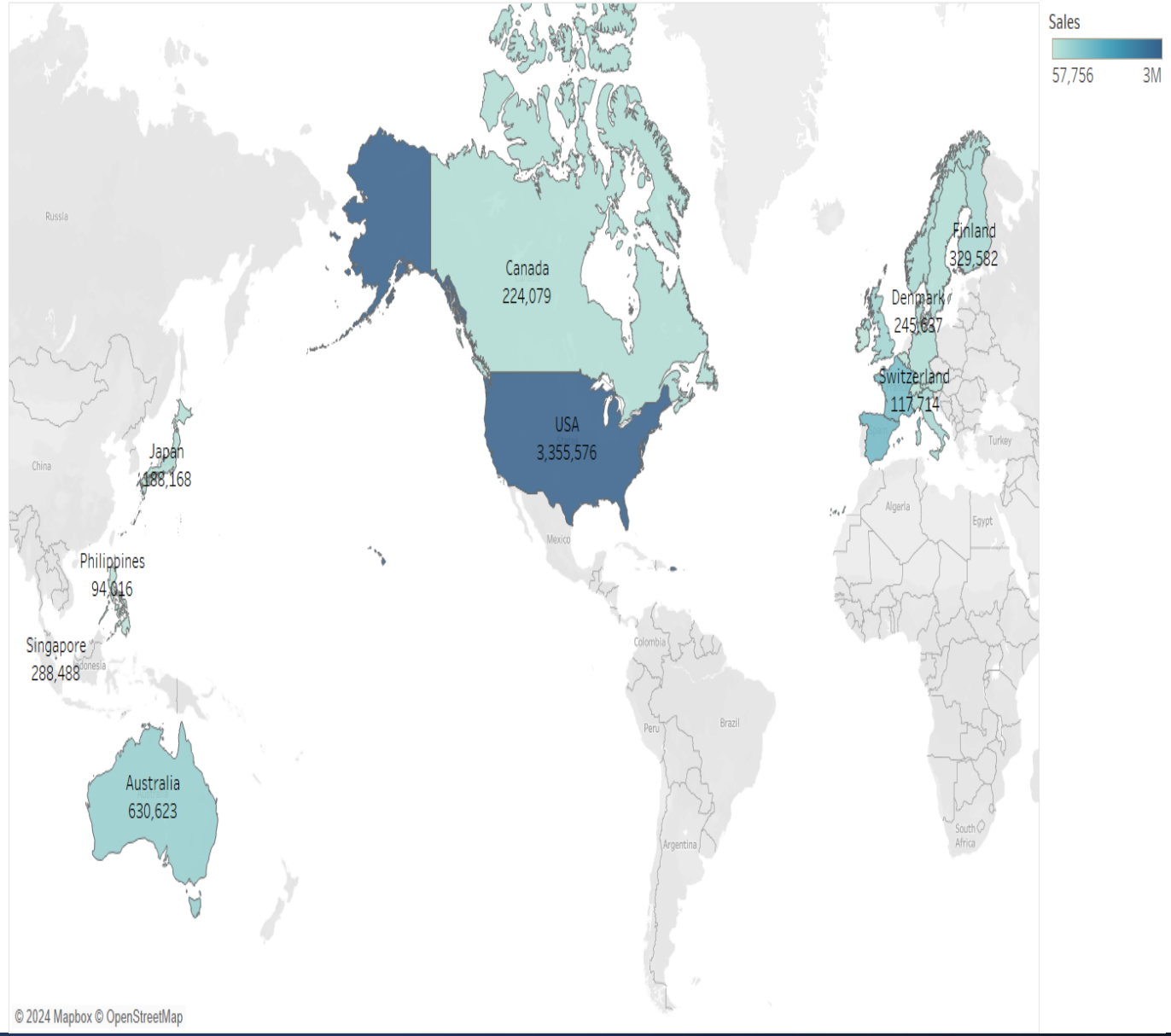
Sales by Week

Inference:

* Sundays have consistently shown the highest sales (2,175,121) over the three-year period, indicating that customers tend to be more active and likely prefer shopping on weekends.

* Wednesdays, on the other hand, have the lowest sales (1,374,952) among the weekdays, suggesting that customers may be less engaged or influenced by factors that affect their buying behaviour on this specific day.

Bivariate - Sales by Country

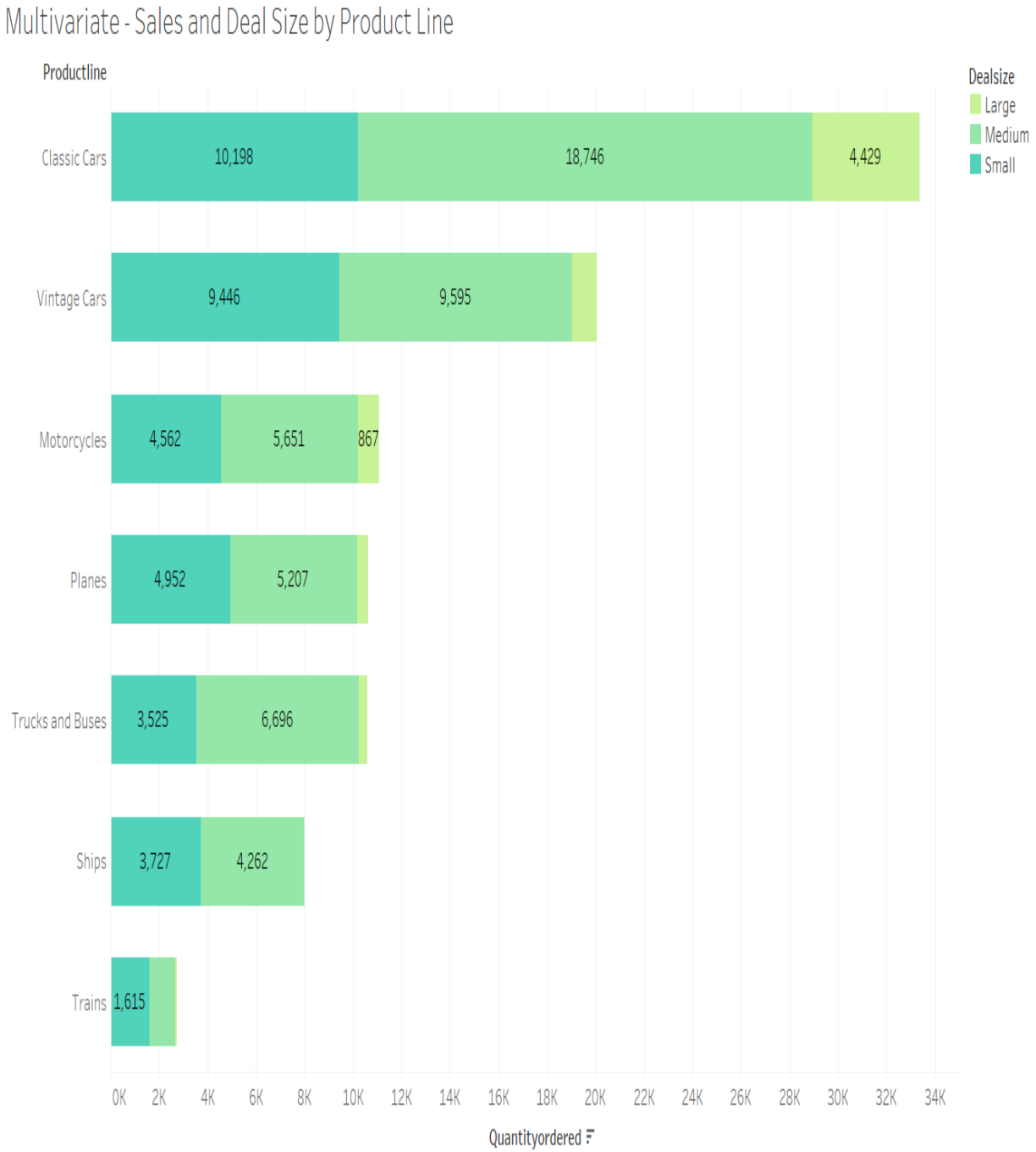


Sales by Country

Inference:

* The sales data by country indicates that the USA (3,355,576) has significantly higher sales compared to other countries, suggesting a strong market presence and potentially a larger customer base.

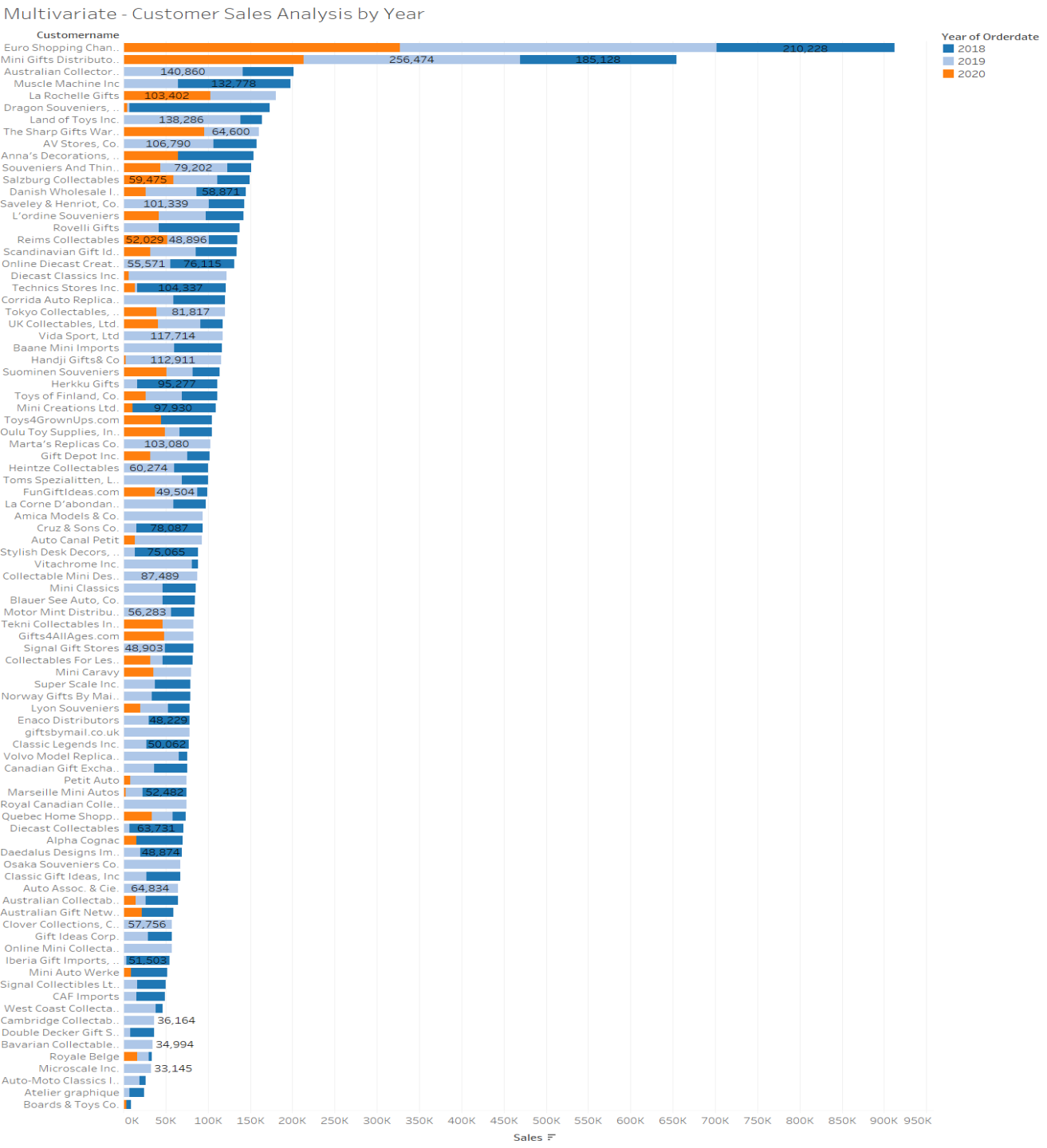
* Canada (224,079), Singapore (288,488), Denmark (245,637), and Finland (329,582) also demonstrate notable sales figures, indicating significant market opportunities and customer demand in these respective countries



Sales and Deal size by Product line

Inference:

The graph reveals that Classic Cars and Vintage Cars are the most frequently ordered product lines, followed by Motorcycles, Planes, Trucks & Buses, and Ships. In contrast, Trains had the lowest number of orders. Interestingly, across all product lines, most orders fell into the medium and small deal size categories. Higher deal sizes were relatively less common among the popular product lines.



Customer Sales Analysis by Year (2018-2020):

1. Euro Shopping Channel:
   - Highest sales across all three years.
   - Maintained consistent sales performance.
2. Mini Gift Distributor Ltd:
   - Second-highest sales.
   - Demonstrated steady sales growth over the years.
3. Australian Collectors Co:
   - Ranked third in sales.
   - No sales recorded in 2020.
4. Muscle Machine Inc:
   - Fourth in terms of sales.
   - No sales recorded in 2020.
5. La Rochelle Gifts:
   - Joined as a new company in 2019.
   - Showed promising sales in the year of establishment.

Summary:

- **Product Line Analysis**: Classic cars and vintage cars are the most popular product lines, indicating a strong demand for these vehicle models. Conversely, ships, trucks, buses, and trains have relatively lower counts, suggesting a niche market or specialized customer base for these product lines.

- **Deal Size Analysis:** The majority of the deals fall into the small and medium categories, with counts of 1,246 and 1,349, respectively. This indicates that the company primarily engages in transactions of moderate to smaller deal sizes.

- **Order Status Analysis:** The significant count of 2,541 successfully shipped orders reflects the company's outstanding order fulfilment capabilities and commitment to delivering orders promptly.

- **Quality Ordered Analysis:** The bins with quality orders of 21, 28, 35, and 42 have the highest counts, indicating their popularity among customers. With counts ranging from 576 to 631, these quantities meet the demands of a significant portion of customers.

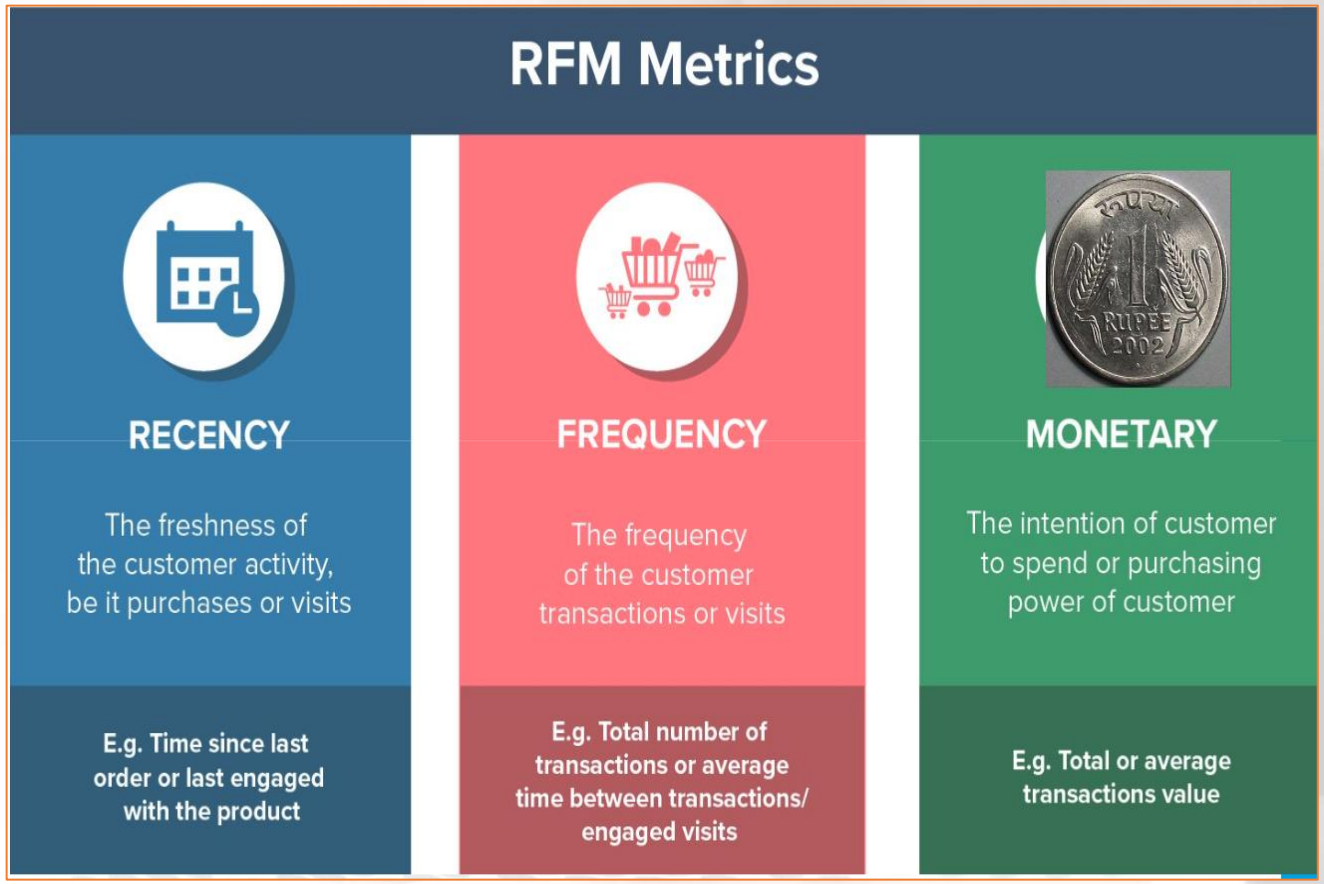
- **Sales Analysis:** Sales showed a positive trend and grew significantly from 2018 (3,353,014) to 2019 (4,669,925), suggesting that the business performed well and experienced an increase in customer demand. However, since we only have sales data until May for 2020 (1,737,283), it is difficult to make conclusive statements about the entire year.

03. 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?

RFM

RFM analysis is a customer segmentation technique commonly used in marketing and retail to analyse and categorize customers based on their purchasing behaviour. RFM stands for Recency, Frequency, and Monetary Value, which are three key metrics used to evaluate customer engagement and profitability.



PARAMETERS

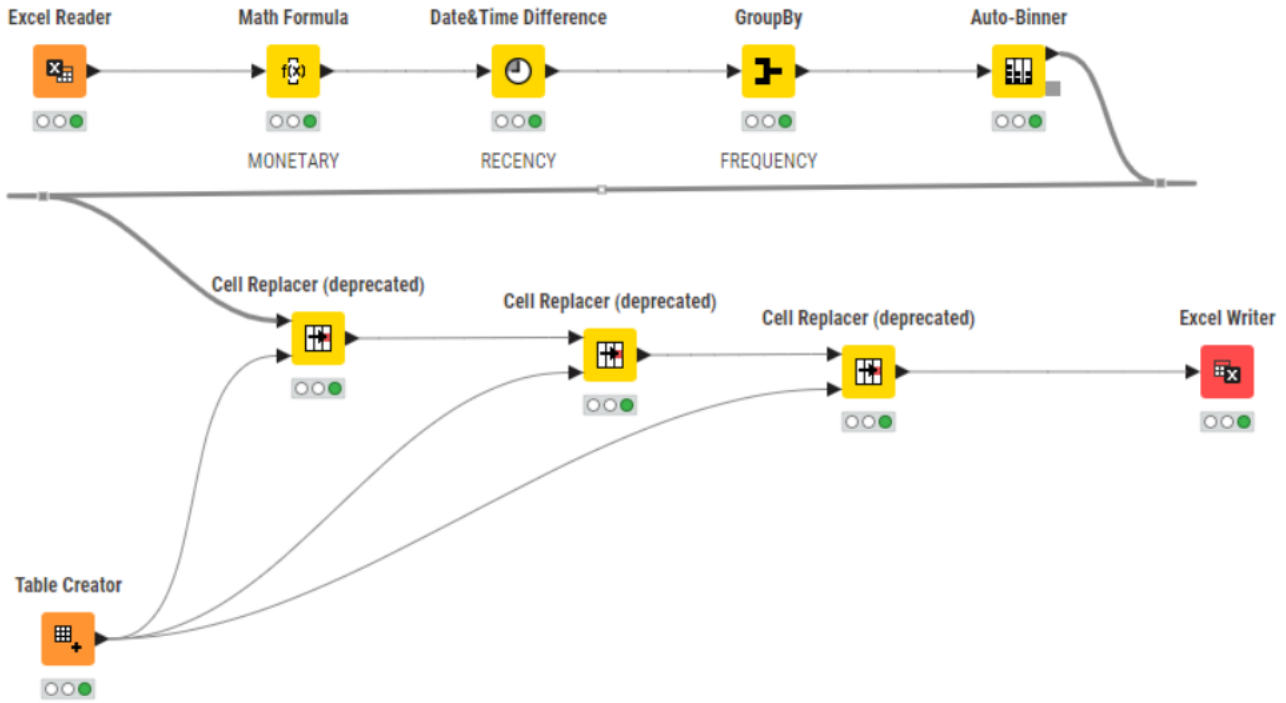
- **Monetary**: Calculated by multiplying the price each with the quantity ordered. It represents the monetary value or revenue generated by each customer.
- **Recency**: Determined by subtracting the order date from a fixed reference date, such as 01-06-2020. It represents how recently a customer made a purchase.
- **Frequency:** Calculated as the count of customer names. It represents how often a customer has made purchases.
- **Auto-binning:** Customers are segmented into four categories based on their RFM scores - High, Moderate, Low, and Very Low. This segmentation helps in categorizing customers based on their value and behaviour.

These parameters are utilized in RFM analysis to evaluate customer behaviour, identify customer segments, and make data-driven marketing and sales decisions.

ASSUMPTION:

- Higher Monetary value indicates a higher spending customer: The assumption is that customers who generate higher monetary value through their purchases are likely to be more valuable and potentially more profitable for the business.
- Recent purchases are more indicative of customer engagement: The assumption is that customers who have made purchases more recently are likely to be more engaged with the company and its offerings. They may have higher potential for repeat purchases or upselling/cross-selling opportunities.
- Higher Frequency of purchases reflects customer loyalty: The assumption is that customers who make purchases more frequently are more loyal to the company. They may have a stronger connection to the brand, higher customer satisfaction, and a higher likelihood of recommending the company to others.

KNIME WORKFLOW



Tool Used

Open for Innovation
KNIME

Output table head

Row ID	OD...	PHONE	ADDRESSLI...	CITY	POST...	COUN...	CONT...	CONT...	DEALS...	MONE...	RECE...	ORDE...	MONE...	RECE...	RECE...	MONE...	FREQ...
Row0	108	(171) 555-1555	Fauntleroy Circus	Manchester	EC2 5NT	51	Ashworth	Victoria	Medium	157,807.81	197	Bin 4	Bin 4	Bin 3	Low	High	High
Row1	757	61.77.6555	1 rue Alsace-L...	Toulouse	31000	20	Roulet	Annette	Medium	70,488.44	65	Bin 1	Bin 1	Bin 1	High	Very Low	Very Low
Row2	949	011-4988555	Via Monte Bian...	Torino	10100	26	Accorti	Paolo	Large	94,117.26	266	Bin 2	Bin 3	Bin 4	Very Low	Moderate	Low
Row3	949	02 9936 8555	201 Miller Street	North Syd...	2060	46	O'Hara	Anna	Small	153,996.13	84	Bin 4	Bin 4	Bin 2	Moderate	High	High
Row4	016	40.32.2555	54, rue Royale	Nantes	44000	7	Schmitt	Carine	Medium	24,179.96	189	Bin 1	Bin 1	Bin 3	Low	Very Low	Very Low
Row5	342	61-9-3844-6...	7 Allen Street	Glen Wave...	3150	23	Connery	Sean	Medium	64,591.46	23	Bin 2	Bin 1	Bin 1	High	Very Low	Low
Row6	678	03 9520 4555	636 St Kilda Ro...	Melbourne	3004	55	Ferguson	Peter	Medium	200,995.41	185	Bin 4	Bin 4	Bin 2	Moderate	High	High
Row7	949	61-7-3844-6...	31 Duncan St. ...	South Bris...	4101	15	Calaghan	Tony	Large	59,469.12	120	Bin 1	Bin 1	Bin 2	Moderate	Very Low	Very Low
Row8	949	30.59.8555	67, avenue de ...	Versailles	78000	18	Tonini	Daniel	Large	64,834.32	234	Bin 1	Bin 1	Bin 4	Very Low	Very Low	Very Low
Row9	678	(1) 47.55.6555	25, rue Lauriston	Paris	75016	27	Perrier	Dominique	Medium	93,170.66	55	Bin 3	Bin 3	Bin 1	High	Moderate	Moderate
Row10	029	6175558428	16780 Pompto...	Brickhaven	58339	8	Taylor	Leslie	Medium	26,479.26	181	Bin 1	Bin 1	Bin 2	Moderate	Very Low	Very Low
Row11	678	07-98 9555	Erling Skakkes ...	Stavern	4110	32	Bergulfsen	Jonas	Medium	116,599.19	209	Bin 3	Bin 3	Bin 3	Low	Moderate	Moderate
Row12	662	+49 89 61 08 ...	Hansastr. 15	Munich	80686	14	Donnerme...	Michael	Medium	34,993.92	260	Bin 1	Bin 1	Bin 4	Very Low	Very Low	Very Low
Row13	099	+49 69 66 90 ...	Lyonerstr. 34	Frankfurt	60528	22	Keitel	Roland	Medium	85,171.59	209	Bin 2	Bin 2	Bin 3	Low	Low	Low
Row14	380	3105552373	4097 Douglas ...	Glendale	92561	3	Young	Leslie	Medium	9,129.35	114	Bin 1	Bin 1	Bin 2	Moderate	Very Low	Very Low
Row15	108	+34 913 728 ...	Merchants Hou...	Madrid	28023	13	Fernandez	Jesus	Large	49,642.05	440	Bin 1	Bin 1	Bin 4	Very Low	Very Low	Very Low
Row16	949	6175555555	4658 Baden Av.	Cambridge	51247	11	Tseng	Kyung	Medium	36,163.62	390	Bin 1	Bin 1	Bin 4	Very Low	Very Low	Very Low
Row17	949	(604) 555-3392	1900 Oak St.	Vancouver	V3F 2K1	22	Tannamuri	Yoshi	Large	75,238.92	223	Bin 2	Bin 2	Bin 3	Low	Low	Low
Row18	949	2155554695	782 First Street	Philadelphia	71270	21	Cervantes	Francisca	Medium	67,506.97	231	Bin 2	Bin 1	Bin 3	Low	Very Low	Low
Row19	949	2125558493	5905 Pompton...	NYC	10022	20	Hernandez	Maria	Medium	77,795.2	193	Bin 1	Bin 2	Bin 3	Low	Low	Very Low
Row20	108	+353 1862 15...	25 Maiden Lane	Dublin	2	16	Cassidy	Dean	Large	57,756.43	259	Bin 1	Bin 1	Bin 4	Very Low	Very Low	Very Low

04. 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

CUSTOMERNAME	RECENCY_HML	MONETARY_HML	FREQUENCY_HML
Euro Shopping Channel	High	High	High
Mini Gifts Distributors Ltd.	High	High	High
La Rochelle Gifts	High	High	High
Souveniers And Things Co.	High	High	High
Reims Collectables	High	High	High

The top 5 best customers are determined by calculating the RFM score, which is the sum of scores for Monetary, Recency, and Frequency. Higher RFM scores indicate more valuable customers.

Customers at Risk of Churning: Identifying Potential Churners

CUSTOMERNAME	RECENCY_HML	MONETARY_HML	FREQUENCY_HML
Saveley & Henriot, Co.	Very Low	High	High
Vida Sport, Ltd	Very Low	Moderate	Moderate
Herkku Gifts	Very Low	Moderate	Moderate
Marta's Replicas Co.	Very Low	Moderate	Moderate
AV Stores, Co.	Low	High	High

The customers with very low to low recency, high to moderate monetary value, and high to moderate purchase frequency are at an increased risk of churning.

Top 5 Lost customer

CUSTOMERNAME	RECENCY_HML	MONETARY_HML	FREQUENCY_HML
Double Decker Gift Stores, Ltd	Very Low	Very Low	Very Low
West Coast Collectables Co.	Very Low	Very Low	Very Low
Signal Collectibles Ltd.	Very Low	Very Low	Very Low
Daedalus Designs Imports	Very Low	Very Low	Very Low
CAF Imports	Very Low	Very Low	Very Low

- Customers with low frequency of purchases, indicating reduced engagement.
- Customers with low recency, implying prolonged inactivity.
- Customers with low monetary value, suggesting decreased spending.

Top 5 Loyal Customer

CUSTOMERNAME	RECENCY_HML	MONETARY_HML	FREQUENCY_HML
Handji Gifts& Co	High	Moderate	High
Reims Collectables	High	High	High
L'ordine Souveniers	High	High	High
Danish Wholesale Imports	High	High	High
Salzburg Collectables	High	High	High

Loyal customer refers to customers who are consistent and repeat purchases over time. They may not necessarily have the highest monetary value, but they demonstrate a strong commitment to the brand by repeatedly choosing to do business with the company.

05. RECOMMENDATIONS:

- Customer Retention Strategies: Engage and retain customers who are at risk of churning. Offer personalized incentives, loyalty programs, or exclusive promotions to encourage their continued loyalty

- Continuous Monitoring and Analysis: Regularly monitor customer behavior, sales trends, and key performance indicators to identify changing patterns and proactively address any potential issues.

- Upselling and Cross-Selling Opportunities: Identify opportunities to upsell and cross-sell to existing customers, especially those with a high monetary value. Recommend relevant products or services based on their purchase history to increase their average order value and enhance customer loyalty.


Thank You!