Marketing and Retail Analytics project

MILESTONE 1 – LAVANYA N RAO

Content

- Agenda & Executive Summary of the data
- Exploratory Analysis and Inferences
- Customer Segmentation using RFM analysis
- Univariate, Bivariate, and Multivariate Analysis Using Data Visualization
- Sales Across Different Categories of Different Features
- Output Table Head
- Inferences from RFM Analysis

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.

FACT CHECK ABOUT GIVEN DATA

```
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   datetime64[ns]
 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: datetime64[ns](1), float64(2), int64(5), object(12)
memory usage: 429.3+ KB
```

```
ORDERNUMBER           0
QUANTITYORDERED       0
PRICEEACH             0
ORDERLINENUMBER       0
SALES                 0
ORDERDATE             0
DAYS_SINCE_LASTORDER  0
STATUS                0
PRODUCTLINE           0
MSRP                  0
PRODUCTCODE           0
CUSTOMERNAME          0
PHONE                 0
ADDRESSLINE1          0
CITY                  0
POSTALCODE            0
COUNTRY               0
CONTACTLASTNAME       0
CONTACTFIRSTNAME      0
DEALSIZE              0
dtype: int64
```

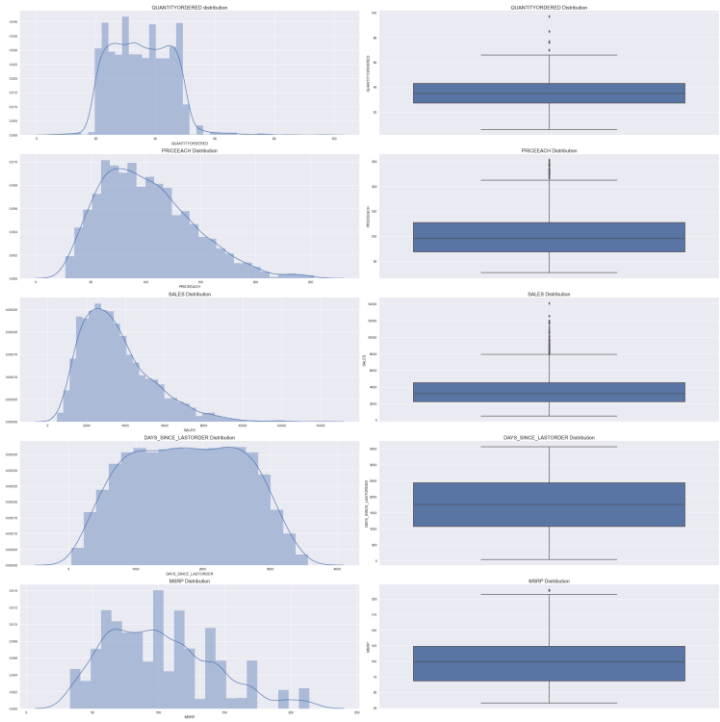
FACT CHECK ABOUT GIVEN DATA

- The dataset consists of 2747 rows and 20 columns.
- No null values are present within the dataset.
- The data types include datetime64 (1), Float64 (2), int64 (5), and Object (12).
- A data description is provided, indicating that the data appears to be well-structured.
- The dataset demonstrates consistency across the two iterations, indicating stability in the data.

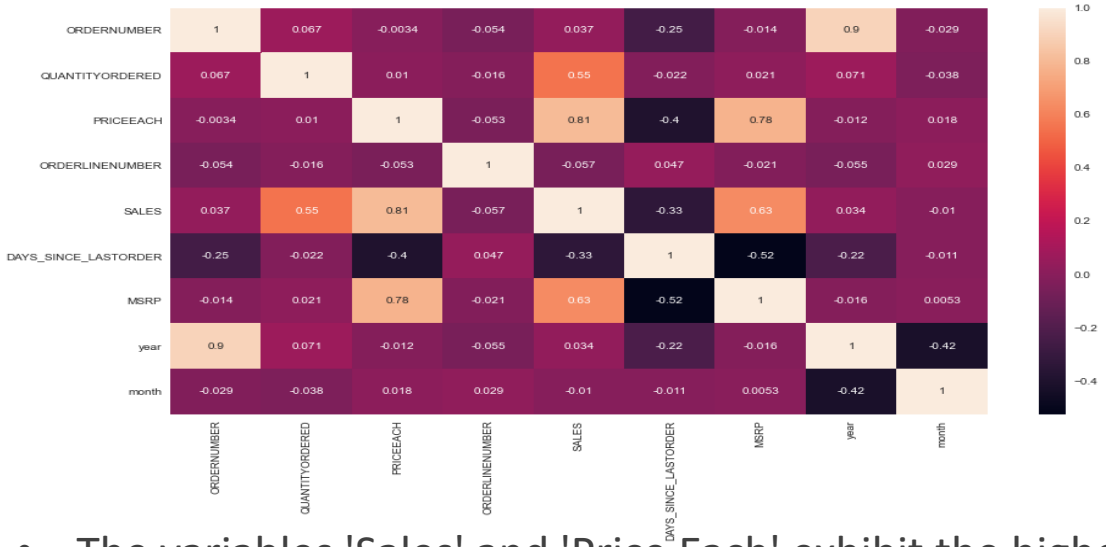
	count	mean	std	min	25%	50%	75%	max
ORDERNUMBER	2747.0	10259.761558	91.877521	10100.00	10181.000	10264.00	10334.500	10425.00
QUANTITYORDERED	2747.0	35.103021	9.762135	6.00	27.000	35.00	43.000	97.00
PRICEEACH	2747.0	101.098951	42.042548	26.88	68.745	95.55	127.100	252.87
ORDERLINENUMBER	2747.0	6.491081	4.230544	1.00	3.000	6.00	9.000	18.00
SALES	2747.0	3553.047583	1838.953901	482.13	2204.350	3184.80	4503.095	14082.80
DAYS_SINCE_LASTORDER	2747.0	1757.085912	819.280576	42.00	1077.000	1761.00	2436.500	3562.00
MSRP	2747.0	100.691664	40.114802	33.00	68.000	99.00	124.000	214.00

Distplot and Boxplot

- Outliers are observed in variables such as Quantity Ordered, Price, and Sales.
- Bivariate analysis using boxplots on Sales and Product Line variables reveals the presence of outliers in each product line category.
- Bivariate analysis using boxplots on Sales and Deal Size variables highlights the presence of outliers specifically in the Large deal size category.

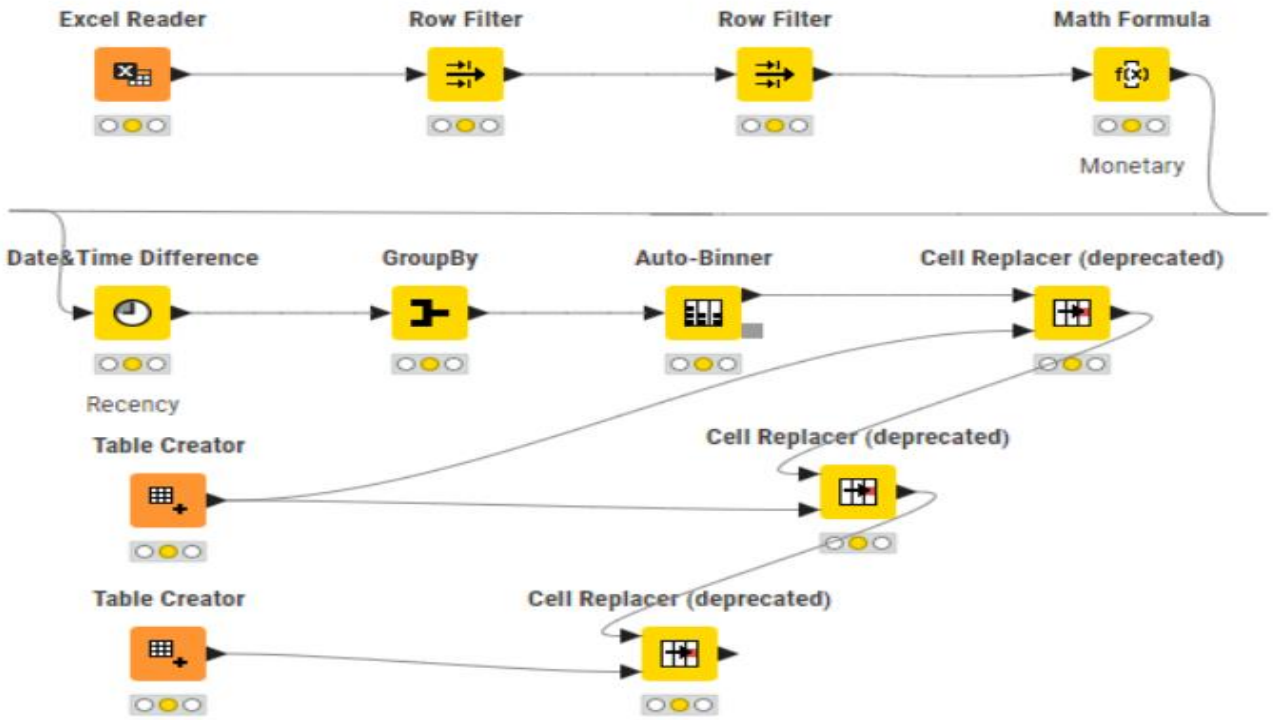


Multi Variate Analysis

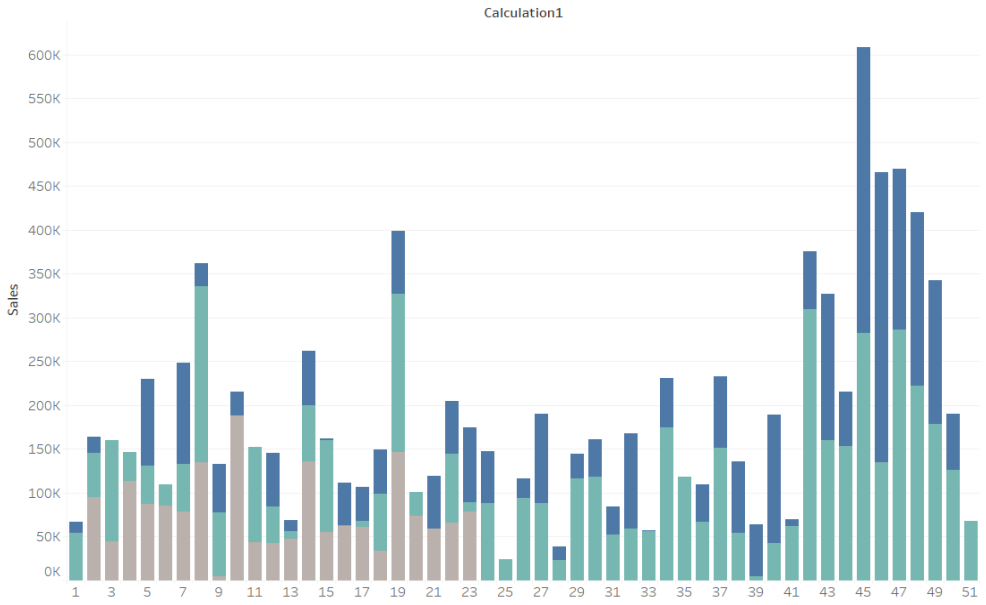


- The variables 'Sales' and 'Price Each' exhibit the highest positive correlation of 0.81, while 'Days_since_lastorder' and 'MSRP' demonstrate the highest negative correlation of -0.52.
- None of the variables display symmetric distribution as observed from the plot.
- A nearly linear relationship is evident between the variables 'Sales' and 'Price Each'.
- The plots indicate a wider spread of data along the trend line for MSRP compared to Price Each, suggesting the need to identify items with significant price changes to maximize sales.

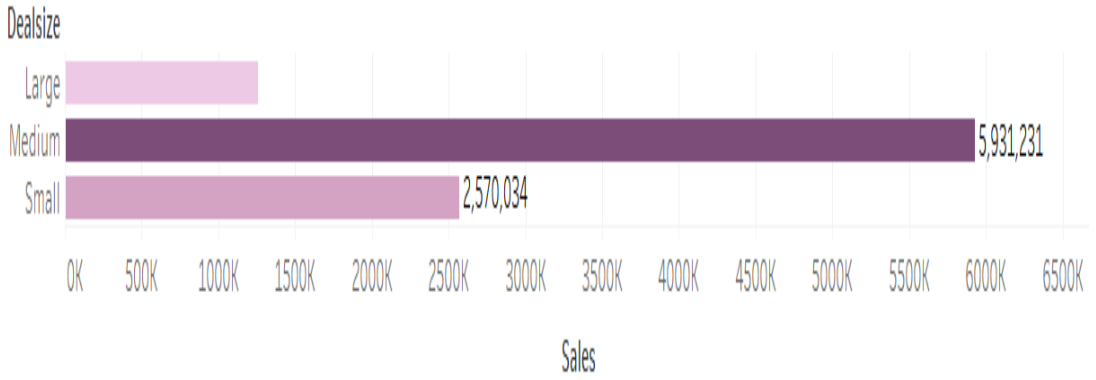
KNIME WORKFLOW and Output table -



WEEKLY TREND IN SALESVARIABLE

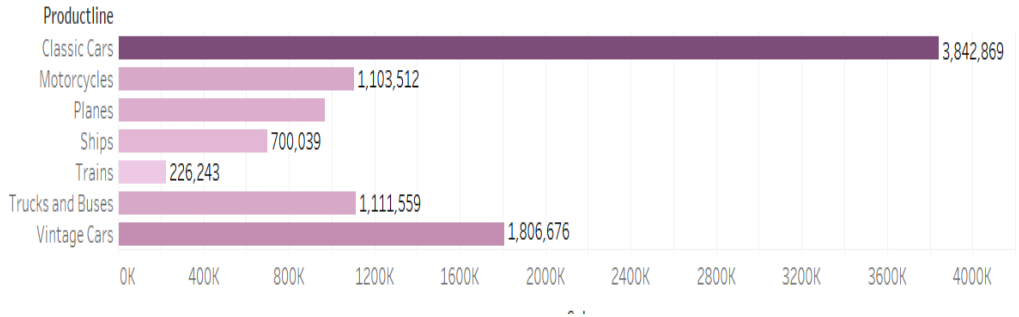


BIVARIATE ANALYSIS SALES ACROSS DIFFERENT CATEGORIES OF DIFFERENT FEATURES IN THE GIVEN DATA SALES VS DEALSIZE :



BIVARIATE ANALYSIS

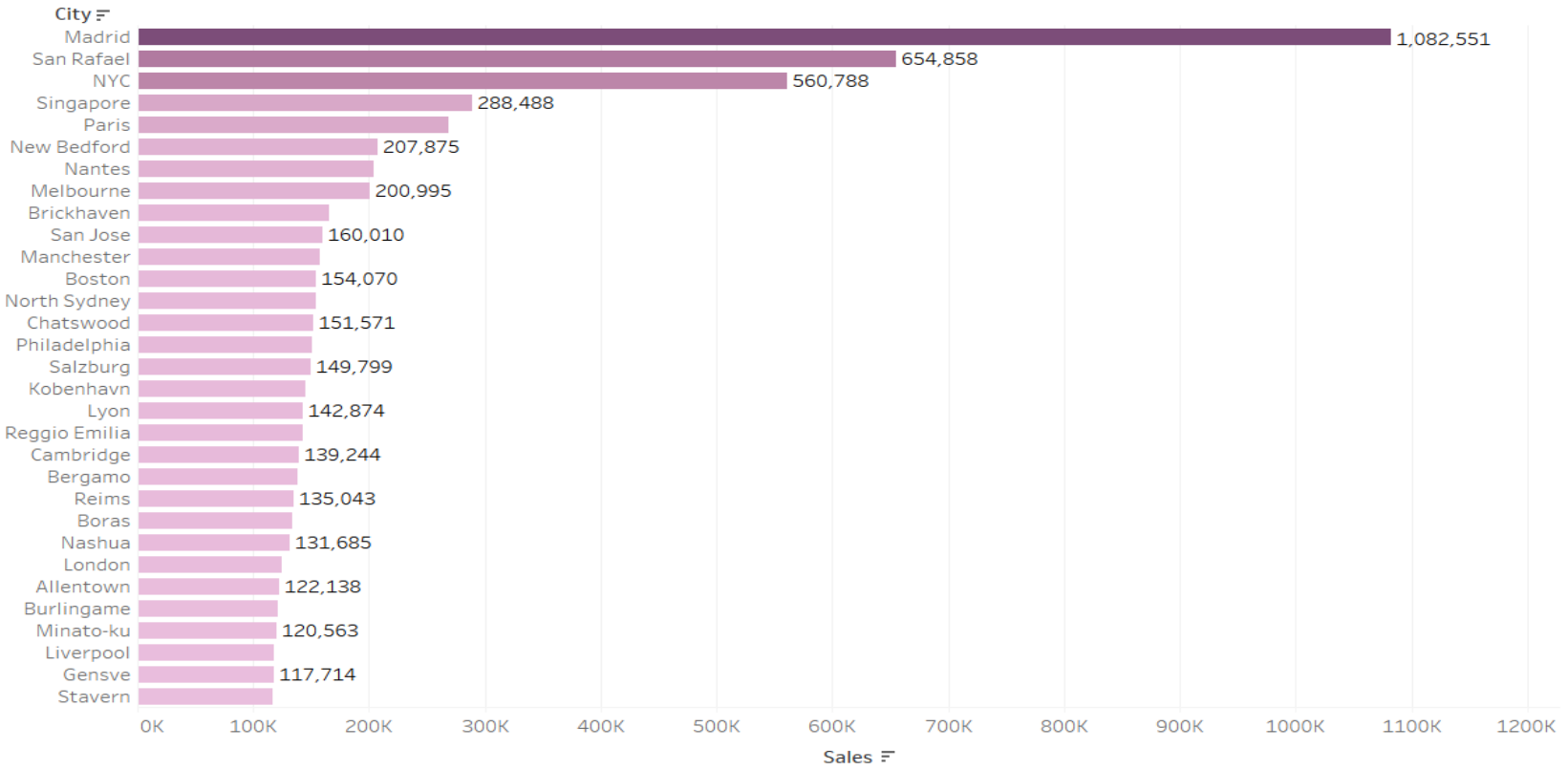
SALES ACROSS DIFFERENT CATEGORIES OF DIFFERENT FEATURES IN THE GIVEN DATA SALES VS PRODUCTLINE :



INFERENCE FROM THE ABOVE ANALYSIS

- Sales of product lines like Trucks/Buses and Motorcycles have reached saturation, indicating limited growth opportunities.
- Classic Cars outperform other car categories in terms of sales across product lines.
- Medium-sized deals constitute the majority of deals.
- Larger deals are relatively scarce compared to medium and smaller deals, suggesting a focus on expanding large-size projects.
- Sales of large-sized deals show minimal growth over time, indicating a potential need for the company to concentrate on securing larger-scale projects.

BIVARIATE ANALYSIS
SALES VS CITY



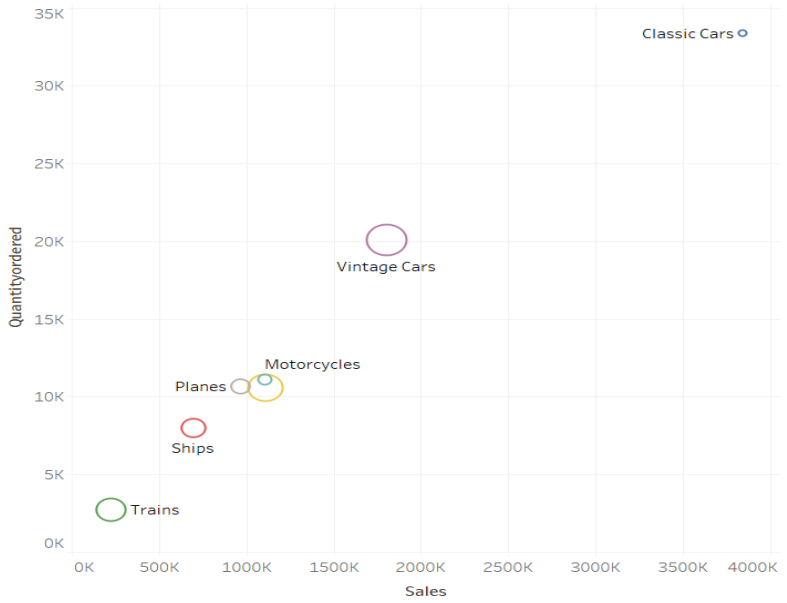
MULTIVARIATE ANALYSIS
NUMARICAL VS NUMARICAL

SALES VS MSRP :

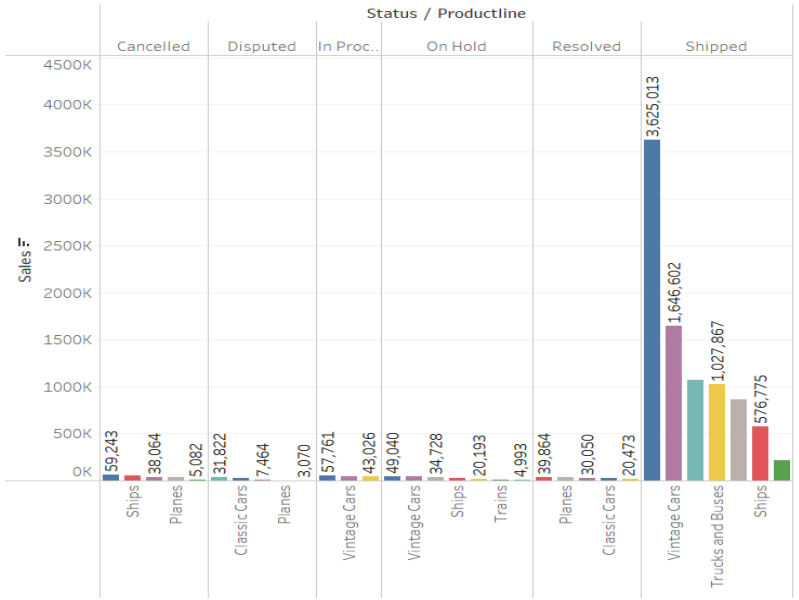
SALES VS PRICEEACH :

MULTIVARIATE ANALYSIS

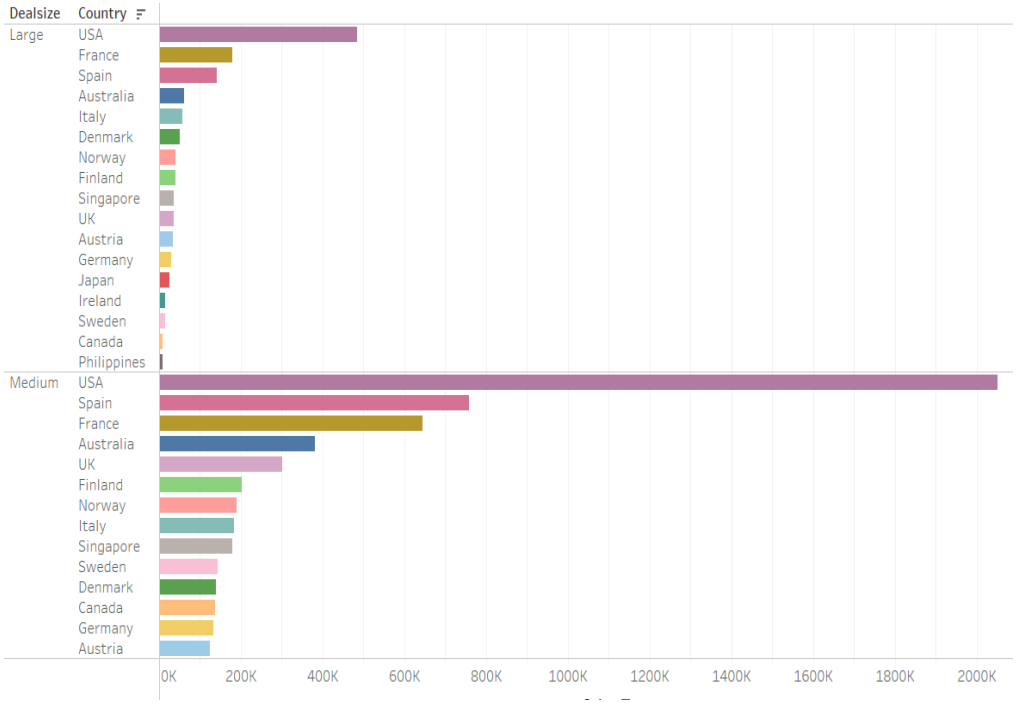
NUMARICAL VS NUMARICAL SALES VS QUANTITYORDERED :

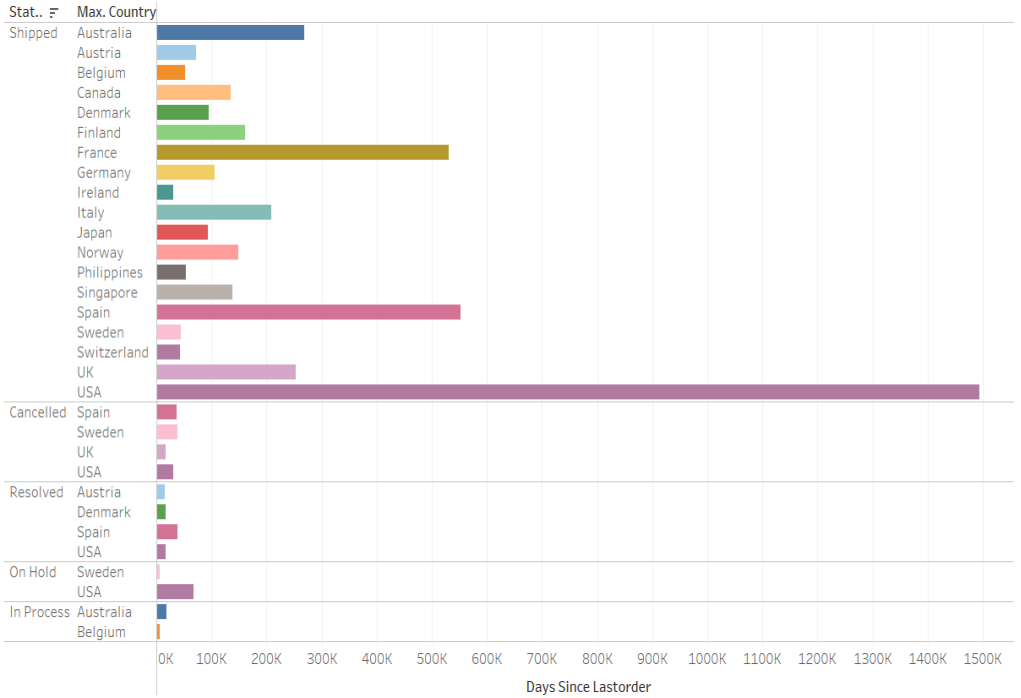


SALES ACROSS STATUS AND PRODUCTLINE



MULTIVARIATE ANALYSIS
SALES ACROSS DEALSIZE ACCORDING TO COUNTRY :





INFERENCE FROM THE ABOVE ANALYSIS:

- With a significant portion of sales originating from 4 – 5 key customers, the company's operations are heavily influenced by customer preferences. Therefore, a strategic focus on customer acquisition is crucial to mitigate the impact of potential client churn on overall sales.
- The company registers a higher volume of sales in the USA compared to other countries, indicating a strong market presence in its domestic market.
- Export revenue is primarily driven by large-sized deals, followed by medium and small-sized deals. This suggests that larger deals are predominantly sourced from international clients. However, there's a need for the company to also prioritize domestic sales alongside its export strategies.

Customer Segmentation Using RFM Analysis

Introduction:

- RFM Analysis: A sophisticated methodology for customer segmentation that evaluates Recency, Frequency, and Monetary Value metrics.
- Significance of Customer Segmentation: Crucial for targeted marketing, personalized experiences, and maximizing customer lifetime value.
- Approach Overview: Employing KNIME, a powerful data analytics platform, for comprehensive RFM Analysis.

Segmentation Methodology:

- RFM Analysis Workflow: Rigorous process encompassing data preparation, segmentation, and interpretation.
- Segmentation Categories: Customers categorized into four segments based on their RFM scores: Low, Medium, High, and Extremely High.
- Bin Allocation: Customers are assigned to bins based on their Recency, Frequency, and Monetary metrics, facilitating granular analysis.

Key Findings: Top Customers

- "Mini Gifts Distributors": Eminent presence with high scores across Recency, Frequency, and Monetary metrics, indicating strong engagement and significant contribution to revenue.
- "Muscle Machine Inc.": A notable customer with a balanced profile, displaying moderate scores across key metrics, particularly noteworthy for their high Recency.
- "Land of Toys Inc.": Another prominent customer demonstrating high Recency and moderate Frequency, highlighting consistent engagement and potential for further growth.

Key Findings: Lost Customers

- Identification of Lost Customers: Customers with low scores across Recency, Frequency, and Monetary metrics, indicating disengagement or attrition risk.
- Strategies for Re-engagement: Tailored initiatives required to rekindle relationships and regain lost revenue opportunities, focusing on personalized communication and incentives.

Key Findings: Loyal Customers

- Recognition of Loyal Customer Segments: Customers exhibiting high scores in Recency, Frequency, and Monetary metrics, reflecting strong brand loyalty and long-term value.
- Loyalty Enhancement Strategies: Strategies aimed at nurturing and retaining loyal customers, such as VIP programs, exclusive offers, and personalized services.

Inferences from RFM Analysis:

- Segmentation Insights: Identification of distinct customer segments, including Champions, Potential Loyalists, New Customers, At Risk Customers, and Can't Lose Them Customers.
- Strategic Implications: Tailored strategies and initiatives recommended for each segment to optimize marketing, sales, and customer retention efforts.

Conclusion::

- Strategic Insights: RFM Analysis provides actionable insights for refining marketing strategies, enhancing customer engagement, and driving revenue growth.
- Business Impact: Implementation of RFM-driven strategies can lead to improved customer satisfaction, increased retention rates, and sustainable business growth.

Jupyter Notebook-
Attaching the code book - MRA Milestone 1_Project_Lavanya N Rao.ipynb

Tableau Workbook-
MRA | Tableau Public

KNIME Workbook-
Attaching with the PPT Submission along with the output excel file-
Lavanya RFM.knwf