Business report on MRA

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.

Data info:

```
<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   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
```

Data shape, head and tail with summary.

```
In [6]:  df.shape
Out[6]: (2747, 20)
```

```
In [7]:  df.head()
Out[7]:
```

	ORDERNUMBER	QUANTITYORDERED	PRICEEACH	ORDERLINENUMBER	SALES	ORDERD
0	10107	30	95.70	2	2871.00	2018-(
1	10121	34	81.35	5	2765.90	2018-(
2	10134	41	94.74	2	3884.34	2018-(
3	10145	45	83.26	6	3746.70	2018-(
4	10168	36	96.66	1	3479.76	2018-1

```
In [8]:  df.tail()
Out[8]:
```

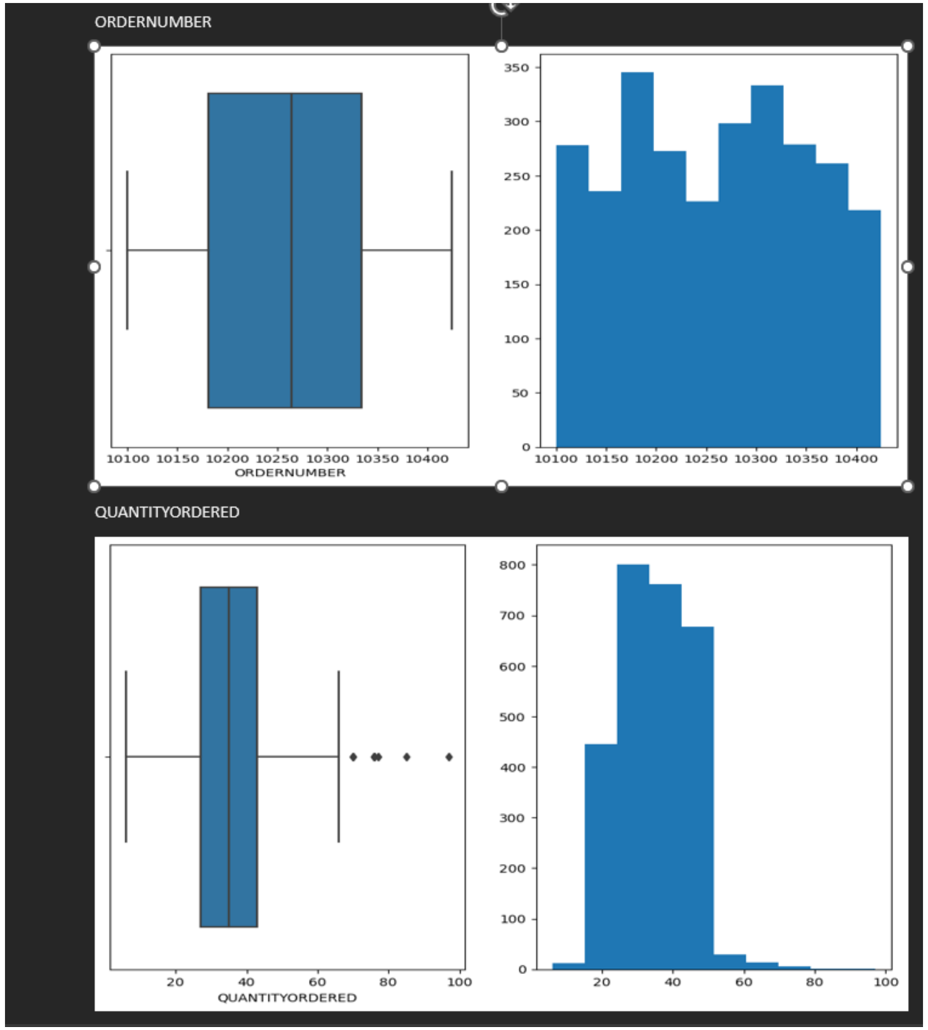
	ORDERNUMBER	QUANTITYORDERED	PRICEEACH	ORDERLINENUMBER	SALES	ORDI
2742	10350	20	112.22	15	2244.40	20
2743	10373	29	137.19	1	3978.51	20:
2744	10386	43	125.99	4	5417.57	20:
2745	10397	34	62.24	1	2116.16	20:
2746	10414	47	65.52	9	3079.44	20:

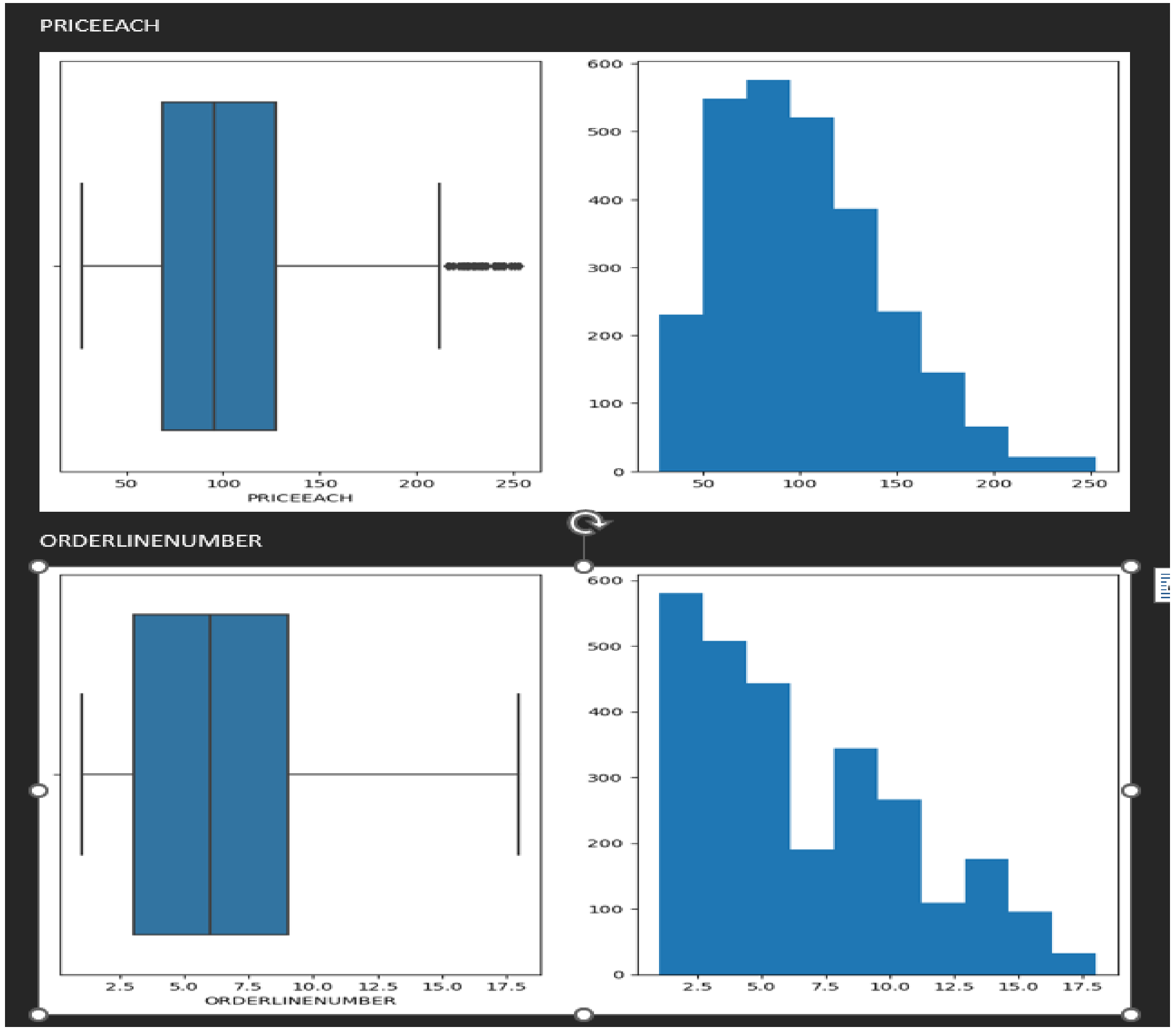
	count	mean	std	min	25%	50%	75%	max
ORDERNUMBER	2747.0	10259.761558	91.877521	10100.00	10181.000	10264.00	10334.500	10425.00
QUANTITYORDERED	2747.0	35.103021	9.762135	6.00	27.000	35.00	43.000	97.00
PRICEEACH	2747.0	101.098951	42.042548	26.88	68.745	95.55	127.100	252.87
ORDERLINENUMBER	2747.0	6.491081	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

Observation:

1) The Data has 2747 rows with 20 columns.
2) The quantity ordered is min 6 and max 97.
3) Sales are max with 14082.

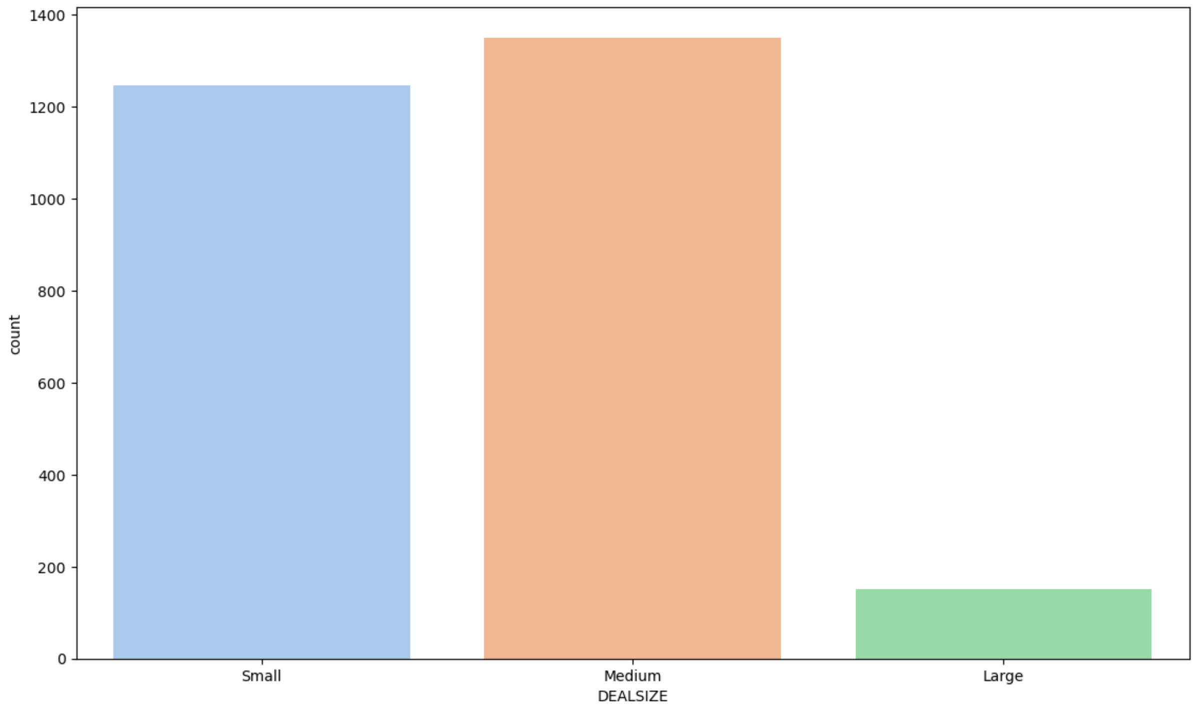
Univariant analysis



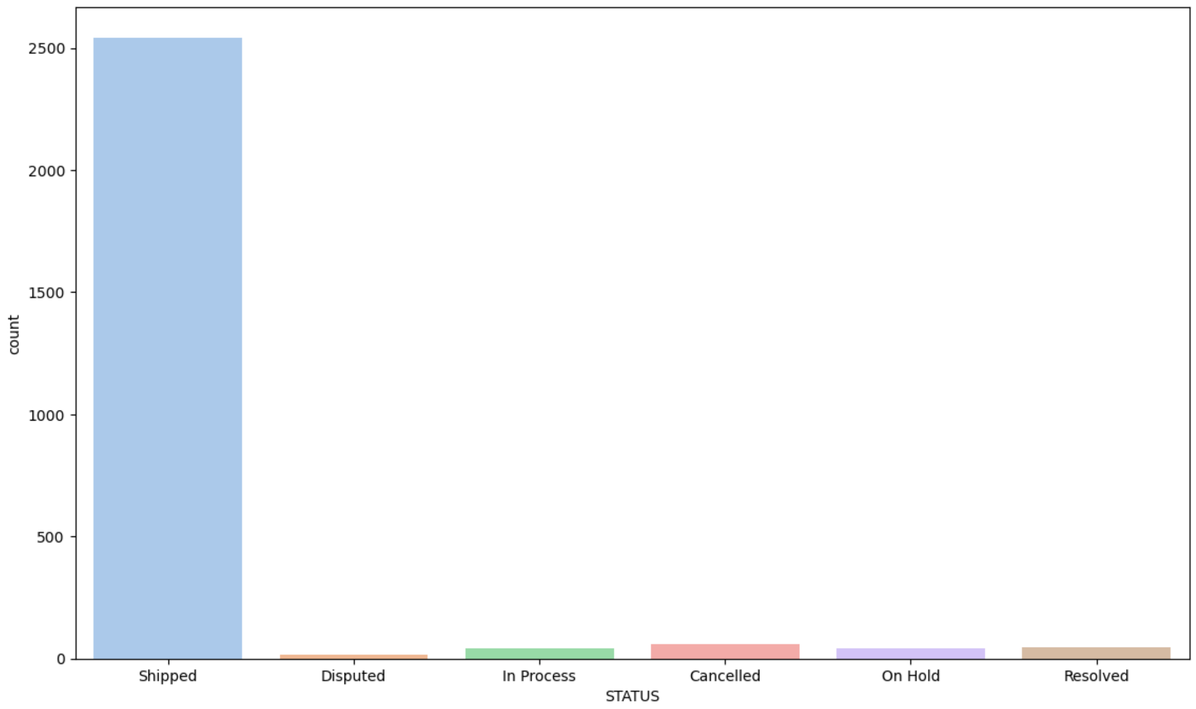


Univariant analysis on categorical.

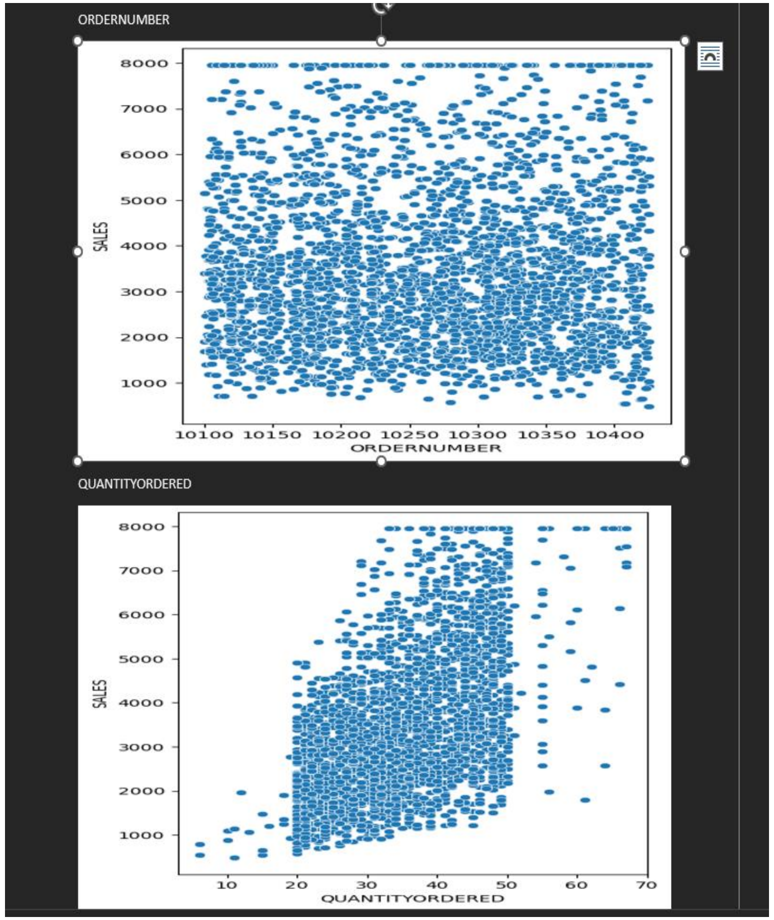
DEALSIZE

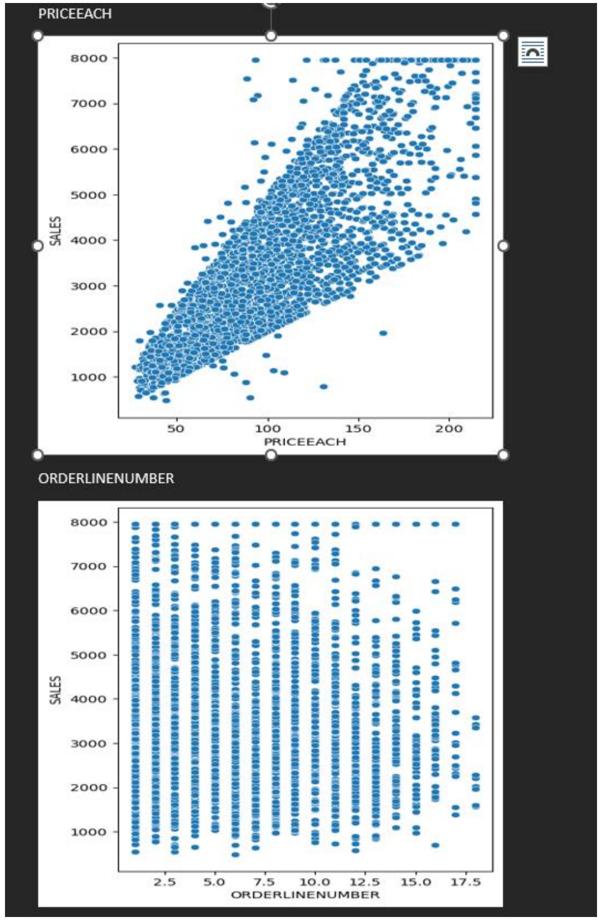


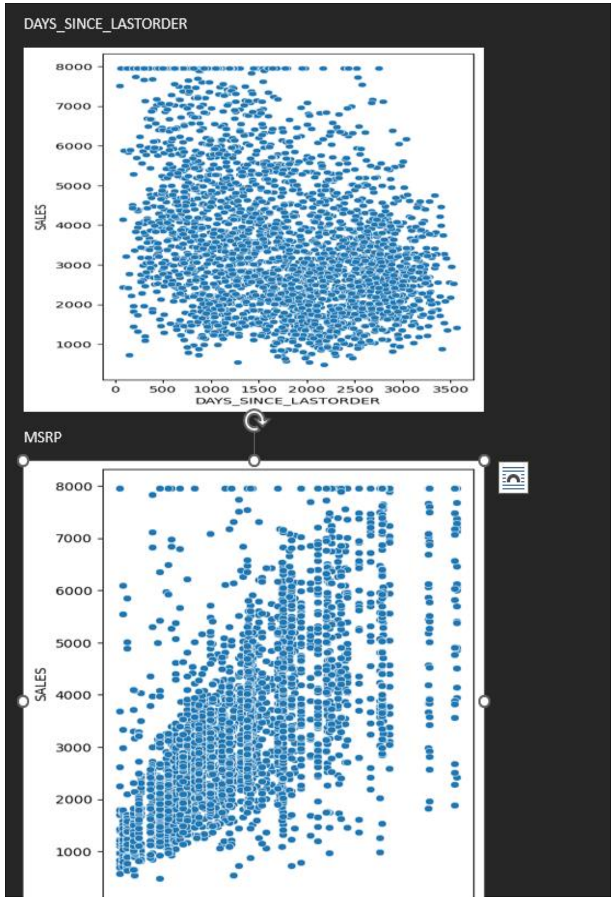
STATUS



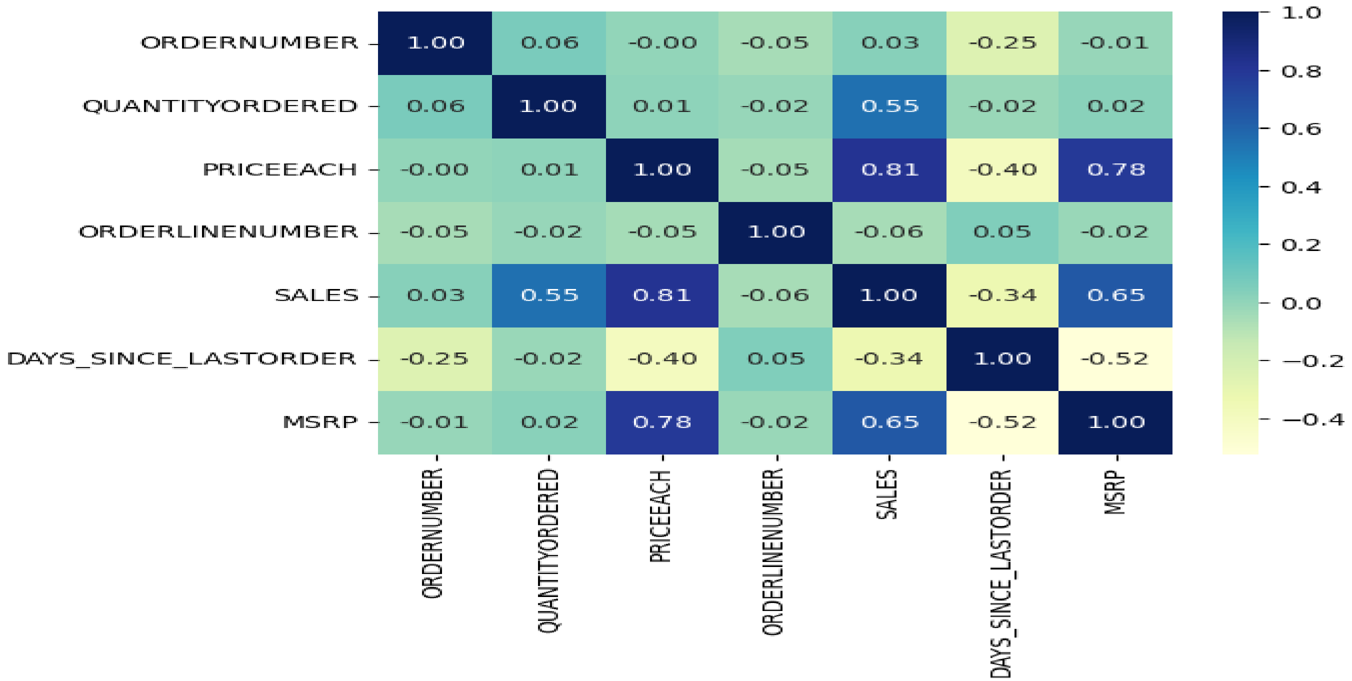
Bivariant analysis



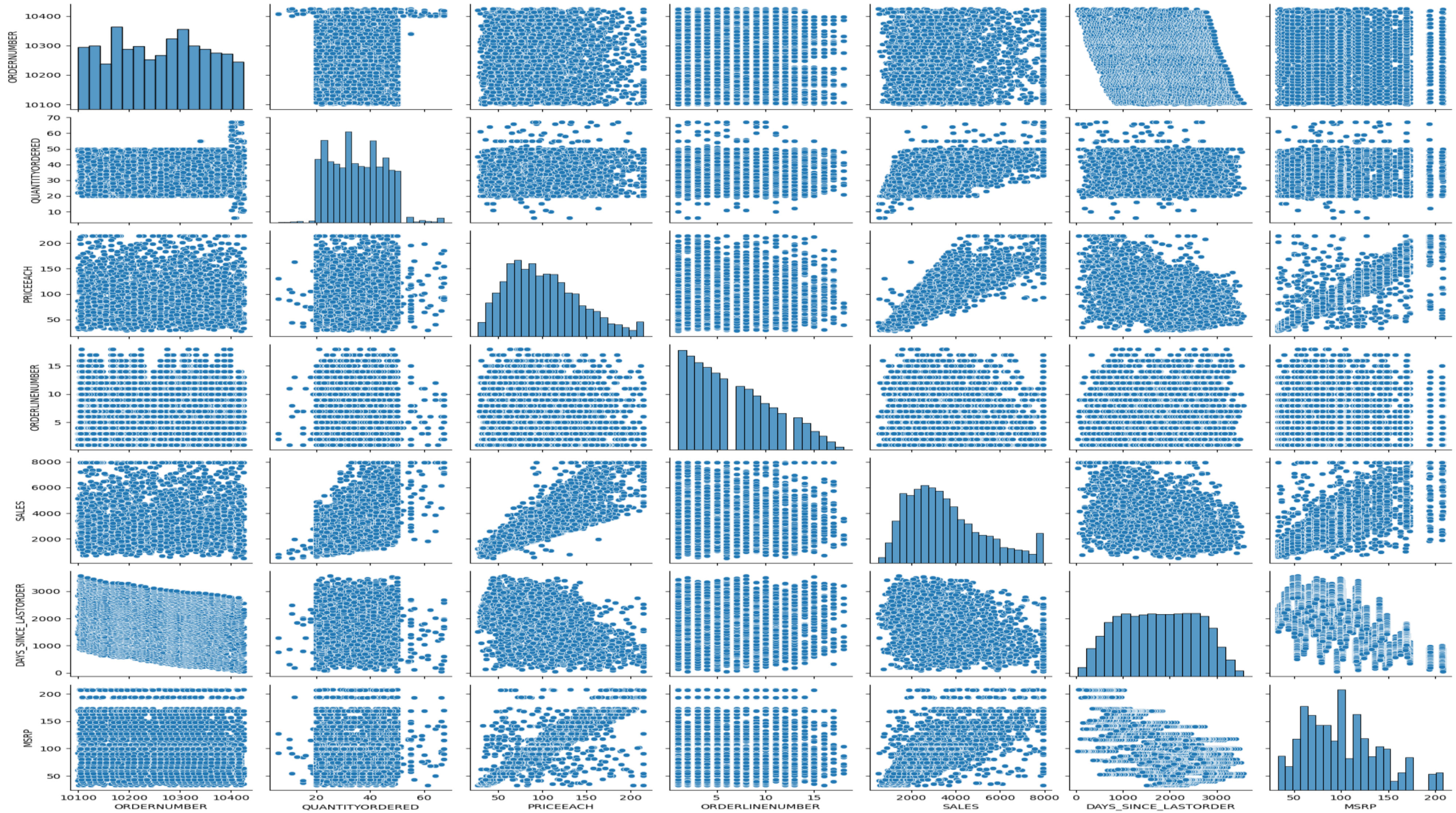




Relationship between all numeric variables.

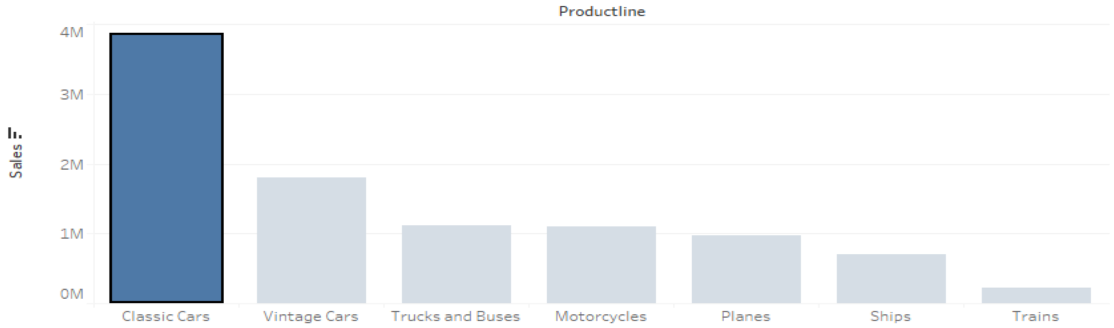


Multivariant analysis

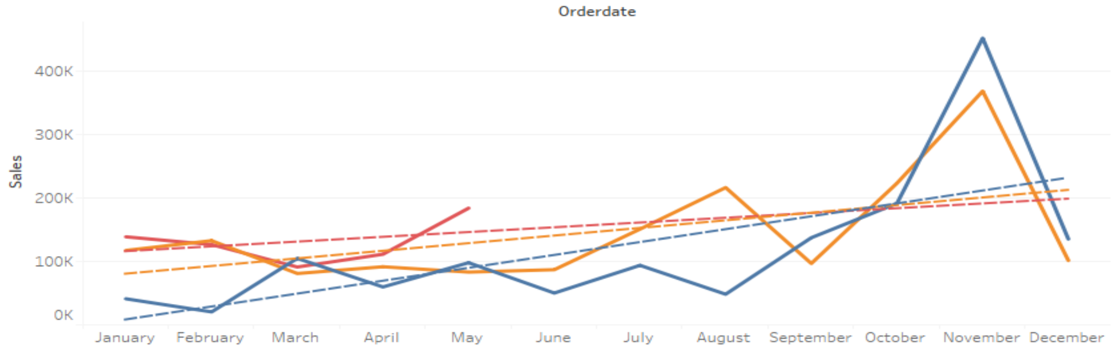


Dashboard representation of most sold automobile across different year and country.

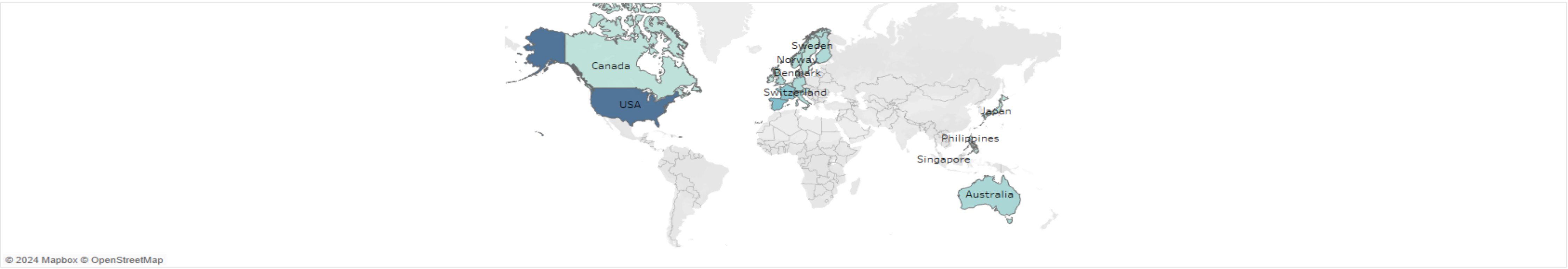
Sales vs product
Classic cars have more sales than Trains. i.e means usage of cars are more when compared to other automobiles.



Sales trend
The sales are more for 2018,2019 and peak during November december. However the sales for 2020 is yet unkonw due to lack of data.



Sales across countries
USA have high market for automobiles. The sales are top in US.



Inference from EDA

1) The sales depend on the price of each product. Its positively correlated.
2) Quantity of products ordered is making positive trends showing right direction on sales.
3) The medium size products deals are more in count.
4) The status of product shipped influence more on sales.
5) The sale has max count of 14802.
6) Days since last ordered is 3553 days ago. With min days of 48.
7) The classic cars have most sales across USA.
8) The highest sales comes from November and December holiday season.
9) Train automobile has low sales. Due to usage purpose.

RFM

RFM (Recency, Frequency, Monetary) analysis is a customer segmentation technique commonly used in marketing to evaluate customer behavior and segment customers based on their purchasing patterns.

RFM

- **Recency (R):** How recently a customer made a purchase.
  - Assumption: Customers who purchased more recently are more likely to return.
- **Frequency (F):** How often a customer makes purchases.
  - Assumption: Customers who purchase more frequently are more engaged.
- **Monetary (M):** How much money a customer spends.
  - Assumption: Higher spending indicates more valuable customers.

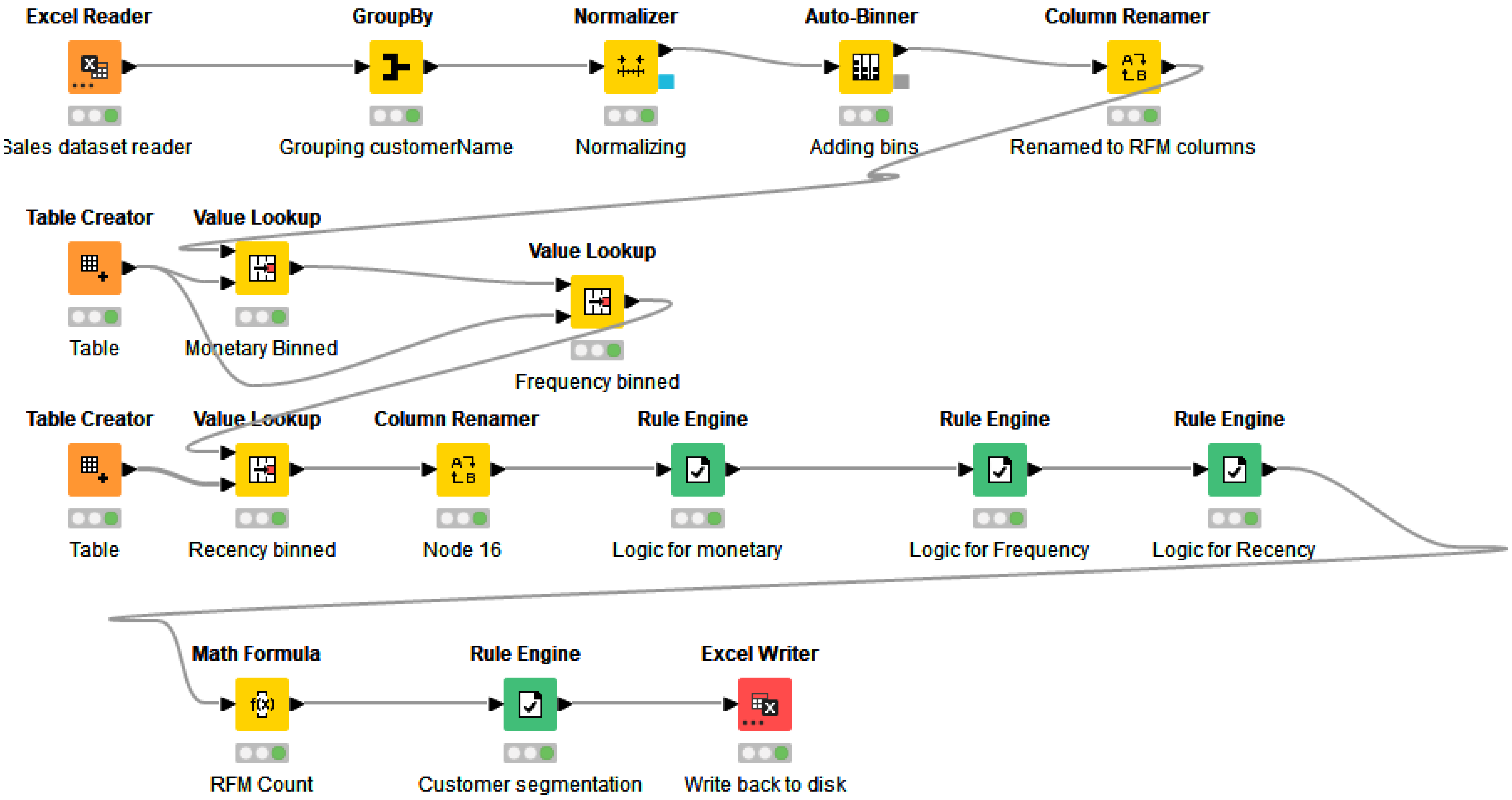
Parameters Used in RFM Analysis

1. **Recency:**.Parameter: Transaction Date and a reference date (e.g., today's date).
2. **Frequency:** Parameter: Customer ID and Transaction Count.
3. **Monetary:**Parameter: Customer ID and Transaction Amount.

Assumptions Made

1. **Customer Value Patterns:** Assumes customer behavior follows the recency, frequency, and monetary patterns.
2. **Data Cleanliness:** Transaction data is accurate and complete.
3. **Segmentation:** Divides customers into 4 segments based on RFM scores:
   1. High Recency, High Frequency, High Monetary (Top Customers).
   2. High Recency, Low Frequency/Monetary (New Customers).
   3. Low Recency, High Frequency/Monetary (Loyal Customers).
   4. Low Recency, Low Frequency/Monetary (Churned Customers).

RFM workflow image



Inferences from RFM Analysis and identified segments

1. Customer Segmentation Breakdown

•Customers were categorized into **High-Value**, **Medium-Value**, and **Low-Value** segments based on their Recency, Frequency, and Monetary scores.
•**Key Observations**:
    •**High-Value Customers:** These customers are actively purchasing frequently, spending high amounts, and engaging recently. They represent a small but critical portion of the customer base and contribute significantly to revenue.
    •**Medium-Value Customers:** Customers with moderate spending and engagement. They may be influenced to increase their purchases with targeted marketing strategies.
    •**Low-Value Customers:** Customers with infrequent purchases, low spending, and poor recent engagement. They are either at risk of churning or have already disengaged.

2. Insights on Customer Behavior

•**Best Customers** are those in the **High-Value** segment. They exhibit loyal purchasing behavior and drive the most business value.
•**Customers on the Verge of Churning**: Customers in the Medium-Value segment with declining frequency and higher recency indicate they are losing interest.
•**Lost Customers** represent those who have stopped engaging altogether. **Loyal Customers** often overlap with best customers but include those who buy frequently and consistently, even if their monetary contribution is moderate.

3. Strategic Recommendations

For Best Customers:

Reward loyalty with exclusive offers, discounts, or membership programs to retain them.
Personalize their experiences to ensure continued engagement.

For Customers on the Verge of Churning:

Send re-engagement campaigns like discounts, personalized emails, or limited-time offers.
Understand potential reasons for disengagement and address them.

For Lost Customers:

Use win-back campaigns to re-engage them with compelling offers or reminders.
Analyze patterns to prevent similar customers from becoming lost.

For Loyal Customers:

Encourage referrals through reward programs.
Offer upselling or cross-selling opportunities to maximize their contribution.

4. Overall Impact of RFM Analysis

- This analysis helps prioritize marketing and customer relationship efforts effectively.
- By focusing on high-value customers and re-engaging potential churners, businesses can improve revenue and customer retention.

Interpretation of RFM Analysis

1. Best Customers
•Criteria:
- **Monetary = H (3)**
- **Frequency = H (3) or M(2)**
- **Recency = H(3)** (indicating recent engagement).
- **Segmentation:** Likely classified as **High-Value Customers**

TOP 5 customers: RFM 333 or 323

Best 5 – CUSTOMERNAME(RFM score 333)
Anna's Decorations, Ltd
Australian Collectors, Co.
Land of Toys Inc.
Euro Shopping Channel
Diecast Classics Inc. – RFM(323)

2. Customers on the Verge of Churning
•Criteria:
- **Monetary = L (1)** or **M (2)**
- **Frequency = L (1)**
- **Recency = H (1 or 2)** .
- **Segmentation:** Likely classified as **Medium-Value Customers** or **Low-Value Customers**.

Top 5 Customers on the Verge of Churning:
Customers with RFM_Score like 232 233 or 222.

CUSTOMERNAME
AV Stores, Co.
Amica Models & Co.
Australian Gift Network, Co
Auto Canal Petit
Baane Mini Imports

3. Lost Customers

•Criteria:

- **Monetary = L (1)**
- **Frequency = L (1)**
- **Recency = H (1)**
- **Segmentation:** Likely classified as **Low-Value Customers**.

Top 5 Lost Customers:

1.Customers with RFM_Score like 211 or 111.

CUSTOMERNAME
Alpha Cognac
Atelier graphique
Australian Collectables, Ltd
Auto Assoc. & Cie.
Auto-Moto Classics Inc.

4. Loyal Customers

•Criteria:

- **Monetary = H (3)** or **M (2)**
- **Frequency = H (3) or M(2)**
- **Recency = H (1)** or **M (2)** (indicating consistent engagement).
- **Segmentation:** Often overlaps with **High-Value Customers** or **Medium-Value Customers**.

Top 5 Loyal Customers:

1.Customers with RFM_Score like 322

CUSTOMERNAME
AV Stores, Co.
Amica Models & Co.
Anna's Decorations, Ltd
Australian Collectors, Co.
Australian Gift Network, Co

Problem Statement:

A grocery store shared the transactional data with you. Your job is to conduct a thorough analysis of Point of Sale (POS) data, identify the most commonly occurring sets of items in the customer orders, and provide recommendations through which a grocery store can increase its revenue by popular combo offers & discounts for customers.

Data info:

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 20641 entries, 0 to 20640
Data columns (total 3 columns):
 #   Column    Non-Null Count  Dtype
---  ------    --------------  -----
 0   Date      20641 non-null  object
 1   Order_id  20641 non-null  int64
 2   Product   20641 non-null  object
dtypes: int64(1), object(2)
memory usage: 483.9+ KB
```

Data shape, head and tail with summary.

```
(20641, 3)
```

```
df.head()
```

	Date	Order_id	Product
0	01-01-2018	1	yogurt
1	01-01-2018	1	pork
2	01-01-2018	1	sandwich bags
3	01-01-2018	1	lunch meat
4	01-01-2018	1	all- purpose

```
df.tail()
```

	Date	Order_id	Product
20636	25-02-2020	1138	soda
20637	25-02-2020	1138	paper towels
20638	26-02-2020	1139	soda
20639	26-02-2020	1139	laundry detergent
20640	26-02-2020	1139	shampoo

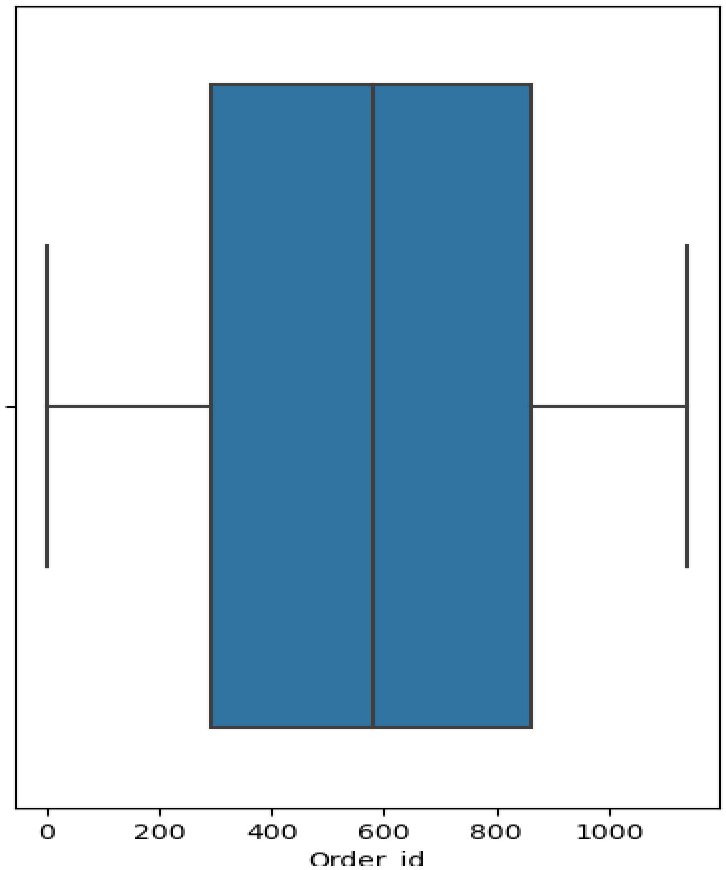
	count	mean	std	min	25%	50%	75%	max
Order_id	20641.0	575.986289	328.557078	1.0	292.0	581.0	862.0	1139.0

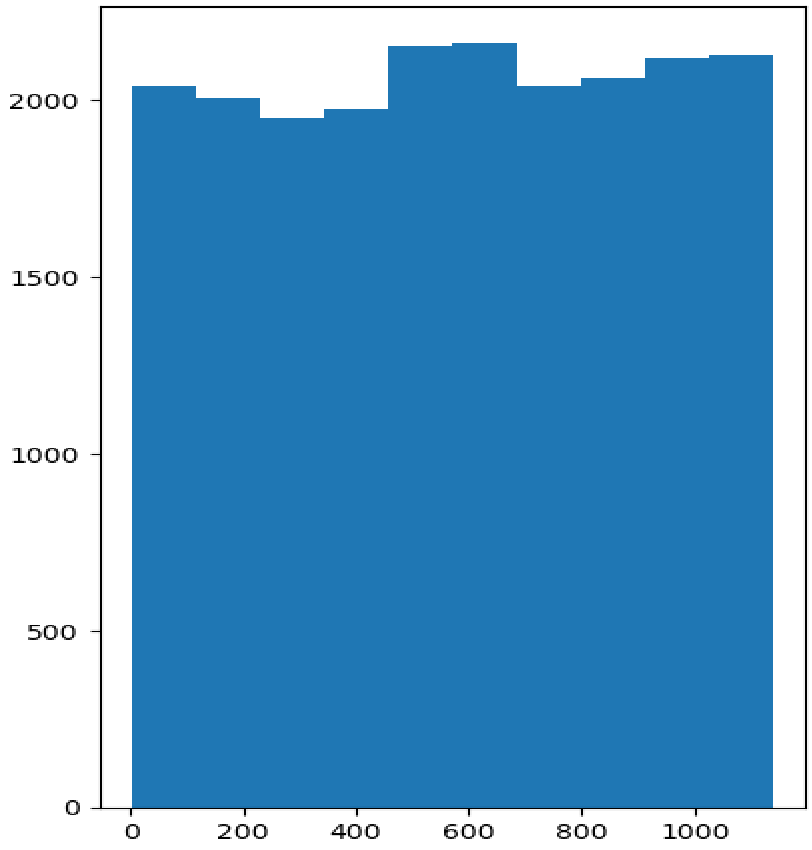
Observation:

1) The Data has 20641 rows with 3 columns.
2) There are wide variety of products.

Univariant analysis

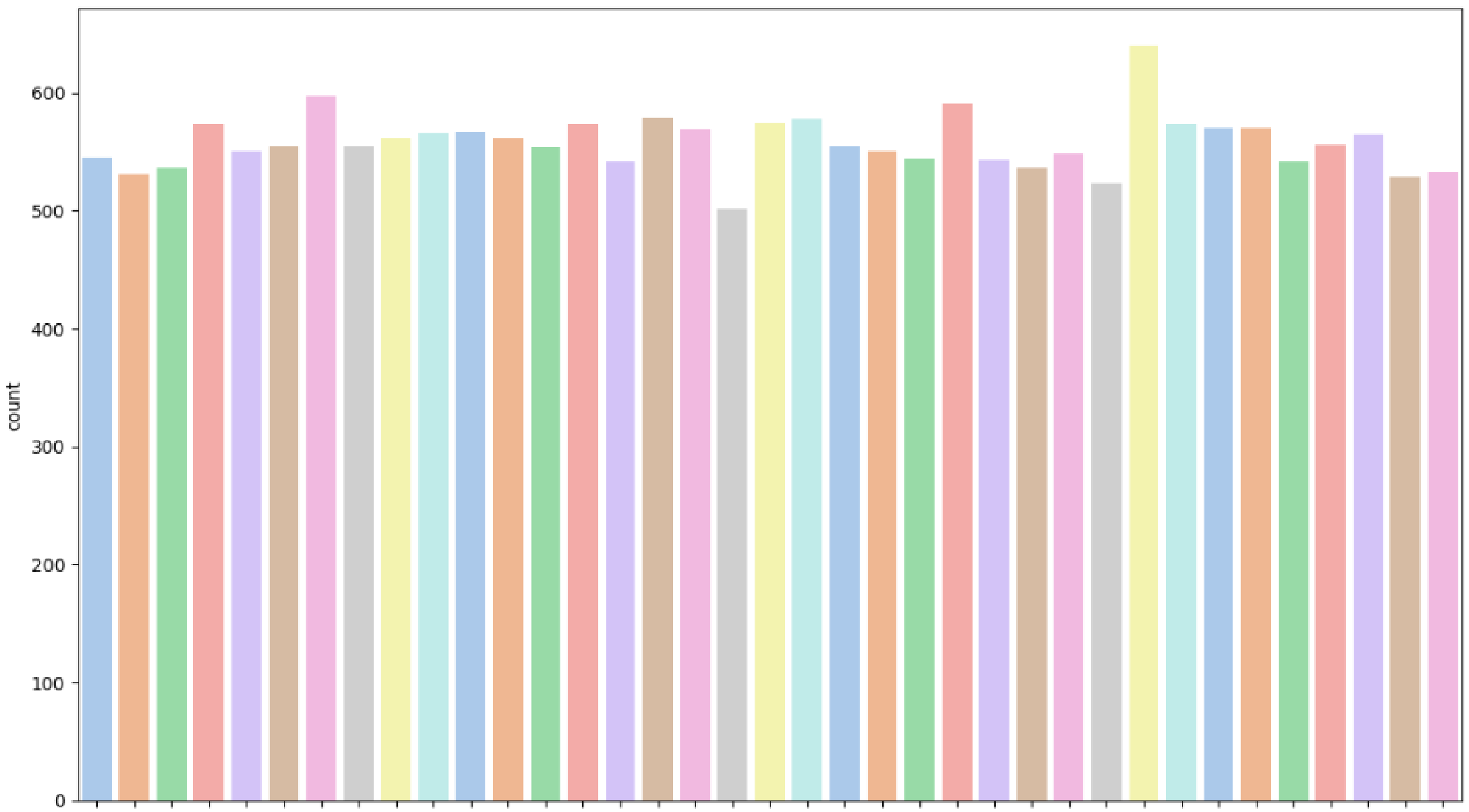
Order_id



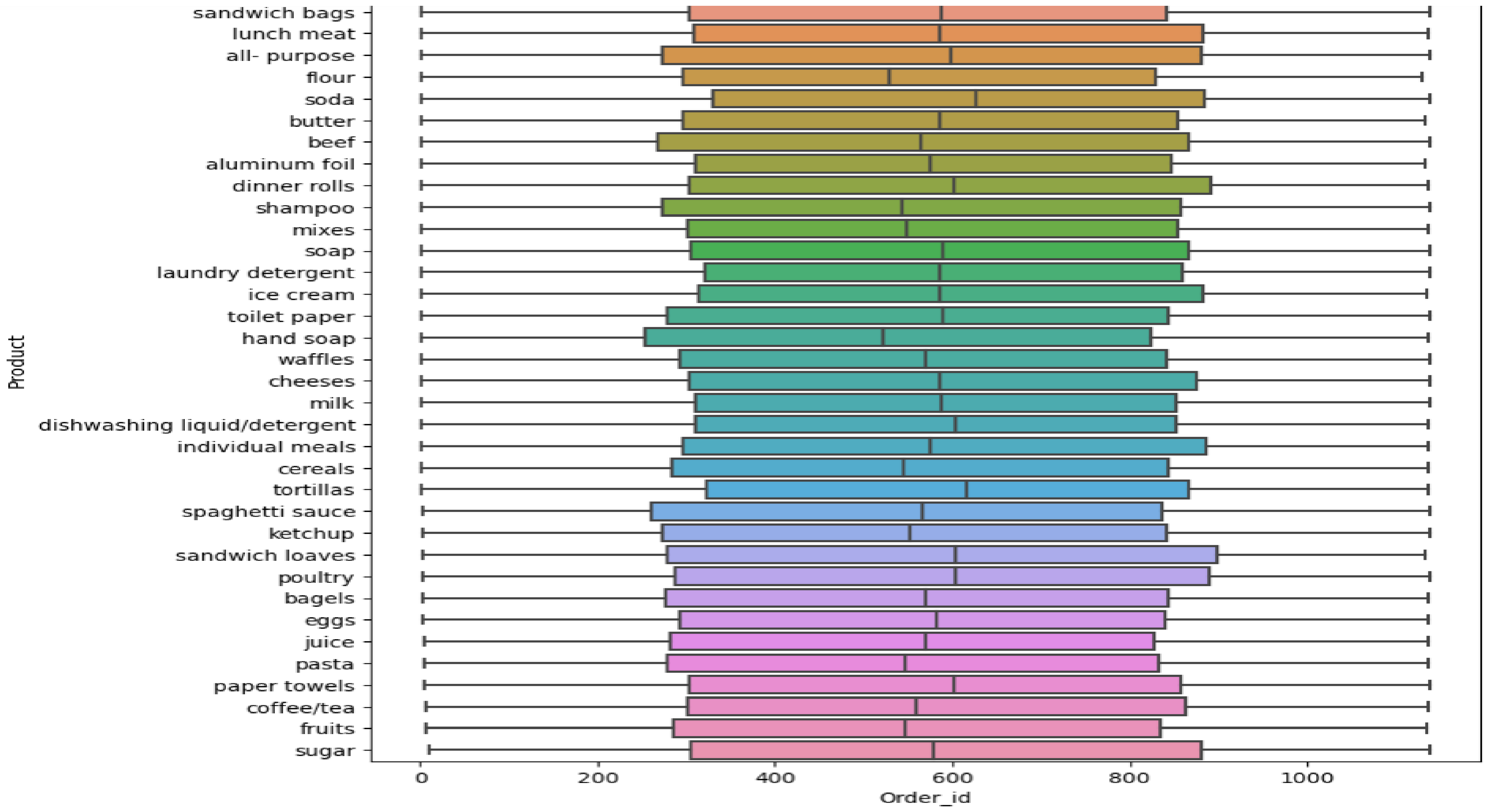


Univariant analysis on categorical.

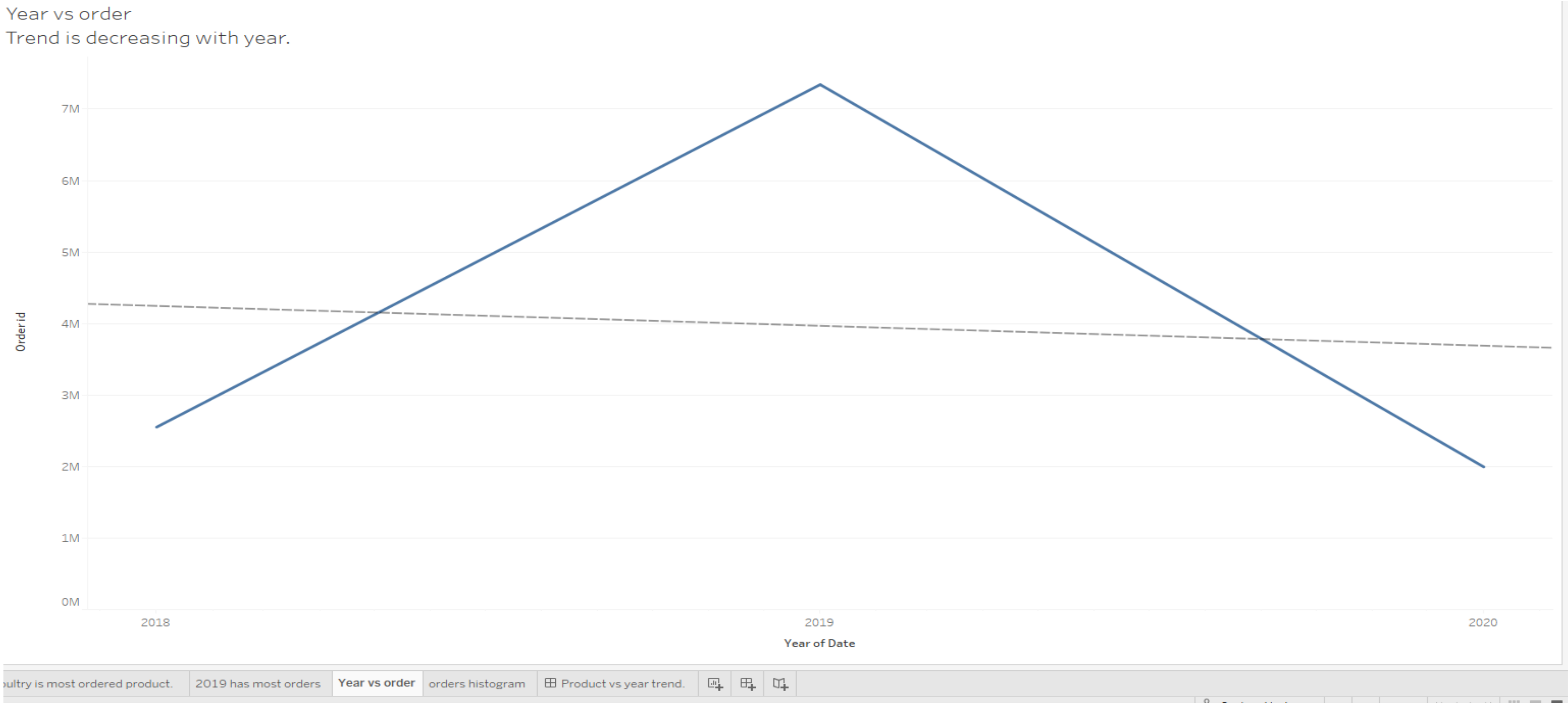
Product



Bivariant analysis

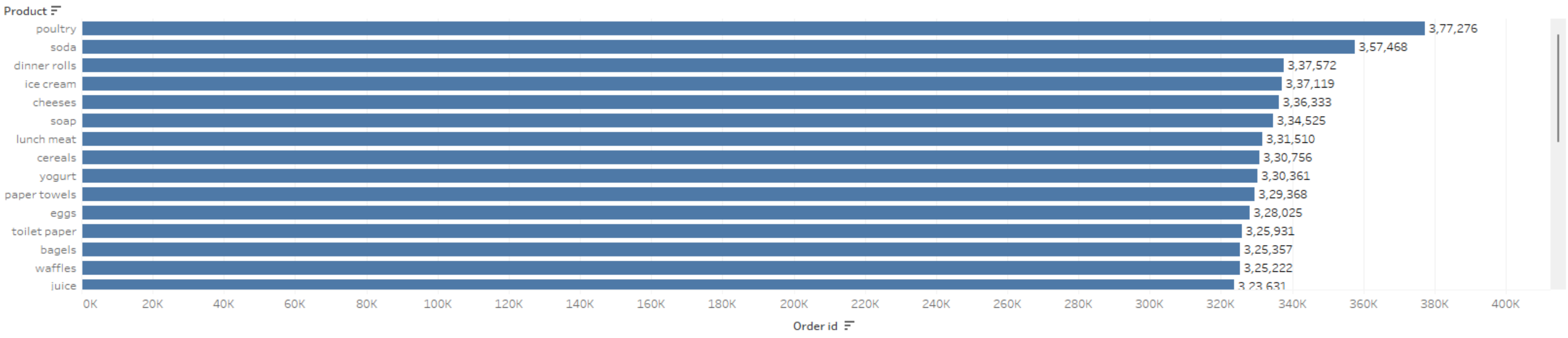


The Purchase of product is peak in 2019 year but slightly decreased in 2018. But the trend line shows more decrease in 2020.

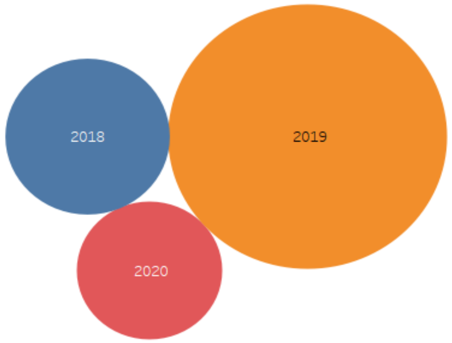


Most ordered product over the months and year.

Poultry is most ordered product.

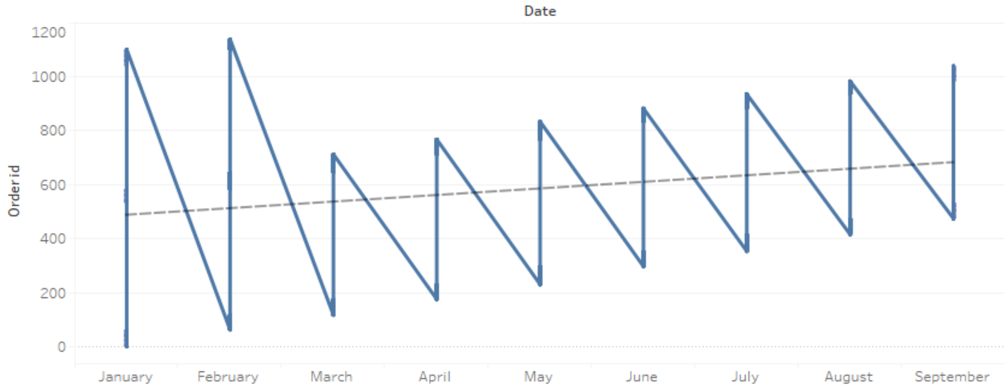


2019 has most orders



Orders in the month

Constant increase in the trend of shopping.



Inference on EDA

1) The Shopping for daily essentials keep increasing over the period of years.
2) The Poultry essentials are most bought product.
3) Soda consummation from few years have increased tremendously.
4) There are some seasonal trends in buying few of the household items like spaghetti and waffles.
5) 2019 has most orders when compared to other years. May be we need more promotions or discount to address this.

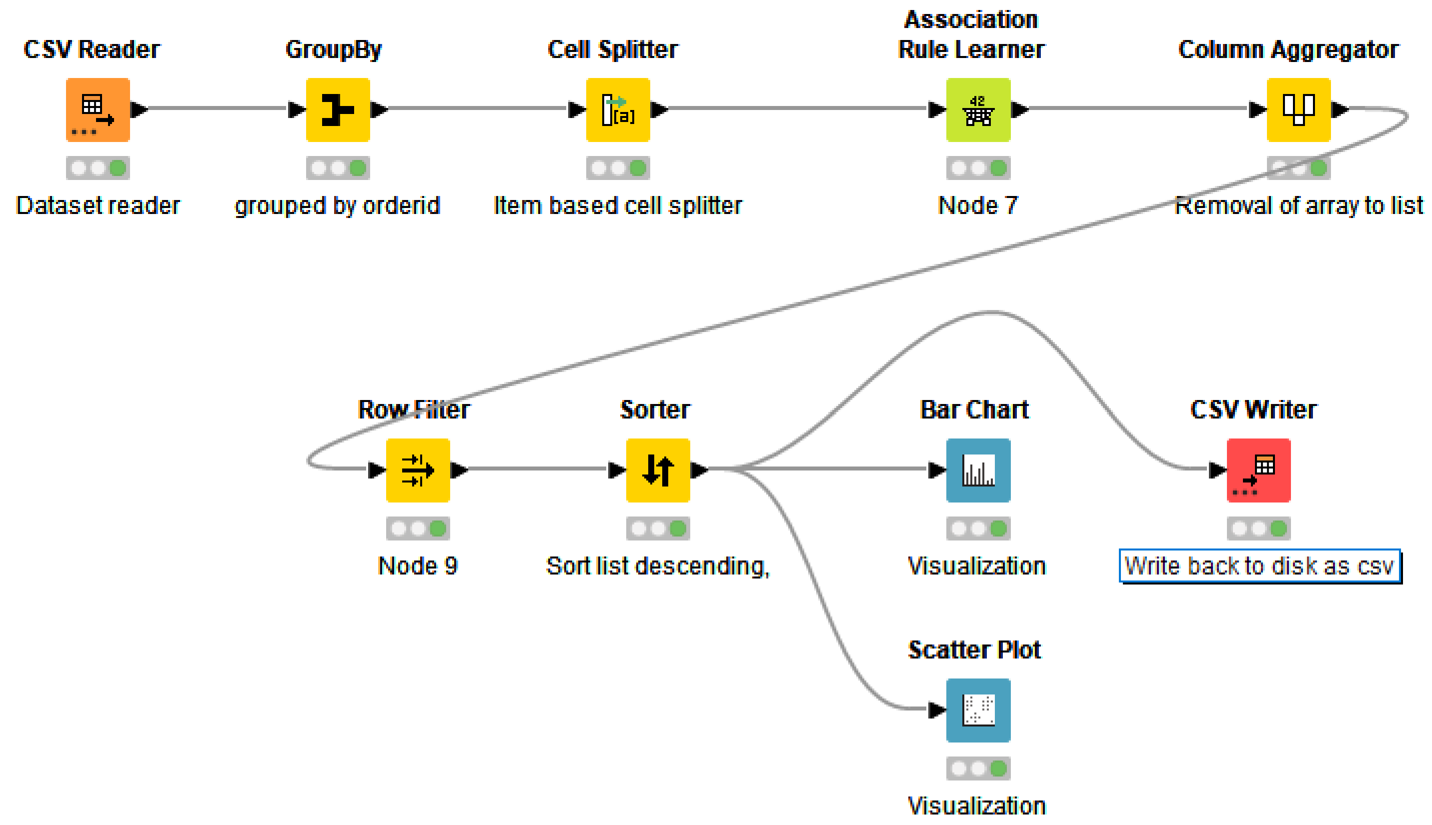
MBA

Market Basket Analysis (MBA) is a data mining technique used to discover patterns of product co-purchases by analyzing transactional data. It employs **association rules** to identify relationships between items that customers frequently buy together

Association Rules:

1. Support:
The proportion of transactions in which a particular itemset appears. It indicates how popular an itemset is in the dataset.
2. Confidence:
The likelihood of purchasing an item given that another item is already in the basket.
3. Lift:
Measures the strength of an association compared to random chance. A lift > 1 indicates a strong association.

MBA Workflow image



Threshold Values of Support and Confidence

Threshold Values of Support and Confidence:

1. **Support Threshold**:
   - Minimum Support helps filter out infrequent combinations.
   - Typical threshold: 0.01 to 0.05 (1%–5%) depending on dataset size and business objectives.

   In this analysis, the lowest Support is 0.039 – 0.3, indicating rules generated focus on frequent combinations.
2. **Confidence Threshold**:
   - Minimum Confidence ensures that only strong and reliable rules are generated.

   In this analysis, Confidence is from 0.7-1.0, which is the maximum, signifying that the consequents are always present with the antecedents.

Inference on analysed dataset.

Cross-Selling Opportunities
•Observation: Products like *soap* and *towels* appear with unrelated items (e.g., *beef* or *shampoo*).
•Inference: These could indicate surprising cross-selling potential, worth testing with combo offers.

Basket Size Insights
•Observation: Rules with high confidence but low lift suggest that items are often bought together but are already common in most baskets.
•Inference: Focus on upselling higher-margin items or increasing basket size with offers like "Spend X and get a discount.“

Customer Preferences
•Observation: High confidence in rules involving *yogurt, mixes, and bags*.
•Inference: Customers prefer bundling convenience items like mixes with reusable bags, indicating a preference for easy, sustainable shopping experiences.

Seasonal Trends
•Observation: If certain products (e.g., *spaghetti, sauce*) show spikes in sales during specific periods.
•Inference: These could indicate seasonal buying trends, which can be used for timely promotions or stock adjustments.

Popular Categories
•Observation: High confidence and lift for products like *beef, butter, paper* in various combinations.
•Inference: Food items and daily essentials dominate sales, highlighting their importance in driving revenue.

Revenue Trends

•**Growth or Decline**: The **SALES** trends across time periods to identify growth or areas requiring improvement.

Increasing sales in a quarter indicates effective strategies, while a decline may signal competition or unmet customer needs. This we could see in 2020 year.

Top-Selling Products

•Observation: Items like *towels, paper, and shampoo* appear frequently in the association rules.

•Inference: These items are likely high-demand products and can be prioritized for stock replenishment or promotional campaigns.

Suggested Combos with Lucrative Offers:

Combo 1: Towels + Paper + Shampoo Offer: Buy any two (towels, paper, shampoo) and get the third item free.
•**Rationale**: High lift and confidence suggest these items are strongly associated. Bundling them can encourage bulk purchases and improve customer satisfaction.

Combo 2: Rolls + Spaghetti + Sauce Offer: 20% discount on the bundle (all three items together).
•**Rationale**: These items are likely bought together for cooking purposes. Offering discounts on the bundle can increase the average transaction value.

Combo 3: Milk + Eggs + Soap Offer: Buy 2 items and get 50% off on the third.
•**Rationale**: These are daily essentials, and offering an attractive discount could make the store a preferred choice for customers' regular purchases.

Combo 4: Yogurt + Mixes + Bags Offer: Buy one "Yogurt and Mixes" pack and get a reusable shopping bag free.
•**Rationale**: This encourages sustainable practices and makes the purchase more appealing, particularly for eco-conscious customers.

Combo 5: **Towels + Beef + Butter Offer**: Flat 15% off when purchased together.

Rationale: Towels and beef might seem unusual together, but the data shows a strong association. Highlighting this bundle in a unique way can drive curiosity and sales.

Campaign Ideas:

1. “**Home Essentials Pack**”:
    1. Items: Milk, eggs, paper, soap, towels.
    2. Offer: Buy all 5 items and save 25%.
    3. Rationale: Appeals to families purchasing everyday necessities.

2. “**Pasta Night Special**”:
    1. Items: Spaghetti, sauce, flour.
    2. Offer: Buy 2 and get a free recipe book or seasoning pack.
    3. Rationale: Encourages purchases for a themed meal, increasing basket size.

3. **Loyalty Points Booster**:
    1. Double loyalty points on combos featuring frequently associated items like shampoo, paper, towels.
    2. Rationale: Enhances customer retention and rewards frequent buyers.