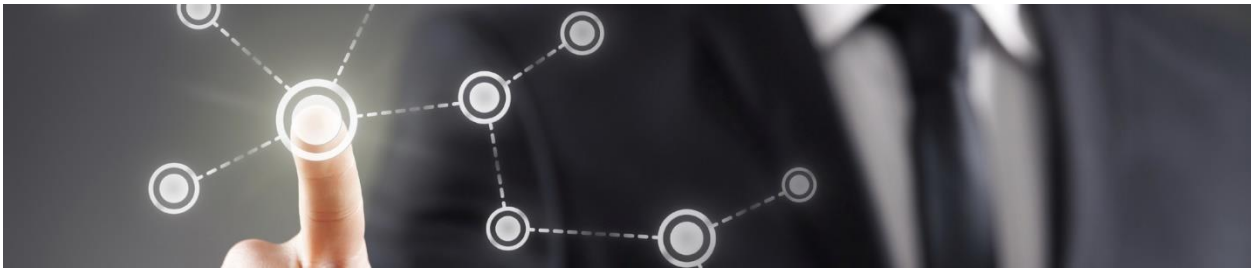
Retail Sales Analysis



Context

The datasets are from a unnamed business' online retail sales. Unfortunately, they have had to cut back from 25 employees to 5 because of the economic downturn with the corona virus. Lets help them out with some informative kernels and visualizations! Anything and everything helps!

Data Overview

Dataset contains 2 table **Transaction** and **Monthly Total Sales** from 2017 to 2029

Transaction contains 6 columns and 1775 rows:

- Product Type: Type/Category of the product
- Net Quantity: Number of products sold in 1 transaction
- Gross Sales: total revenue generated from selling the products before any deductions such as discounts or returns
- Discounts: total amount of discounts applied to the gross sales
- Returns: the quantity of products that were returned by customers after purchase
- Total Net Sales: the final amount of money earned after accounting for discounts and returns

	Product Type	Net Quantity	Gross Sales	Discounts	Returns	Total Net Sales
0	Art & Sculpture	34	14935.0	-594.00	-1609.0	12732.00
1	Basket	13	3744.0	-316.80	0.0	3427.20
2	Basket	12	3825.0	-201.60	-288.0	3335.40
3	Basket	17	3035.0	-63.25	0.0	2971.75
4	Art & Sculpture	47	2696.8	-44.16	0.0	2652.64

Figure 1. Sample of Transaction dataset

Monthly Total Sales contains 9 columns and 36 rows, each row represents 1 month:

- Month: Order month
- Year: Order year
- Total Orders: Total Orders each month
- Gross Sales: Gross Sales each month
- Discounts: Discounts each month
- Returns: Returns each month
- Net Sales: Net Sales each month
- Shipping: Shipping each month
- Total Sales: Total Sales by month

	Month	Year	Total Orders	Gross Sales	Discounts	Returns	Net Sales	Shipping	Total Sales
0	January	2017	73	8861.5	-129.40	-448.45	8283.65	1088.30	9371.95
1	February	2017	56	6908.5	-104.70	-416.20	6387.60	892.45	7280.05
2	March	2017	60	5778.5	-172.20	-1017.20	4589.10	707.43	5296.53
3	April	2017	70	8814.0	-281.40	0.00	8532.60	1068.30	9600.90
4	May	2017	54	6677.0	-185.75	-253.80	6237.45	866.46	7103.91

Figure 2. Sample of Monthly Sales dataset

Data Preprocessing

Before jumping to the analysis process, we shoud do data preprocessing, including cleaning and tranformation steps. This dataset is amost clean so I just need to remove 8 rows contain NaN value and 1 row with negative Net Quantity then change the value of Discounts and Returns to absolute value. After preprocessing step, the Transaction dataset has 1666 rows remaining.

Monthly Total Sales Analysis

What months does the business get the most sales?

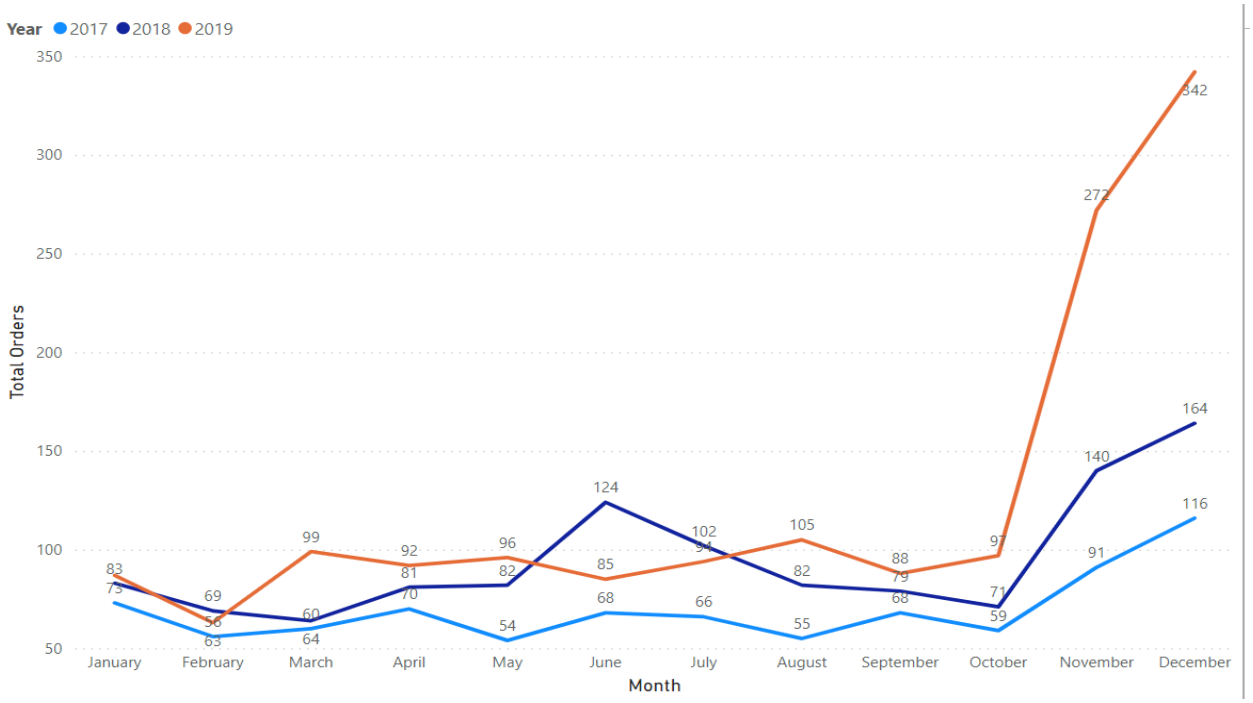


Figure 3. Total Orders by Month

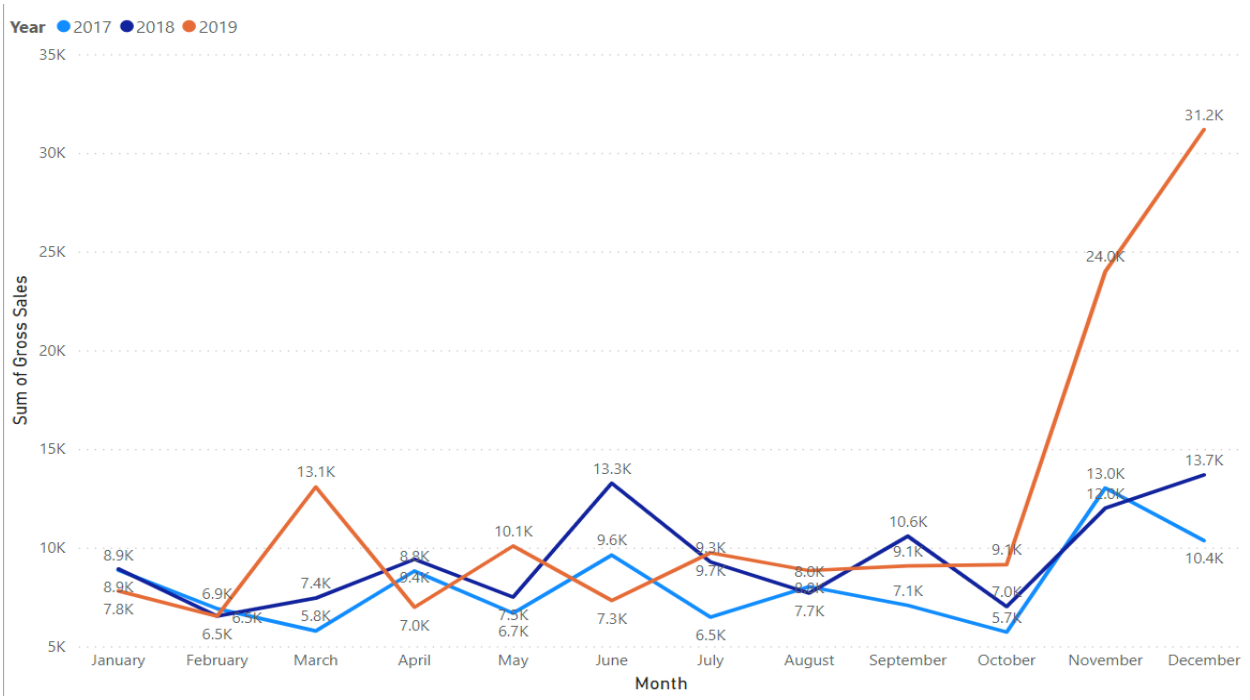


Figure 4. Gross Sales by Month

As you can see, the monthly Gross Sales have been quite **consistent in the first 10 months** it averages about 100 orders per month. The total orders start to **pick up October onwards and peak in December** before declining signficantly back to the average level in January. Likewise, the gross and total sales averages about 8k-10k per month before picking up October onwards and peak in Nov/Dec with an average of 15k, meaning that customers purchased more at the end of the year. However, in June 2018, both Gross Sales and Total orders is about 30 higher than the average sales in others month. Potential reasons might be, season specific products, cheap products in that season, or more attention in the business during Q4.

Is the discount effect the sales?

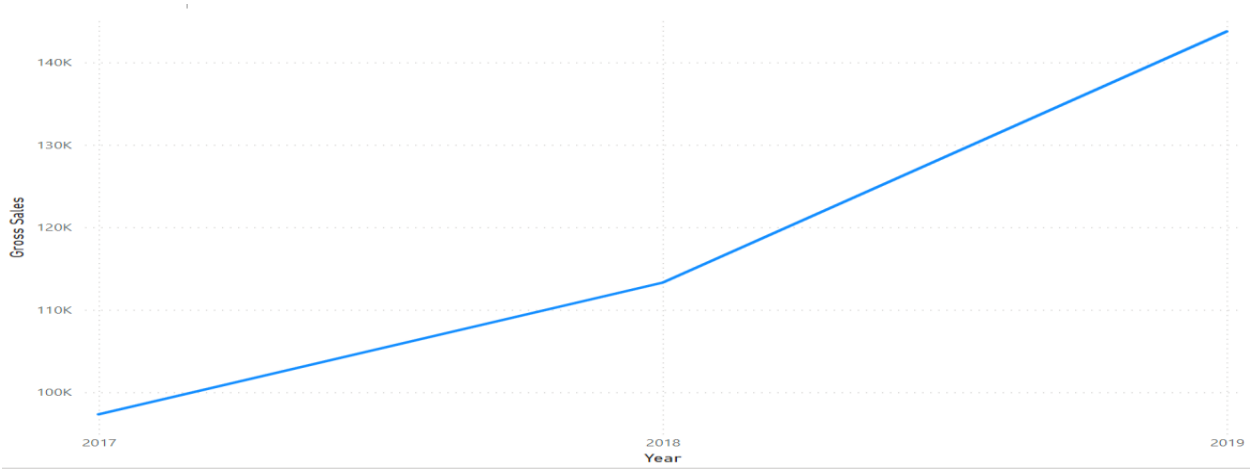


Figure 5. Gross Sales by Year

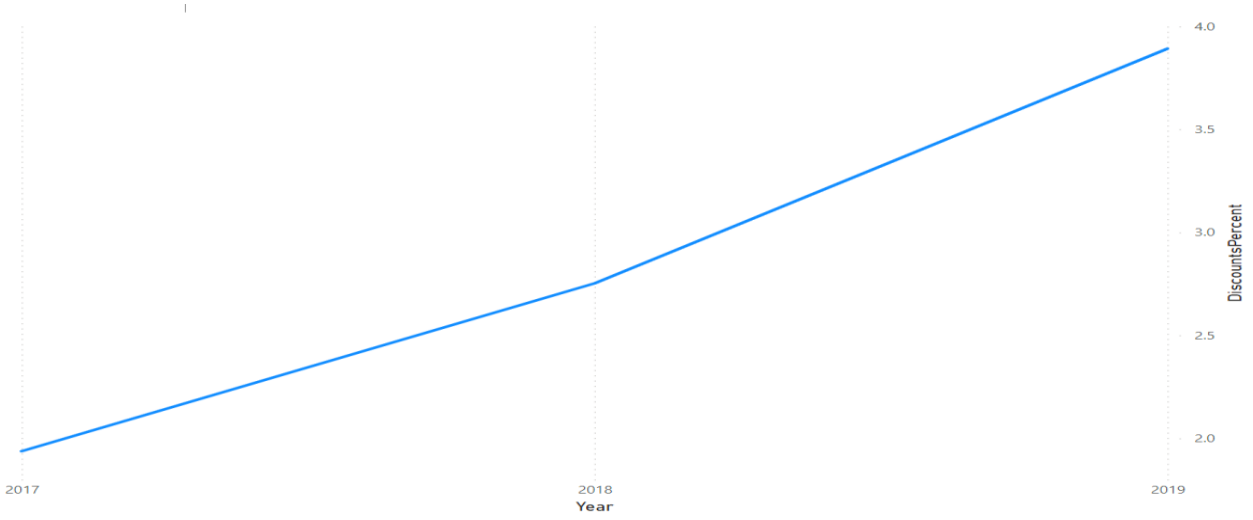


Figure 6. DiscountPercent by Year

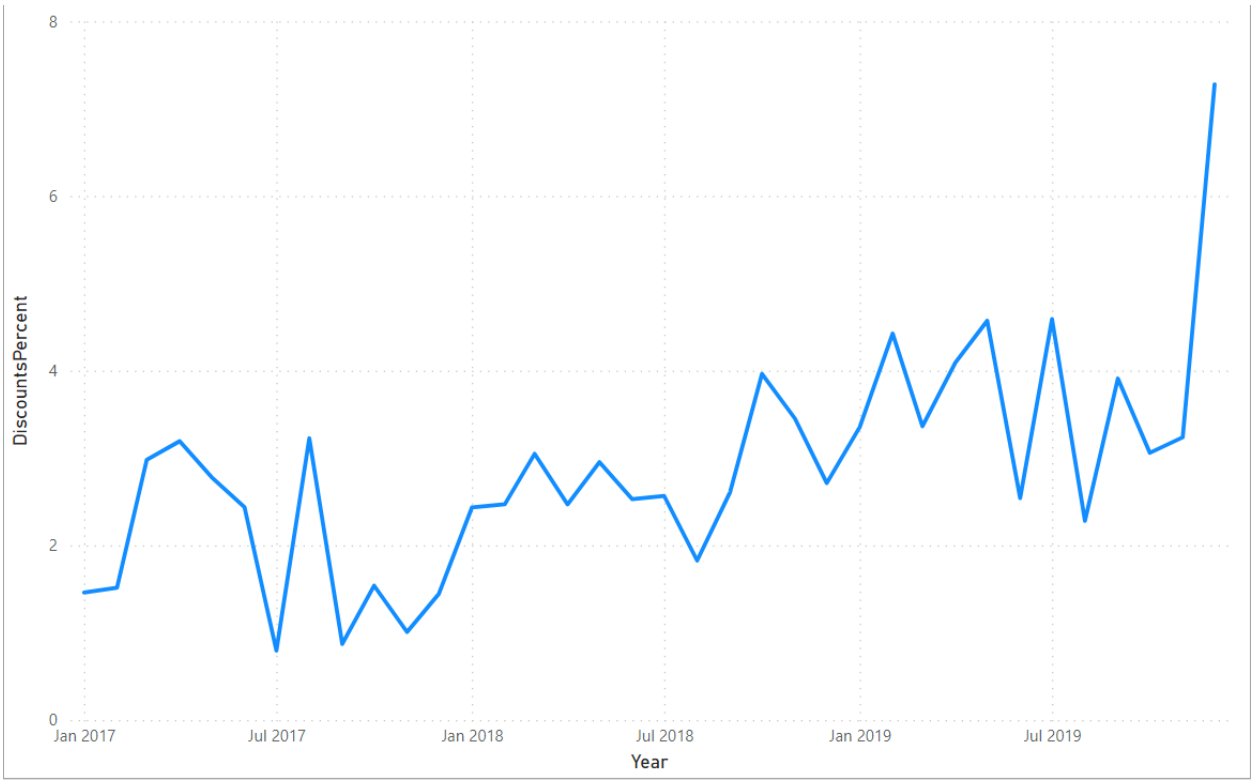


Figure 7. DiscountPercent by Month

The discount rate is increasing year on year. This means, the business is losing its profitability each year. From the customer's point of view, they are being habituated to receive more and more discount. If you disappoint them, they may not buy from you again. Discount rate should be zero or atleast constant across the years and for all customers. Discount is an expense to the business.

Discount rate in last quarter of 2017 is quite low but the sales still increased. On the contrary, discount rate in the fourth quarter of 2018 and 2019 was much higher and the sales increased too. This means that **customers will buy more products regardless there is a discount or not at the end of the year.** However, the line chart of discount rate and gross sales by year is quite similar. So increased discounts will help increase year-round sales.

How is the return rate?

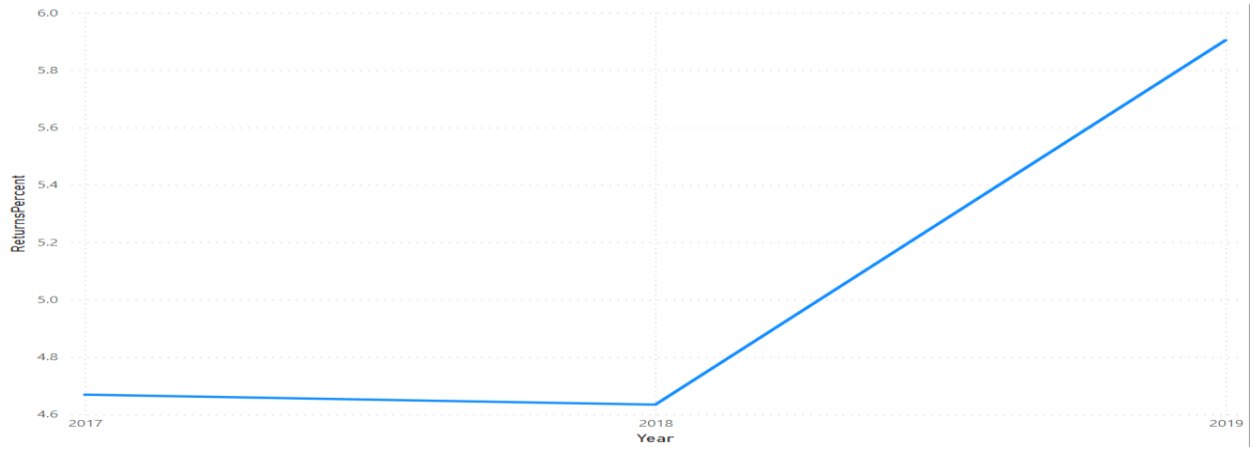


Figure 8. ReturnPercent by Year

Clearly, with the increase in sales every year, the returns % also increases. This is a bad sign for the business. Returns % has to be decreased or atleast be maintained across the years. The business has to find out the reason for the returns and has to work on it immediately, otherwise, with increase in business, most of the effort of the owners would go towards dealing with returns.

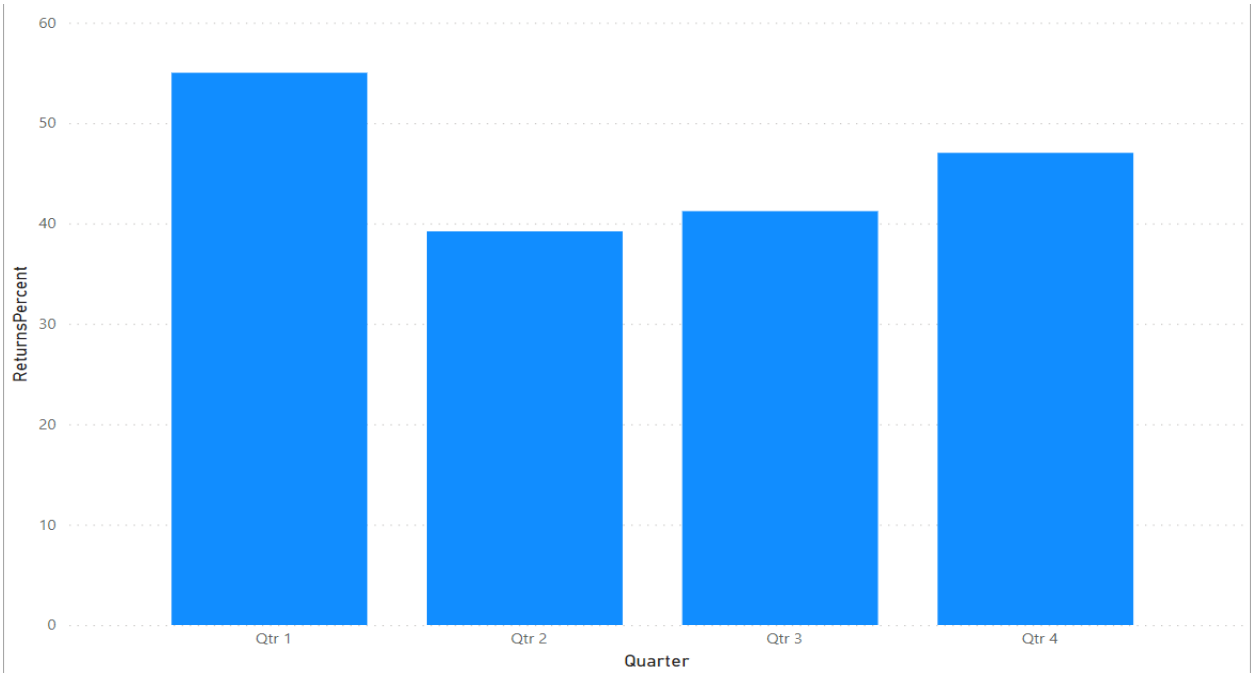


Figure 9. ReturnPercent by Quarter

Most returns occur in the first quarter and most sales occur in the last quarter (Figure 3 and 4). This could mean that **customers buy products in Q4 then return it in the next 1-2 months.**

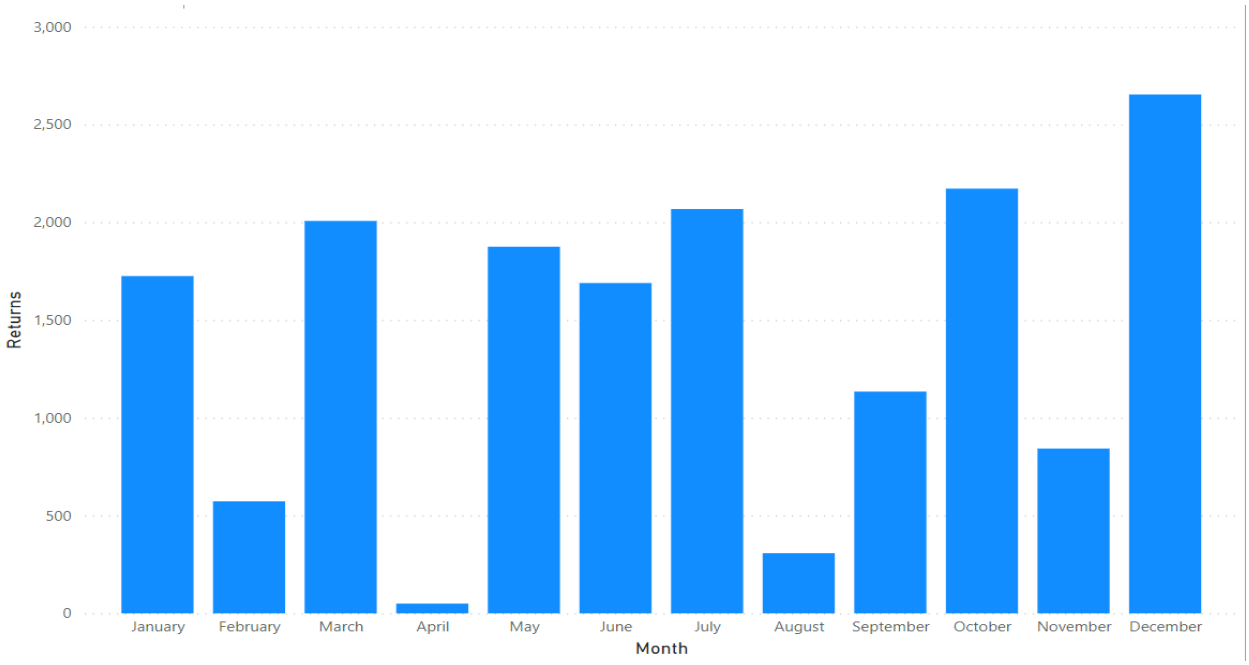


Figure 10. ReturnPercent by Month

Let drill down to return by month, we can see that after **April has the lowest returns**. The reason why cannot be found with the existing dataset. Maybe when having to return too many products, customers will reduce the number of purchases so the number of returned products is less.

Bussiness Growth Rate vs Shipping Growth Rate

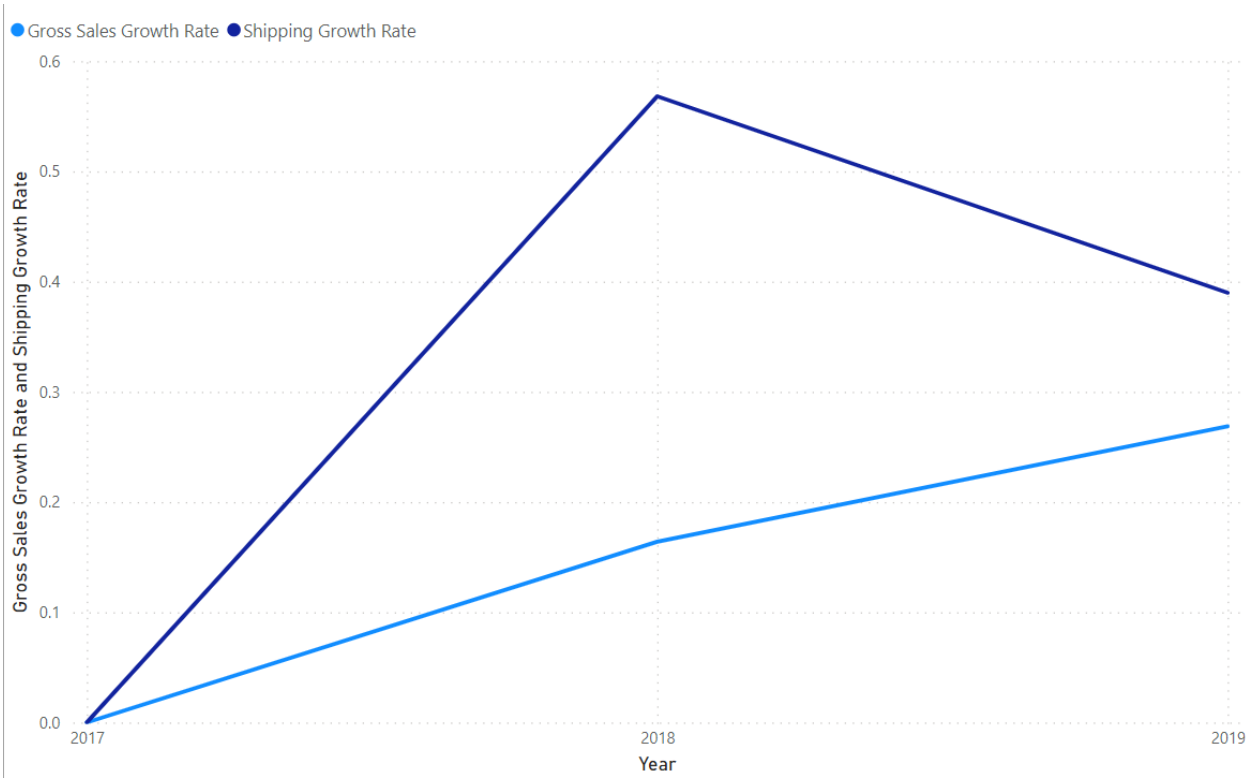


Figure 11. Growth Rate of Shipping Cost and Revenue

It is now very clear that the growth rate of shipping cost is very high compared to the growth rate of the Sales. This means that the growth of the business does not keep pace with the growth of other costs.

Product Analysis

What products were sold the most (in quantity)?

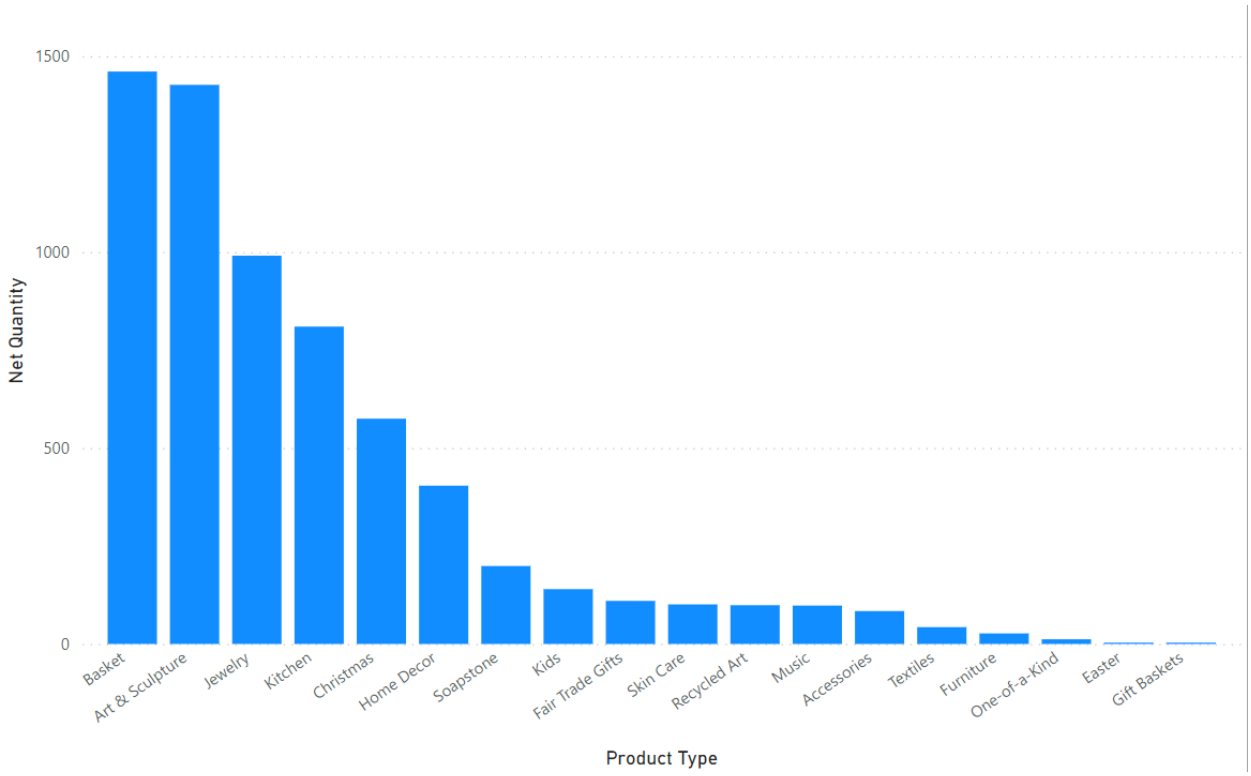
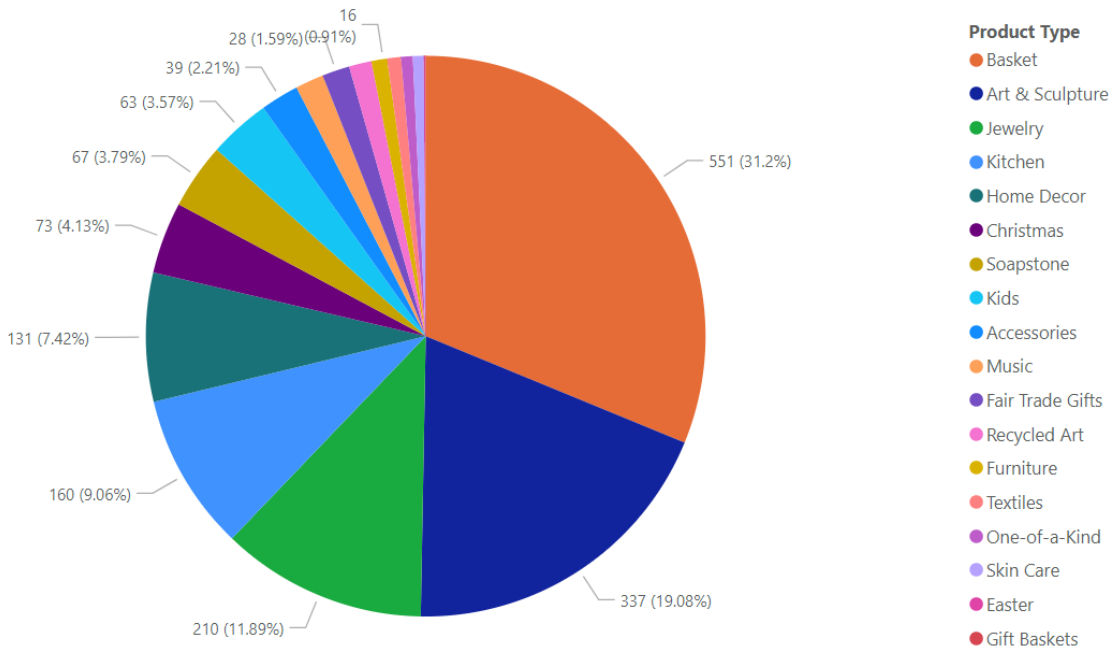


Figure 12. Net Quantity by Product Type

Baskets and Art & Sculpture are sold the most, followed by Jewelry, Kitchen and Chrismas.

What products were the most frequently bought?



As we see, the story is pretty much the same here too. Here, though, we can see a very interesting plot: **Just 5 products make up more than 75% of the demand**. The bussiness need to focus on this 5 products and may ignore product with low purchases.

What products lose us the most money in returns?

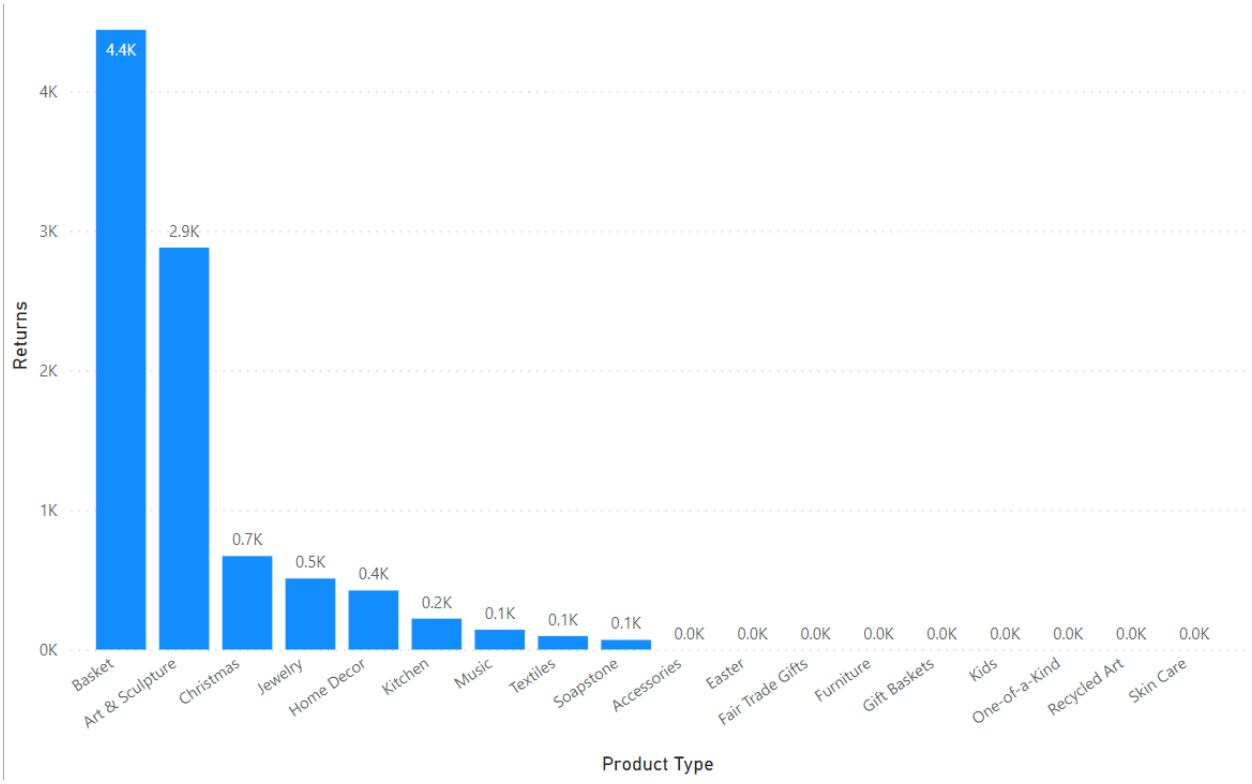


Figure 13. Returns by Product Type

As we can see from the chart above, the product that loses us the most money in returns are **Baskets Art & Sculpture**. This can cause businesses to miss out on a huge number of potential customers. On the contrary, we can also see that we have a bunch of products which are losing us nothing at all in returns.

What is the discount rate of this business?

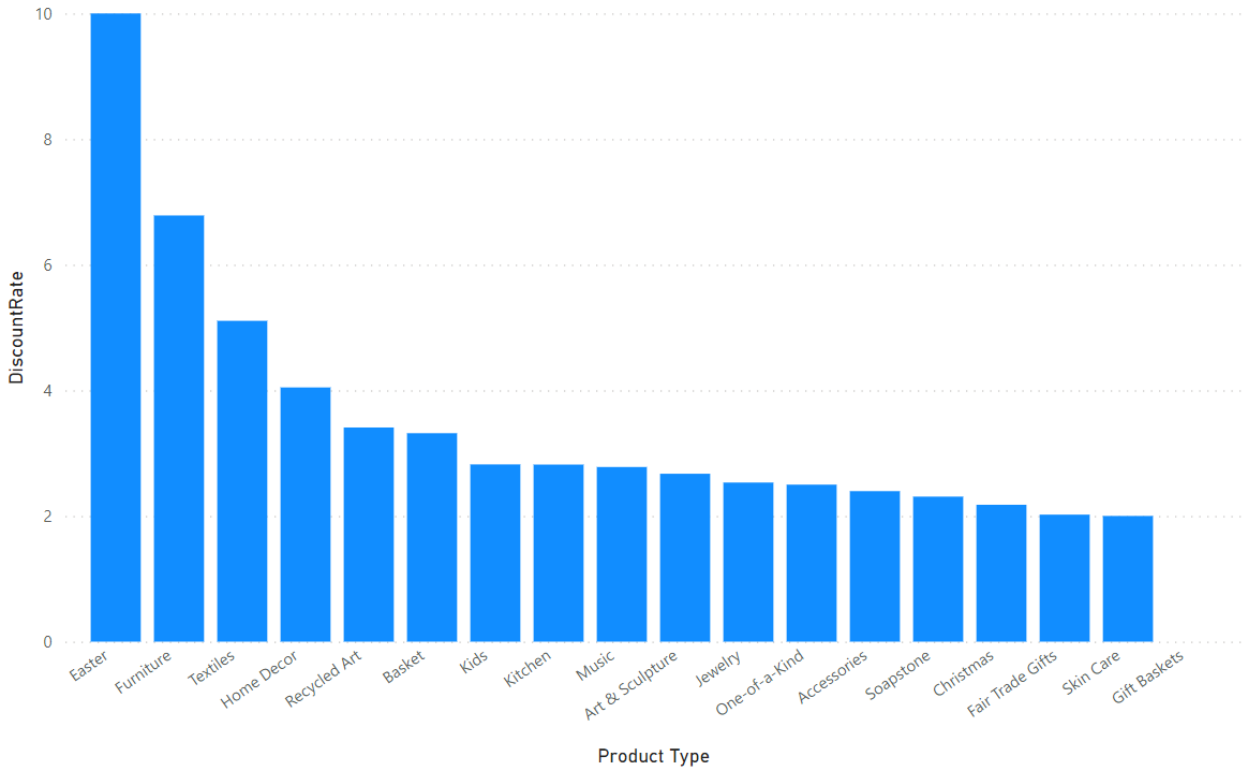


Figure 14. DiscountRate by Product Type

Products with high discount rates (**Easter and Furniture**) do not bring in much revenue for businesses in terms of both quantity and gross sales (Fig 12).

What products perform the worst in terms of gross sales vs net sales performance?

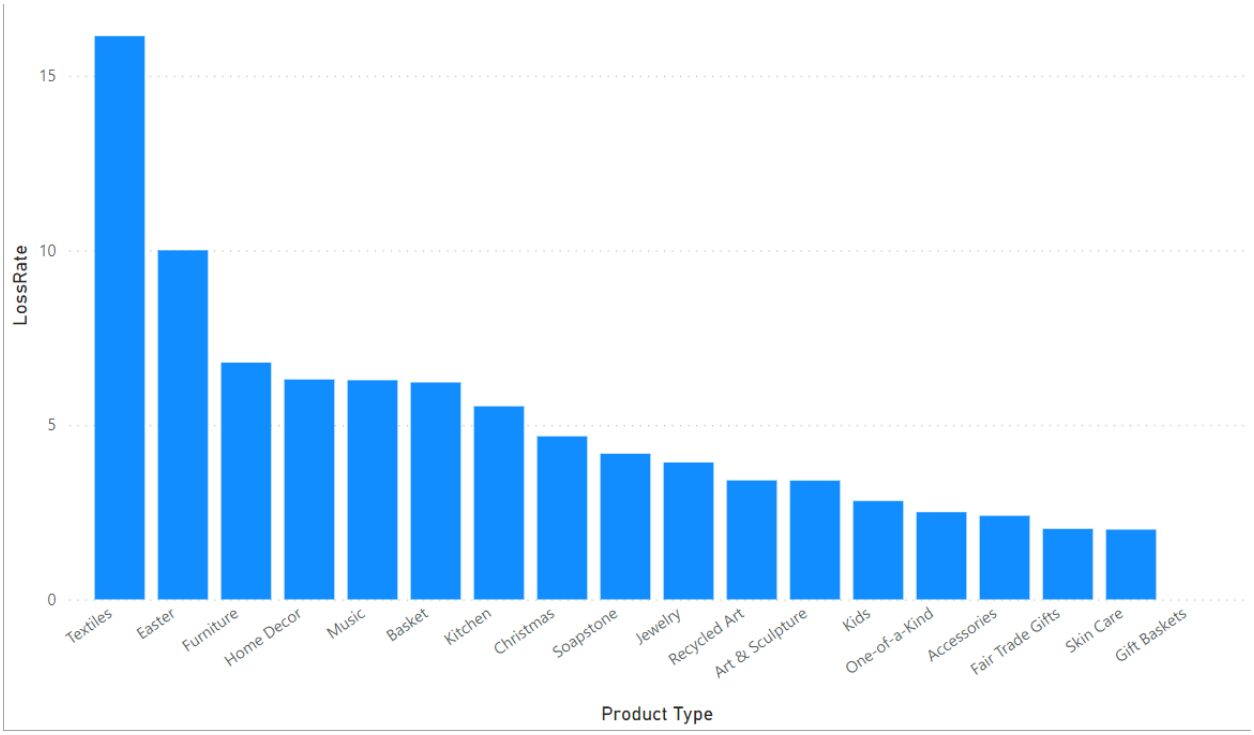


Figure 15. LossRate by Product Type

Let check what products is worst in terms of LossRate (LossRate is calculated by sum of Returns and Discounts). **Textiles** are the worst product because business will lose about 15% of revenue per order on discounts and returns.

Does more discount lead to more sales and more sales lead to more Return Rate?

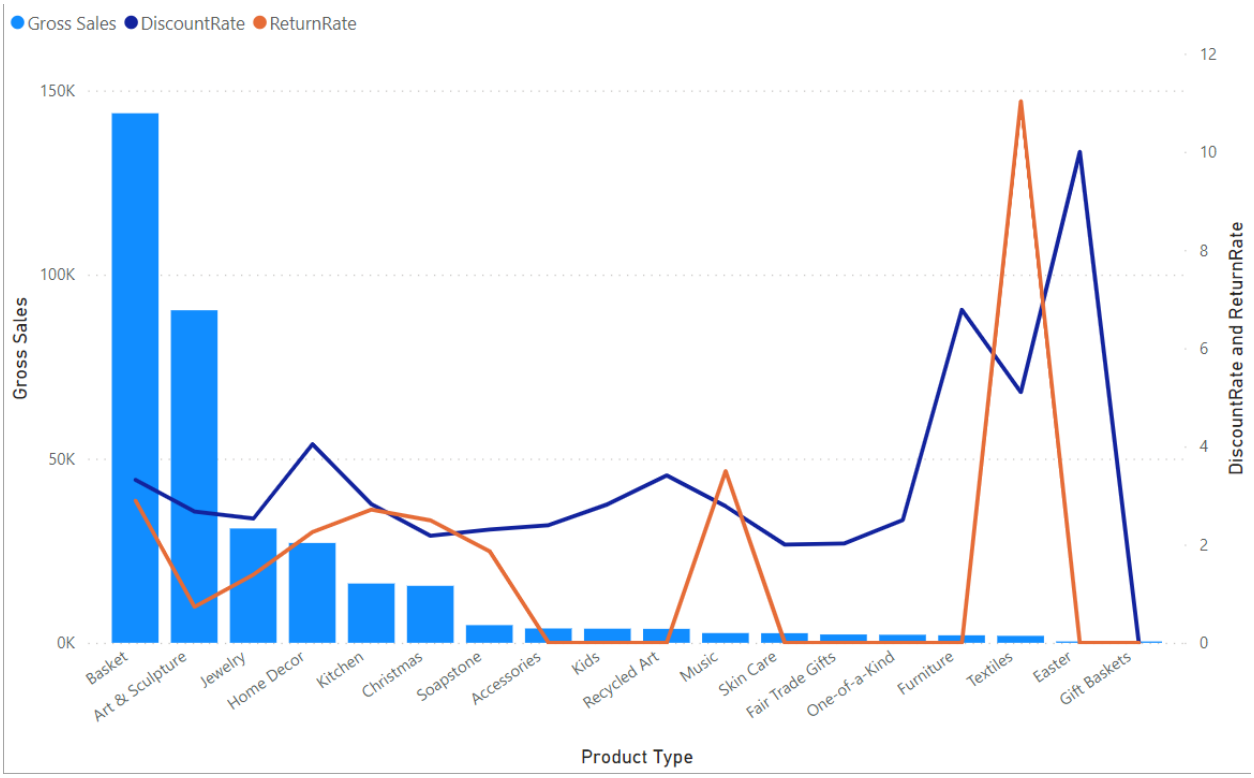


Figure 16. Gross Sales, DiscountRate and ReturnRate by Product Type

The answer is No, Basket and Art & Sculpture have high Returns but not Return Rate and they don't really need high discounts rate to reach the highest sales. Many others product cost business money (Easter and Furniture have highest DiscountRate) without bringing in revenue. Likewise, Textiles and Music have high Return Rate despite very low turnover.