

Summary of Dataset --- Transaction Analysis

Gu&Cui

- Basic summary of the products selling quantity

```
sell_products_summary['sell_num'].describe()
```

```
count      104547.000000
mean         304.057735
std          791.266060
min           1.000000
25%          14.000000
50%           65.000000
75%          286.000000
max         50287.000000
Name: sell_num, dtype: float64
```

```
sell_products_summary['sell_num'].quantile([0.90,0.95,0.99,0.999])
```

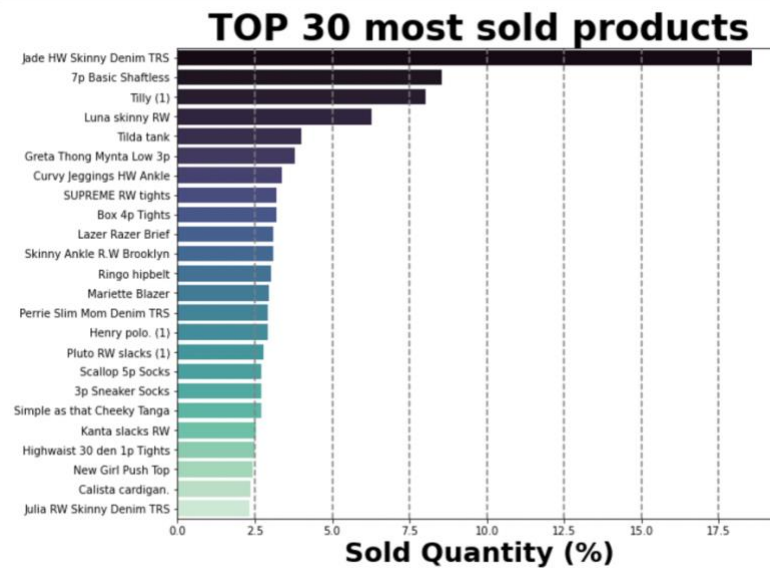
```
0.900      793.000
0.950     1318.000
0.990     3185.080
0.999     9282.534
Name: sell_num, dtype: float64
```

From the above picture, we can have some summary about the products selling quantity.

1. Totally, there are 104547 types of different products sold in the transactions dataset.
2. The mean value of selling quantity is 304 while the least sold product only sold 1 time while the most sold product sold 50287 times.
3. 25% of products were sold 14 times(or less times), 50% of products were sold 65 times(or less times), 75% of products were sold 286 times (or less times)
4. 90% of products were sold 793 times (or less times), 95% of products were sold 1318 times (or less times), 99% of products were sold 3185 times (or less times)

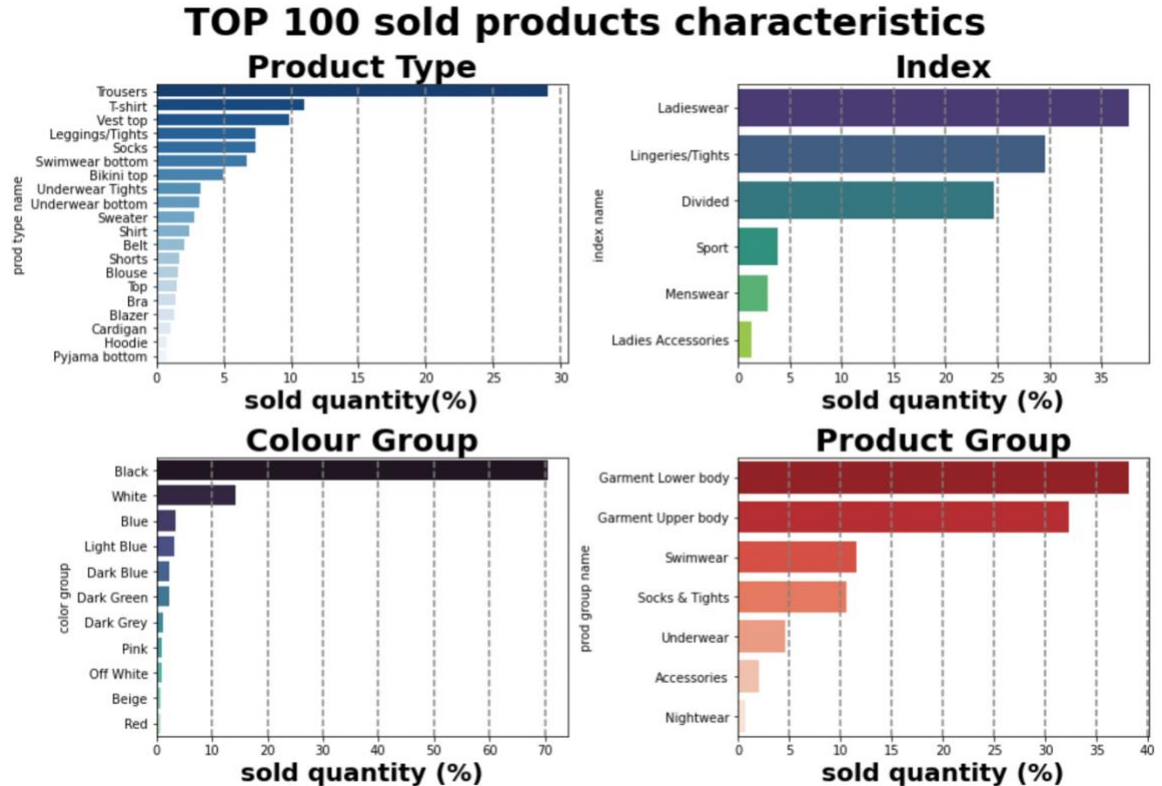
- Q1 - Which are the TOP 100 sold items? & Top 100 sold products characteristic.

Considering the size of the figure, we just show the top 30 selling products.



The details of the top 100 selling products can be found in the data frame: top_100_details (as shown in the notebook)

The characteristics of the top 100 selling products:



1. We can find that among the top 100 products' type, Trousers account for the largest proportion, which is almost 30%.
2. The most selling product index is Ladieswear, which accounts for 38%.
3. Black accounts for the largest proportion between the top 100 selling products, which accounts for more than 70%.
4. Garment Lower body accounts for the most proportion, which is 38%.

- **Q2 - Which are the products that sold only once ? & The characteristics of those products.**

We can find from below that there are totally 4491 products that were sold only once.

```
: sell_only_one = sell_products_summary["sell_num"].where(lambda x: x==1).dropna()
print(sell_only_one)
```

| | |
|--------|-----|
| 100056 | 1.0 |
| 100057 | 1.0 |
| 100058 | 1.0 |
| 100059 | 1.0 |
| 100060 | 1.0 |
| ... | |
| 104542 | 1.0 |
| 104543 | 1.0 |
| 104544 | 1.0 |
| 104545 | 1.0 |
| 104546 | 1.0 |

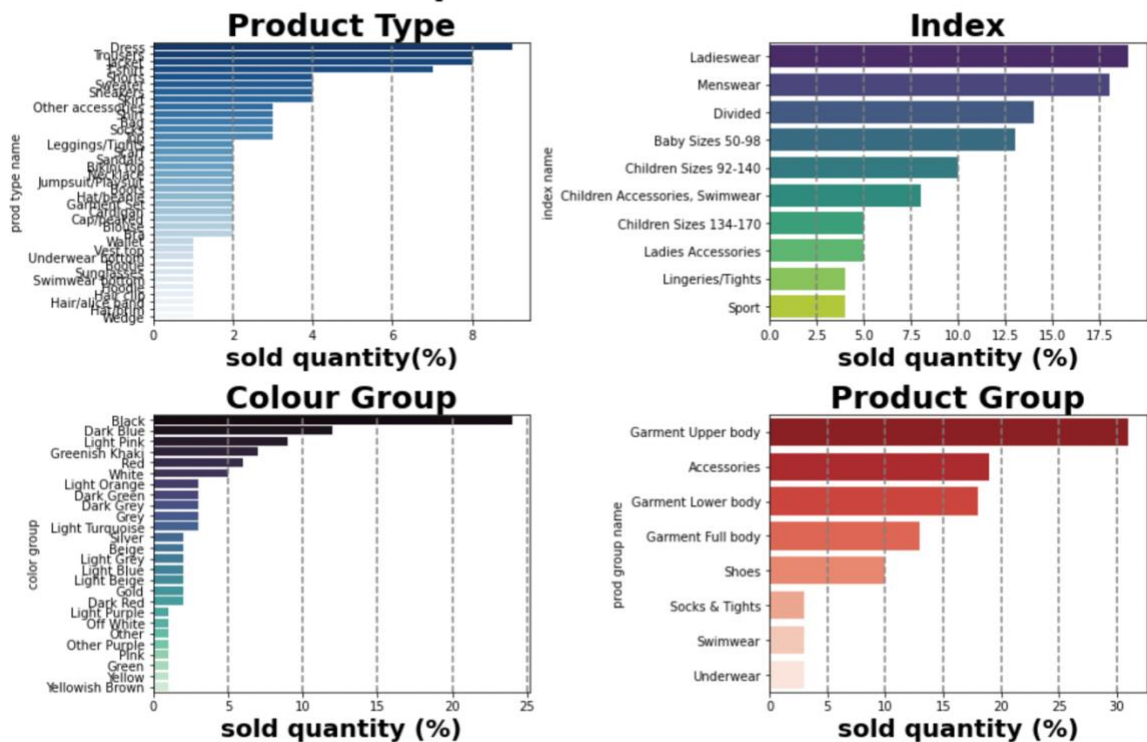
Name: sell_num, Length: 4491, dtype: float64

Since there are 104547 types of different products sold in the transactions dataset, so the products sold only once accounts for 4.3% among all the items in the store.

The characteristics of those selling once products:

1. Considering the large number of product types, we only show the 100 records of selling once products. From the below figure, we find that dress accounts for the most proportion in selling once product type, which is 9.5%.
2. "Ladieswear" accounts for the most in the selling once products' index name.(18.7%)
3. Black accounts for the most in selling once products' colour group.(about 24%)
4. Garment upper body accounts for the most proportion in once sold products, which is nearly 31%.

Sold once products characteristics



• Q3 - Which are the most products that generate most earnings?

Below is the article list ranking by the earnings descending order. The item with article_id "706016001" generates most earnings, which is 1631.7. (We don't know the exact unit of the price.)

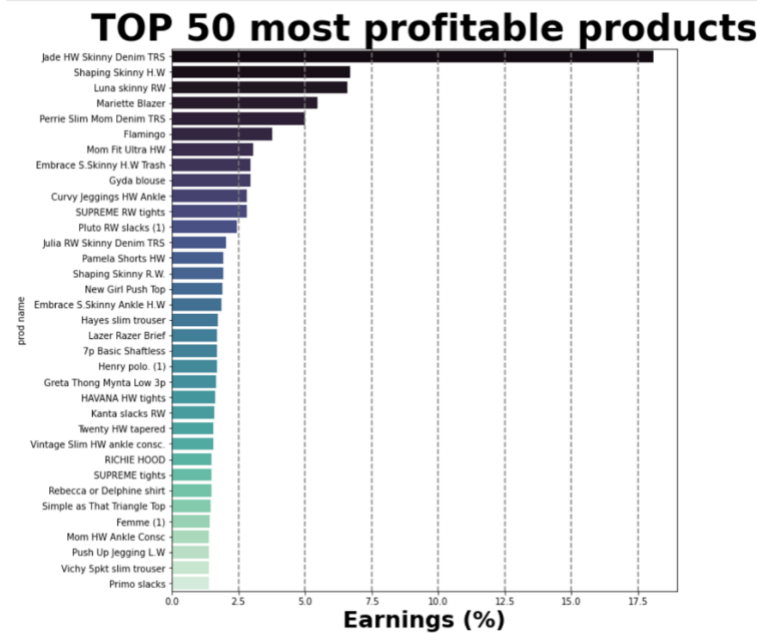
| | article_id | earning |
|--------|------------|-------------|
| 0 | 706016001 | 1631.732102 |
| 1 | 706016002 | 1136.321085 |
| 2 | 568601006 | 939.268593 |
| 3 | 448509014 | 781.478390 |
| 4 | 706016003 | 692.195915 |
| ... | ... | ... |
| 104542 | 407663007 | 0.000678 |
| 104543 | 520431003 | 0.000678 |
| 104544 | 405310006 | 0.000475 |
| 104545 | 569052001 | 0.000441 |
| 104546 | 344625009 | 0.000424 |

[104547 rows x 2 columns]

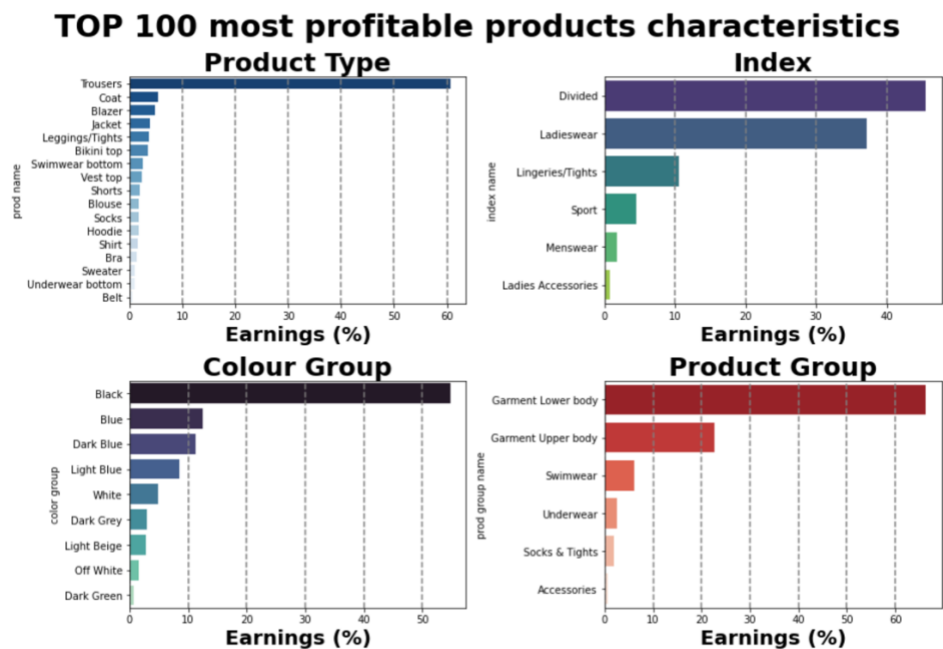
From calculating the top 100 profitable products, we can find that top 100 profitable

products accounts for 4.36 % of the total earnings (104547 different kinds of products). So next we focus on analyzing the top 100 profitable products.

Considering the big size of figure showing top 100 profitable products' name, below I just show the top 50 profitable products' name.



Some characteristics of the top 100 profitable products:

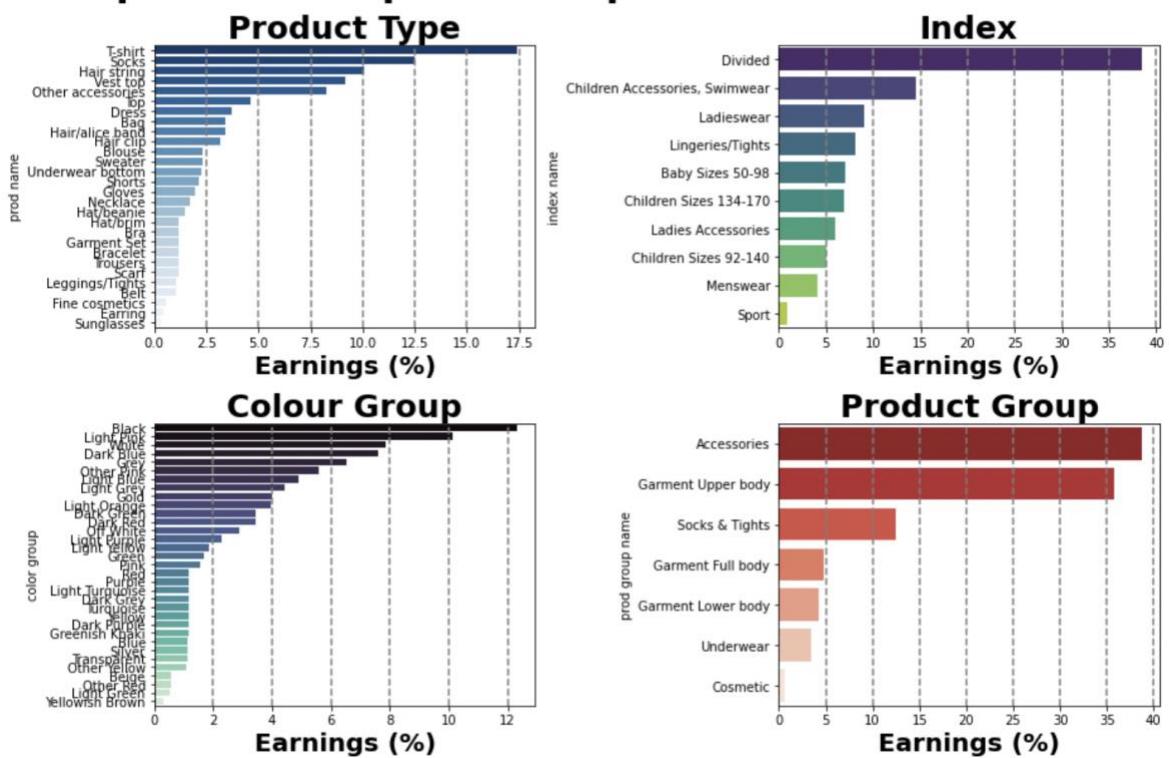


We can find from above figures that:

1. Among the top 100 profitable products' type, Trousers accounts for the most proportion (61%)
2. Among the top 100 profitable products' index, Divided accounts for the most proportion (48%)
3. Among the top 100 profitable products' colour group, Black accounts for the most proportion (56%)
4. Among the top 100 profitable products' group, Garment Lower body accounts for the most proportion (67%)

- **Q4 - Which are worst 100 profitable products?**

Flop 100 worst profitable products characteristics



1. Among the flop 100 worst profitable products' type, T-shirts accounts for the most proportion (17%)
2. Among the flop 100 worst profitable products' index, Divided accounts for the most proportion (38%)
3. Among the flop 100 worst profitable products' colour group, Black accounts for the most proportion (12.5%)
4. Among the flop 100 worst profitable products' group, Accessories accounts for the most proportion (38%)