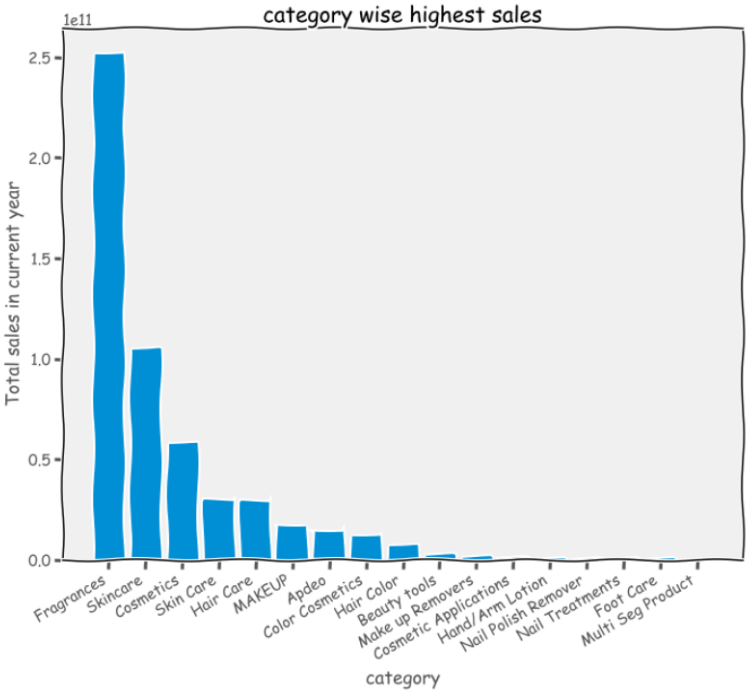
Data Analytics Table 1

What is this data and what information do you get from this data?

Seeing the table1, I find that there are total sales of different categories and different brands. And in tabel2, I find that this data is focused on Revlon products as they specifically have Revlon total sales and also total sales of all other brands through that retailer. So this data can be used to make some interpretations about the sales of those category products. The analysis report will be helpful to the brands' companies to understand how popular their products are in the market. And accordingly, they can make changes to their company.

Data Representation - Table1

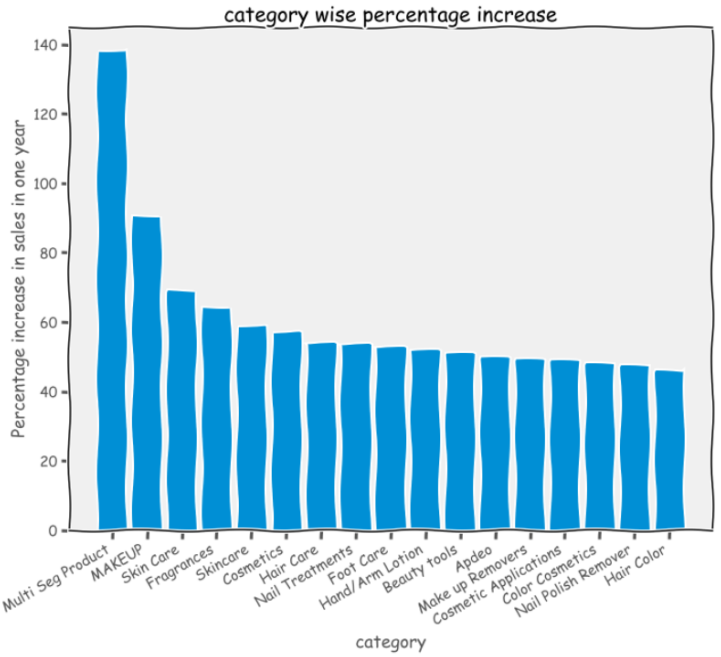
Crowd waiting to buy Fragrances



Crowd waiting to buy Multi seg product the previous year.

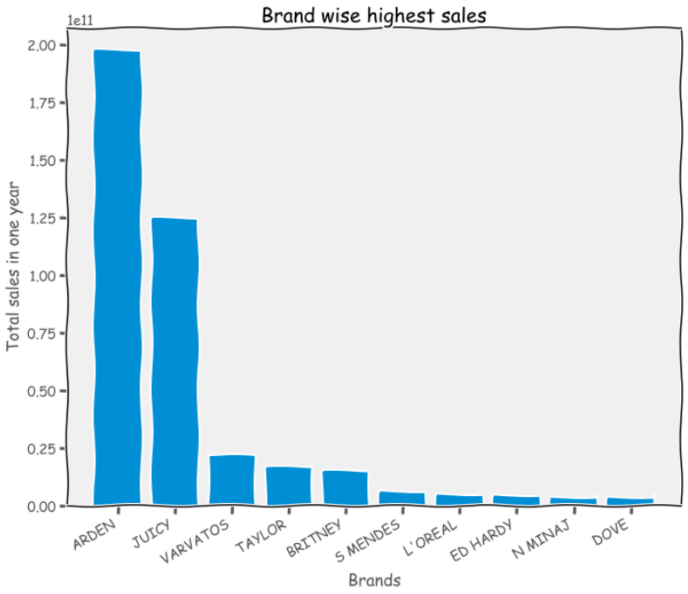
Above is the category-wise graph for total sales in the current year and you can see fragrance has highest and multi seg lowest. Below is the percentage increase from the previous year and multi seg have highest.

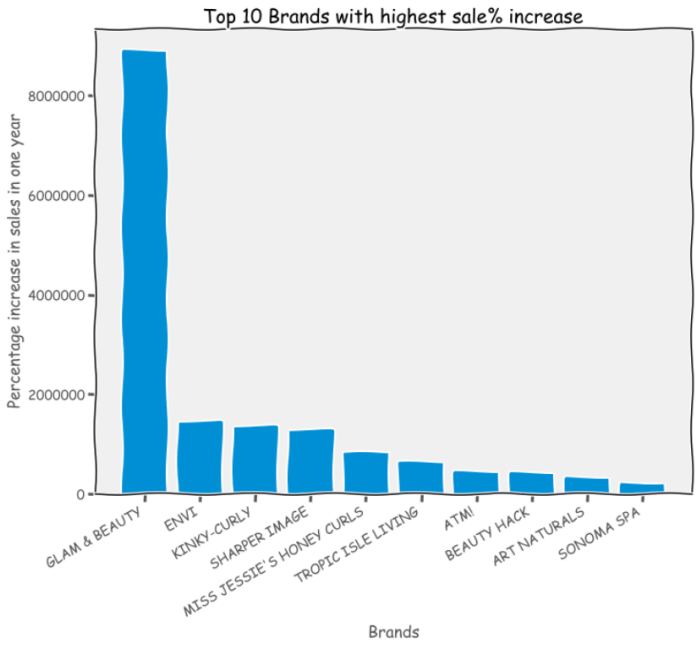
crowd waiting to buy multi seg products the current year(increased)



- This is interpretation 1 using table1, there are 17 unique category types and we have total sales for each category in a particular brand.

  And I calculated the sum of each category's total sales in all the brands and found that fragrances have the highest total sale which means many people are buying fragrance products and very few people are buying multi seg products. Reference table[1] - using Pandas python library and jupyter notebook software.
- However, in interpretation 2, I found something different! Sales difference percentage is highest for multi seg with 137%, i.e; its sales have increased by 1.37 times compared to previous. So it's not totally avoidable, it can be tested with some new changes and if still the crowd stays the same then we can avoid those products, if not it is worth keeping!
- Interpretation3 is about the brands, and in that I observed that Arden is the most popular brand and the brand that is rapidly increasing(percentage increase of sales) it's growth is Glam & Beauty.





The headline and the sub-headline for this data?

1. Headline - Why fragrances are leading products ?

   Sub-heading - Here are the leading fragrances in the market.
2. Headline - High increase in sales percent for multi seg products?

   Sub-heading - Could multi-seg products keep up this level of increment.
3. Headline - What's so good about Arden products?

   Sub-heading - Read to see why is Arden the most selling product.

Additional Information

- I used pandas to make these interpretations. First i had to deal with null values, in brand there are 3 and i deleted those rows. Then in category, total sales there are no null values. In py total sales there are 547 null values and it felt reasonable to replace them with $0.00
- Then i split the total sales column string at $ and took only the float part so to make calculations using them. And made some calculations and made some plots.

Total sales in each category including all the brands - descending order

```
currentyear_category_sales = data.groupby("category")["Total Sales"].sum().sort_values(ascending=False)
currentyear_category_sales
```

```
category
Fragrances               2.518120e+11
Skincare                 1.050536e+11
Cosmetics                5.792872e+10
Skin Care                2.974027e+10
Hair Care                2.905964e+10
MAKEUP                   1.726516e+10
Apdeo                    1.465978e+10
Color Cosmetics          1.238592e+10
Hair Color               7.391551e+09
Beauty tools             2.316441e+09
Make up Removers         1.713359e+09
Cosmetic Applications    1.151018e+09
Hand/Arm Lotion          5.040531e+08
Nail Polish Remover      4.725984e+08
Nail Treatments          4.221229e+08
Foot Care                2.933375e+08
Multi Seg Product        4.852350e+05
Name: Total Sales, dtype: float64
```

```
prevyear_category_sales = data.groupby("category")["PY Total Sales"].sum().sort_values(ascending=False)
prevyear_category_sales
```

```
category
Fragrances               1.534014e+11
Skincare                 6.618835e+10
Cosmetics                3.686800e+10
Hair Care                1.889039e+10
Skin Care                1.760170e+10
Apdeo                    9.793653e+09
MAKEUP                   9.067728e+09
Color Cosmetics          8.377478e+09
Hair Color               5.074470e+09
Beauty tools             1.535149e+09
Make up Removers         1.148575e+09
Cosmetic Applications    7.736976e+08
Hand/Arm Lotion          3.320316e+08
Nail Polish Remover      3.208941e+08
Nail Treatments          2.751674e+08
Foot Care                1.922475e+08
Multi Seg Product        2.038870e+05
Name: PY Total Sales, dtype: float64
```

Total sales percentage of increase for each category compared to previous year.

```
percentage_increase = (((currentyear_category_sales-prevyear_category_sales)/prevyear_category_sales)*100)
percentage_increase
```

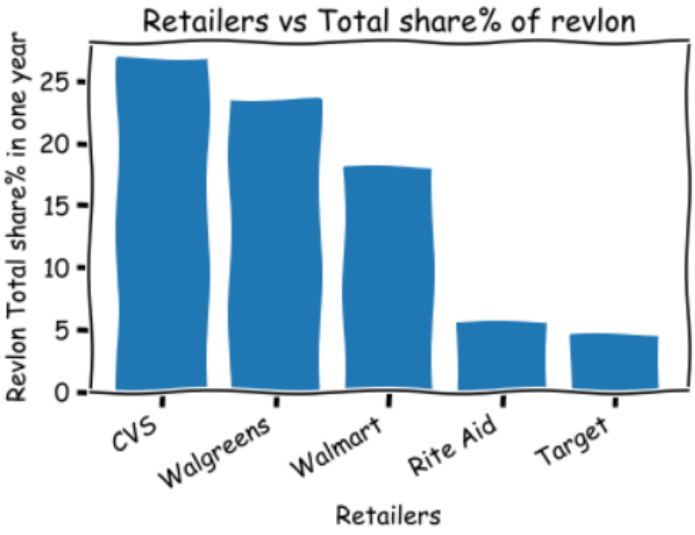
```
category
Multi Seg Product        137.992123
MAKEUP                    90.402235
Skin Care                 68.962501
Fragrances                64.152309
Skincare                  58.719229
Cosmetics                 57.124646
Hair Care                 53.832925
Nail Treatments           53.405820
Foot Care                 52.583269
Hand/Arm Lotion           51.808775
Beauty tools              50.893585
Apdeo                     49.686516
Make up Removers          49.172551
Cosmetic Applications     48.768488
Color Cosmetics           47.847808
Nail Polish Remover       47.275513
Hair Color                45.661548
dtype: float64
```

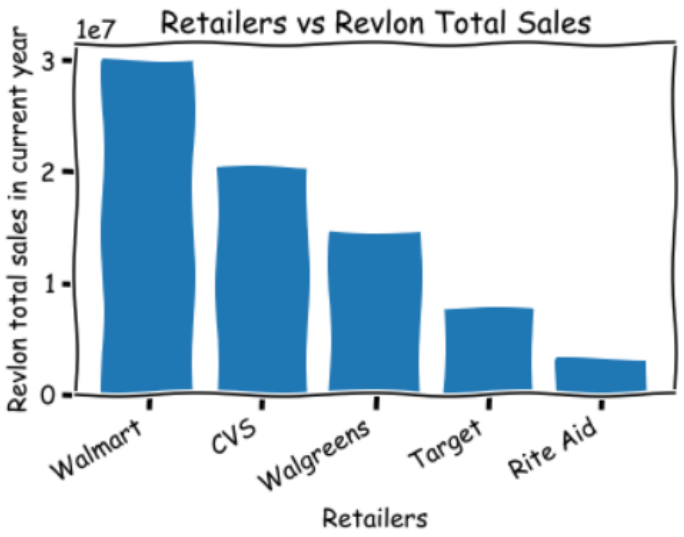
Likewise brand calculations are also done using pandas and in that there was a datetime object(at 3398 index) in brand which gave me error for a while and later found it and deleted that row. And made these interpretations for table 1.

Data Analytics Table 2

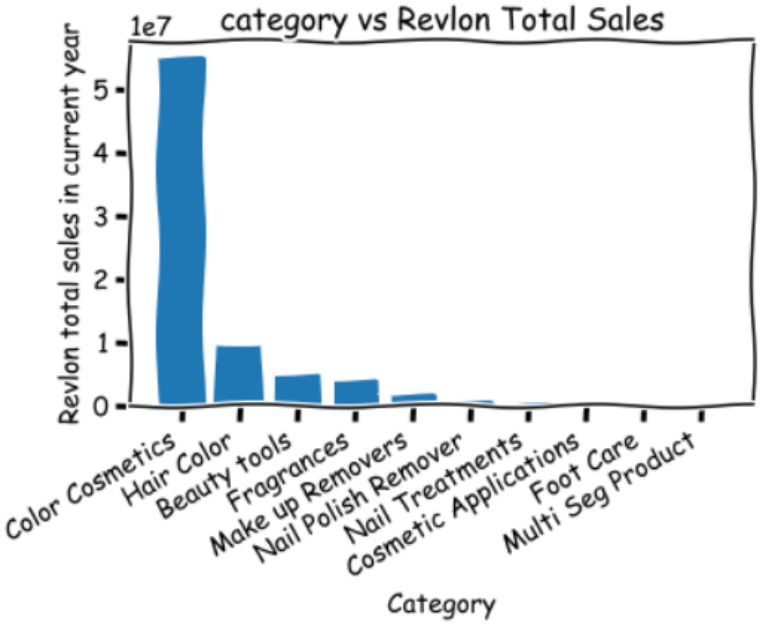
In tabel2, I find that this data is focused on Revlon products as they specifically have Revlon total sales and also total sales of all other brands through that retailer. So this data can be used to make some interpretations about the sales of Revlon products. The analysis report will be helpful to Revlon, like if they find some products have less share % increment then maybe they can consider producing it lesser. We can get these kinds of interpretations from this data. And also Revlon will get insights from what retailers their sales are high. And accordingly, they can divide their products to the retailers based on the analysis.

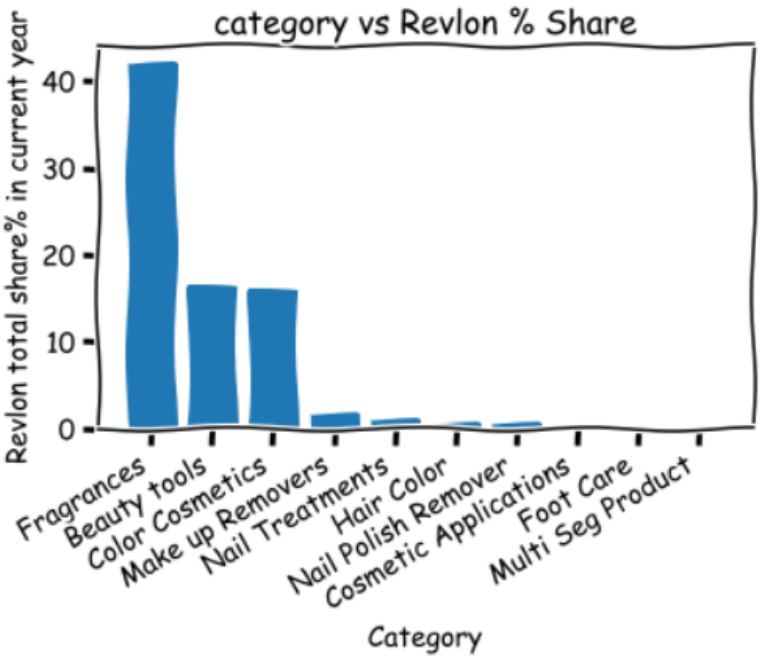
Data Representation - Table2





Interpretation 1, Here we can see total Revlon sales are high in Walmart, but it does mean that Revlon products are selling more in Walmart than other CVS. Comparatively, Walmart has high Revlon sales because Walmart has many branches. But actually “CVS” retailer has more sales of revlon, we can see that in chart1. In CVS there can be many products but more people are choosing Revlon products in CVS. So it is advisable to keep more Revlon products at CVS retailers and keep it constant with Walmart. **Interpreation 2**, Here we can see that revlon total sales are high in color cosmetics category and revlon share% is high in fragrances category. And from this we can say that it is advisable to increase products of color cosmetics as they have high total sales.





Heading and sub-heading of this data

1. Heading - CVS Is famous for Revlon?

   Sub-heading - You can go for CVS to find all Revlon products.

2. Heading - Trending color cosmetics at Revlon.

   Sub-heading - Increasing total sales at every year.

3. Heading - How come fragrances have high revlon share %

   Sub-heading - You must check out revlon fragrances

Additional Information

- Have to change data types of some columns to make use of them. Here total sales, py total sales, Revlon total sales, Revlon py total sales have $ symbol. So changed them to float types and made these interpretations.

```
cat_revlon_ts = insights.groupby("category")["Revlon Total Sales"].sum().sort_values(ascending=False)
cat_revlon_ts
```

```
category
Color Cosmetics          54835416.00
Hair Color                9601279.00
Beauty tools              5049273.00
Fragrances                4051566.00
Make up Removers          1781800.00
Nail Polish Remover        420421.04
Nail Treatments            131243.98
Cosmetic Applications       42018.18
Foot Care                      74.97
Multi Seg Product               0.00
Name: Revlon Total Sales, dtype: float64
```

```
insights.groupby("retailer")["Revlon Share % difference"].sum().sort_values(ascending=False)
```

```
retailer
Walmart      58732.5974
CVS             53.3716
Walgreens       44.3744
Rite Aid        13.6709
Target          -4.9723
Name: Revlon Share % difference, dtype: float64
```

```
revlon_ts = insights.groupby("retailer")["Revlon Total Sales"].sum().sort_values(ascending=False)
revlon_ts
## here we can see that comparitively revlon total sales is high at wallmart retailer.
```

```
retailer
Walmart      30006678.65
CVS          20447985.70
Walgreens    14619930.50
Target        7606106.44
Rite Aid      3232390.88
Name: Revlon Total Sales, dtype: float64
```