Superstore Exploratory Data Analysis (EDA) Report

Overview

This report presents the findings from an Exploratory Data Analysis (EDA) conducted on the "Superstore" dataset. The analysis focuses on understanding sales and profitability trends, identifying key factors influencing sales performance, and analyzing customer behavior and product performance. The data, preprocessed using the `preprocessing.ipynb` notebook, served as the foundation for this analysis.

The primary objectives of this EDA include: * Uncovering geographical patterns in sales and profits. * Understanding the contribution of different customer segments. * Evaluating the performance of various product categories and identifying the most profitable and best-selling items. * Analyzing temporal trends and seasonal patterns in sales and profits. * Assessing the impact of discount strategies on revenue and profitability. * Examining order characteristics such as quantity and shipping mode and their influence.

Key Findings

Below is a summary of the key findings extracted from the Superstore data analysis:

1. Geographical Sales Performance

Sales and Profit by Region:

```
             Sales        Profit
Region
West      725457.8245  108418.4489
East      678499.8680   91534.8388
Central   501239.8908   39706.3625
South     391721.9050   46749.4303
```

- The **West** region shows the highest sales and profits, followed by the **East** region.
- The **Central** region has the lowest profits compared to other regions, despite achieving higher sales than the **South** region. This may indicate profitability issues in the Central region.

Top 5 States by Sales:

```
                Sales        Profit
State
California    457687.6315   76381.3871
```

```
New York       310876.2710  74038.5486
Texas          170188.0458 -25729.3563
Washington     138641.2700  33402.6517
Pennsylvania   116511.9140 -15559.9603
```

Bottom 5 States by Sales:

```
                 Sales     Profit
State
Wyoming        1603.136   100.1960
South Dakota   1315.560   394.8283
Maine          1270.530   454.4862
West Virginia  1209.824   185.9216
North Dakota    919.910   230.1497
```

Top 5 States by Profit:

```
                 Sales       Profit
State
California  457687.6315   76381.3871
New York    310876.2710   74038.5486
Washington  138641.2700   33402.6517
Michigan     76269.6140   24463.1876
Virginia     70636.7200   18597.9504
```

Bottom 5 States by Profit (showing losses):

```
                  Sales        Profit
State
Texas          170188.0458 -25729.3563
Ohio            77976.7640 -16959.3178
Pennsylvania   116511.9140 -15559.9603
Illinois        80166.1010 -12607.8870
North Carolina  55603.1640  -7490.9122
```

- **California** and **New York** lead in terms of both sales and profit.
- Some states like **Texas**, **Pennsylvania**, **Ohio**, **Illinois**, and **North Carolina** show significant losses despite some achieving high sales. This calls for a deeper investigation into the causes of these losses (e.g., high discounts or operational costs).

2. Customer Segment Analysis

Sales and Profit by Customer Segment:

```
                 Sales        Profit
Segment
Consumer     1.161401e+06  134119.2092
Corporate    7.061464e+05   91979.1340
Home Office  4.293718e+05   60310.7373
```

Segment Contribution to Total Sales (%):

```
Segment
Consumer      50.563433
Corporate     30.743192
Home Office   18.693375
```

Segment Contribution to Total Profit (%):

```
Segment
Consumer      46.827848
Corporate     32.114601
Home Office   21.057551
```

- The **Consumer** segment is the largest in terms of sales and profit, followed by the **Corporate** segment, and then **Home Office**.
- The distribution of profit contribution percentages roughly follows the same order as sales.

3. Product Performance Analysis

Sales and Profit by Product Category:

```
                   Sales        Profit
Category
Technology      836154.0330  145454.9481
Furniture       741718.4233   18463.3316
Office Supplies 719047.0320  122490.8008
```

- The **Technology** category achieves the highest sales and profits.
- The **Furniture** category generates good sales, but its profits are very low compared to other categories, indicating slim profit margins or high costs associated with this category.
- The **Office Supplies** category achieves good sales and profits.

Top 5 Sub-Categories by Sales:

```
                 Sales        Profit
Sub-Category
Phones        330007.054   44515.7306
Chairs        328167.731   26602.2251
Storage       223843.608   21278.8264
Tables        206965.532  -17725.4811
Binders       203412.733   30221.7633
```

Bottom 5 Sub-Categories by Sales:

```
                Sales        Profit
Sub-Category
Supplies      46673.538   -1189.0995
Art           27118.792    6527.7870
Envelopes     16476.402    6964.1767
Labels        12486.312    5546.2540
Fasteners      3024.280     949.5182
```

Top 5 Sub-Categories by Profit:

```
                   Sales        Profit
Sub-Category
Copiers        149528.030   55617.8249
Phones         330007.054   44515.7306
Accessories    167380.318   41936.6357
Paper           78479.206   34053.5693
Binders        203412.733   30221.7633
```

Bottom 5 Sub-Categories by Profit (showing losses or very low profits):

```
                  Sales        Profit
Sub-Category
Tables        206965.5320  -17725.4811
Bookcases     114879.9963   -3472.5560
Supplies       46673.5380   -1189.0995
Fasteners       3024.2800     949.5182
Machines      189238.6310    3384.7569
```

- Products like **Phones** and **Chairs** achieve high sales. **Copiers**, **Phones**, and **Accessories** are the most profitable.
- The **Tables** sub-category generates significant sales but incurs substantial losses, making it the largest loss-making sub-category. **Bookcases** and **Supplies** also contribute to losses.

4. Temporal Analysis (Seasonality and Trends)

Sales and Profit by Order Year:

```
               Sales        Profit
Order Year
2019       483966.1261   49556.0329
2020       470532.5090   61618.6037
2021       609205.5980   81795.1743
2022       733215.2552   93439.2696
```

- There is a clear upward trend in both sales and profit over the years from 2019 to 2022, indicating business growth.

Sales and Profit by Order Month (Aggregated across years):

```
                Sales        Profit
Order Month
1            95086.8056    9197.6623
2            61231.4558   10244.1127
3           203363.3144   28581.9690
4           137931.6966   11665.7185
5           155412.0697   22645.5630
6           151884.4813   20985.3168
7           148249.7930   13660.1540
8           160399.9310   22701.3920
```

```
9            306440.7957  36275.2327
10           202270.6767  32536.3976
11           356485.5590  35499.4790
12           318162.9095  42416.0829
```

- Some months perform stronger than others. **September (9)**, **November (11)**, and **December (12)** are the months with the highest sales. **March (3)**, **September (9)**, **October (10)**, **November (11)**, and **December (12)** achieve high profits.
- **February (2)** is the lowest sales month, while **January (1)** is the least profitable.

Sales and Profit by Order Quarter (Aggregated across years):

```
                    Sales        Profit
Order Quarter
2019Q1           74447.7960    3811.2290
2019Q2           86257.3876   11216.1280
2019Q3          143633.2123   12804.7218
2019Q4          179627.7302   21723.9541
2020Q1           68851.7386    9264.9416
2020Q2           89124.1870   12190.9224
2020Q3          131417.9892   17023.3196
2020Q4          181138.5942   23139.4201
2021Q1           93237.1810   11441.3708
2021Q2          136082.3010   16390.3394
2021Q3          143787.3622   15823.6048
2021Q4          236098.7538   38139.8593
2022Q1          123144.8602   23506.2026
2022Q2          133764.3720   15499.2085
2022Q3          196251.9560   26985.1325
2022Q4          280054.0670   27448.7260
```

- Generally, the fourth quarter (Q4) of each year tends to achieve the highest sales and profits, indicating strong year-end seasonality.

5. Discount Strategy Impact

Impact of Discount on Average and Total Sales/Profit:

```
              Sales                        Profit
              mean          sum            mean          sum
Discount
0.00       226.742074   1.087908e+06    66.900292    320987.6032
0.10       578.397351   5.436935e+04    96.055074      9029.1770
0.15       529.971567   2.755852e+04    27.288298      1418.9915
0.20       209.076940   7.645944e+05    24.702572     90337.3060
0.30       455.510102   1.029453e+05   -45.828401    -10357.2186
0.32       536.794770   1.449346e+04   -88.560656     -2391.1377
0.40       565.134874   1.164178e+05  -111.927429    -23057.0504
0.45       498.634000   5.484974e+03  -226.646464     -2493.1111
0.50       892.705152   5.891854e+04  -310.703456    -20506.4281
```

```
0.60          48.150000  6.644700e+03   -43.077212    -5944.6552
0.70          97.177708  4.062028e+04   -95.874060   -40075.3569
0.80          56.545853  1.696376e+04  -101.796797   -30539.0392
```

Correlation between Discount and Sales/Profit: * Correlation between Discount and Sales: -0.03 * Correlation between Discount and Profit: -0.22

- There is a slight negative correlation between discount and sales, which might mean that discounts do not necessarily lead to a significant increase in total sales value.
- More importantly, there is a stronger negative correlation between discount and profit (**-0.22**). As the discount increases, profits decrease noticeably.
- Discounts at 0% and 20% (0.2) achieve significant positive total profits. Higher discounts (30% and above) lead to substantial losses. For instance, discounts of 50%, 70%, and 80% are associated with large losses per unit sold on average, and massive total losses.

6. Order Characteristics Analysis

Average Sales and Profit by Order Quantity:

```
              Sales       Profit
Quantity
1           59.234632     8.276396
2          120.287425    16.018520
3          175.201578    23.667715
4          271.764059    37.131310
5          337.936339    40.257394
6          362.101960    18.051517
7          395.888393    56.579163
8          458.210802    42.244342
9          498.083683    68.557716
10         422.046737    35.862404
11         463.896529   126.275503
12         646.452522    51.524609
13        1197.710704   235.144481
14         452.826138    93.931166
```

- Generally, average sales increase with the quantity ordered. The relationship with average profit is not strictly linear, as some quantities (like 6, 8, and 10) show lower average profit than adjacent quantities. This might indicate that products bought in certain quantities could have lower profit margins or be subject to higher discounts.

Sales and Profit by Ship Mode:

```
                    Sales         Profit
Ship Mode
Standard Class   1.357934e+06   164100.8463
Second Class     4.591936e+05    57446.6354
```

```
First Class       3.514284e+05    48969.8399
Same Day          1.283631e+05    15891.7589
```

Order Count by Ship Mode:

```
Ship Mode
Standard Class    5967
Second Class      1945
First Class       1538
Same Day           543
```

- **Standard Class** shipping is the most used and achieves the highest sales and profits.
- All shipping modes are profitable, and their profit contribution follows the order of their sales contribution.

Recommendations

Based on the analysis above, the following recommendations can be made:

1. **Improve Geographical Profitability**: Investigate the causes of losses in states like Texas, Ohio, Pennsylvania, Illinois, and North Carolina. This might involve reviewing pricing and discount strategies, or operational costs in those areas. Focus on enhancing performance in profitable regions like the West and East.
2. **Customer Segment Strategies**: Continue focusing on the Consumer and Corporate segments. Explore ways to increase profitability from the Home Office segment.
3. **Product Performance Management**:
   - Invest in the Technology and Office Supplies categories due to their high profitability.
   - Urgently review the strategy for the Furniture category, especially loss-making sub-categories like "Tables" and "Bookcases." This might require renegotiating with suppliers, adjusting prices, or even discontinuing unprofitable products.
4. **Temporal and Seasonal Planning**: Leverage annual growth trends. Plan inventory and promotions based on high-performing months and quarters (especially Q4).
5. **Optimize Discount Strategy**: Exercise extreme caution when offering discounts. Discounts exceeding 20% often lead to losses. The company should aim to minimize the use of high discounts and focus on value rather than low price to achieve profitability.
6. **Incentivize Larger Order Sizes (Carefully)**: While increasing quantity boosts sales, ensure that products bought in bulk are profitable. A review of discounts on large quantities might be necessary.
7. **Shipping Modes**: No clear issues with shipping modes, as all are profitable. Continue offering diverse options to customers.

Additional Project Assets

- **Preprocessing Notebook**: `preprocessing.ipynb` (used to prepare the data for this analysis).
- **Preprocessed Data**: `preprocessed_superstore_data.csv` (the output file from preprocessing, used in this analysis).
- **Raw Analysis Findings**: `eda_findings.txt` (contains the textual output from the analysis script).