Basic Data Analysis

The Case of a Superstore

Salau Ayomide

Outline

- Dataset Information and Source
- Case Questions
- Data Cleaning using Spreadsheet
- EDA using MySQL
- Data Visualization with PowerBI
- Findings and Recommendation
- Conclusion.

Dataset Information and Source

THIS IS A SAMPLE SUPERSTORE DATASET, A KIND OF A SIMULATION WHERE YOU PERFORM EXTENSIVE DATA ANALYSIS TO DELIVER INSIGHTS ON HOW THE COMPANY CAN INCREASE ITS PROFITS WHILE MINIMIZING THE LOSSES.

THE DATASET CAN BE FOUND ON KAGGLE.COM.

THE IS THE LINK TO THE DATASET (HTTPS://WWW.KAGGLE.COM/DATASETS/BRAVEHART101/SAMPLE-SUPERMARKET-DATASET)

THE DATASET WILL NOT BE UPDATED.

Case Questions

- Which Category is Best Selling and Most Profitable?
- What are the Best Selling and Most Profitable Sub-Category?
- Which is the Top Selling Sub-Category?
- Which Customer Segment is Most Profitable?
- Which is the Preferred Ship Mode?
- Which Region is the Most Profitable?
- Which City has the Highest Number of Sales?

Data cleaning using Spreadsheet

- The Data has 9,997 rows and 2 deleted empty cells at the bottom.

- There are no whitespaces in the data.

EDA using MySQL

- Which Category is Best Selling and Most Profitable?

```sql
-- Which Category is Best Selling and Most Profitable?
SELECT Category,
ROUND(SUM(Sales),2) sales_by_cat, ROUND(SUM(Profit),2) profit_by_cat
FROM samplesuperstore
GROUP BY 1
ORDER BY 1 DESC;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

Category	sales_by_cat	profit_by_cat
Technology	836154.03	145454.95
Office Supplies	719047.03	122490.8
Furniture	741999.8	18451.27

- What are the Best Selling and Most Profitable Subcategory?

```sql
-- What are the Best Selling and Most Profitable Sub-Category
SELECT Sub_Category,
ROUND(SUM(Sales),2) sales_by_subcat, ROUND(SUM(Profit),2) profit_by_subcat
FROM samplesuperstore
GROUP BY 1
ORDER BY 2 DESC
LIMIT 5;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content: | Fetch rows:

Sub_Category	sales_by_subcat	profit_by_subcat
Phones	330007.05	44515.73
Chairs	328449.1	26590.17
Storage	223843.61	21278.83
Tables	206965.53	-17725.48
Binders	203412.73	30221.76

EDA using MySQL

- Which is the Top Selling Sub-Category?

```
20      -- Which is the Top Selling Sub-Category
21 •    SELECT Sub_Category, ROUND(SUM(Sales),2) Top_Sales
22      FROM samplesuperstore
23      GROUP BY 1
24      ORDER BY 2 DESC
25      LIMIT 5;
26
```

Result Grid | Filter Rows: | Export: | Wrap Cell C

Sub_Category	Top_Sales
Phones	330007.05
Chairs	328449.1
Storage	223843.61
Tables	206965.53
Binders	203412.73

- Which Customer Segment is Most Profitable?

```
27      -- Which Customer Segment is Most Profitable?
28 •    SELECT Segment, ROUND(sum(Profit),2) profit
29      FROM samplesuperstore
30      GROUP BY 1
31      ORDER BY 2 DESC;
32
```

Result Grid | Filter Rows: | Export: | Wra

Segment	profit
Consumer	134119.21
Corporate	91979.13
Home Office	60298.68

EDA using MySQL

- Which is the Preferred Ship Mode?

```
33      -- Which is the Preferred Ship Mode?
34 •    SELECT Ship_Mode, count(*)
35      FROM samplesuperstore
36      GROUP BY 1
37      ORDER BY 2 DESC;
```

Result Grid | Filter Rows: | Export:

Ship_Mode	count(*)
Standard Class	5968
Second Class	1945
First Class	1538

- Which Region is the Most Profitable?

```
39      -- Which Region is the Most Profitable?
40 •    SELECT Region, ROUND(sum(Profit),2) profit
41      FROM samplesuperstore
42      GROUP BY 1
43      ORDER BY 2 DESC;
```

Result Grid | Filter Rows: | Export:

Region	profit
West	108418.45
East	91522.78
South	46749.43
Central	39706.36

EDA using MySQL

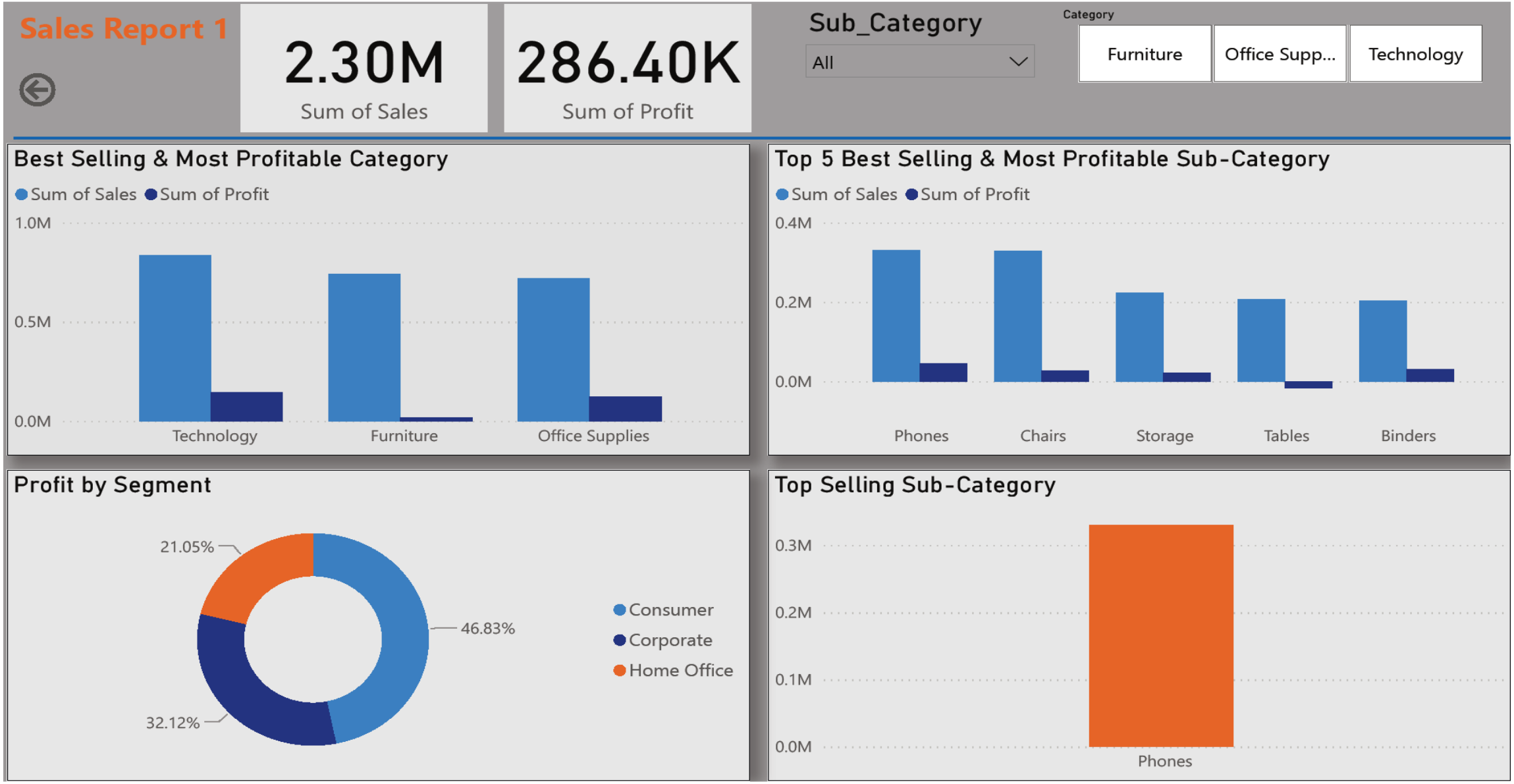
- Which City has the Highest Number of Sales?

```
45      -- Which City has the Highest Number of Sales?
46  •   SELECT distinct City, ROUND(sum(sales),2) Top_Sales
47      FROM samplesuperstore
48      GROUP BY 1
49      ORDER BY 2 DESC
50      LIMIT 10;
```

Result Grid | Filter Rows: | Export: | Wrap Cell Co

City	Top_Sales
New York City	256368.16
Los Angeles	175851.34
Seattle	119540.74
San Francisco	112669.09
Philadelphia	109077.01
Houston	64504.76
Chicago	48539.54

Case Solution 1



Case Solution 2