

Superstore sales analysis and customer segmentation

A Comprehensive Data Analytics Project on Superstore Dataset



*Prepared by: Nihar Nandanwar*  
*June 2025*

# **Project Title:** Sales Analysis for Superstore Dataset **Overview:** This project analyses sales, profit, and discount data for a retail superstore across various regions, product categories, and time periods. The goal is to identify profitable segments and offer actionable business insights using SQL, Python, and Power BI.

**Dataset link:** <https://www.kaggle.com/datasets/vivek468/superstore-dataset-final?resource=download>

## **Objective:** To uncover which regions, product categories, and discount strategies generate the most profit, and whether certain discount levels negatively impact overall revenue.

## **Data Collection:** The dataset was taken from Kaggle, and the single CSV file was divided into 4 files based on our design of a relational database. So I created 4 tables in the superstore\_db database, namely:

* Customers
* Orders
* Order\_details
* Products

All four tables were created and imported into MySQL using **phpMyAdmin**. The data was checked for completeness and structural consistency.

## **Data Cleaning & Preparation:**

* Duplicates were removed from each CSV file **before** importing into MySQL.
* Table contents were verified using SELECT \* FROM table\_name queries.
* Each table was reviewed to ensure appropriate data types for each column (e.g., integers for quantity, decimals for sales, dates for order\_date).
* Encountered a MySQL error due to inserting column headers as values, which was fixed by removing the header row before importing.

## **Key Questions:**

1. **Display the number of rows in each of the 4 tables.**

### **Query Used:**

SELECT COUNT(\*) AS total\_customers FROM customers;

SELECT COUNT(\*) AS total\_products FROM products;

SELECT COUNT(\*) AS total\_orders FROM orders;

SELECT COUNT(\*) AS total\_order\_lines FROM order\_details;

|  |  |
| --- | --- |
| Table Name | Rows |
| customers | 794 |
| products | 1863 |
| orders | 5009 |
| order\_details | 9986 |

1. **Display the Total Sales and Total Profit across all orders.**

### **Query:**

SELECT

ROUND(SUM(sales),2) AS total\_sales,

ROUND(SUM(profit),2) AS total\_profit,

FROM order\_details;

|  |  |
| --- | --- |
| **total\_sales** | **total\_profit** |
| 2296919.70 | 286409.85 |

1. **Display the Overall Profit Margin (%).**

### **Query:**

SELECT

ROUND((SUM(profit) / SUM(sales)) \* 100, 2) AS profit\_margin\_percent

FROM order\_details;

|  |
| --- |
| profit\_margin\_percent |
| 12.47 |

1. **Display the top 10 Customers by Total Sales.**

### **Query:**

SELECT

c.customer\_id,

c.customer\_name,

ROUND(SUM(od.sales), 2) AS total\_sales

FROM customers c

JOIN orders o ON o.customer\_id = c.customer\_id

JOIN order\_details od ON od.order\_id = o.order\_id

GROUP BY c.customer\_id, c.customer\_name

ORDER BY total\_sales DESC

LIMIT 10;

|  |  |  |
| --- | --- | --- |
| [**customer\_id**](http://localhost/phpMyAdmin5/index.php?route=/sql&db=superstore_db&table=customers&sql_query=SELECT%0D%0A%09c.customer_id%2C%0D%0A++++c.customer_name%2C%0D%0A%09ROUND%28SUM%28od.sales%29%2C+2%29+AS+total_sales%0D%0AFROM+customers+c%0D%0AJOIN+orders+o+ON+o.customer_id+%3D+c.customer_id%0D%0AJOIN+order_details+od+ON+od.order_id+%3D+o.order_id%0D%0AGROUP+BY+c.customer_id%2C+c.customer_name+%0AORDER+BY+%60c%60.%60customer_id%60+ASC+LIMIT+10&sql_signature=bafa2b2ad1212efd18f0b9c4a0fef69d694d82b9a109a9d91eb85b38fc74e90e&session_max_rows=25&is_browse_distinct=0) | [**customer\_name**](http://localhost/phpMyAdmin5/index.php?route=/sql&db=superstore_db&table=customers&sql_query=SELECT%0D%0A%09c.customer_id%2C%0D%0A++++c.customer_name%2C%0D%0A%09ROUND%28SUM%28od.sales%29%2C+2%29+AS+total_sales%0D%0AFROM+customers+c%0D%0AJOIN+orders+o+ON+o.customer_id+%3D+c.customer_id%0D%0AJOIN+order_details+od+ON+od.order_id+%3D+o.order_id%0D%0AGROUP+BY+c.customer_id%2C+c.customer_name+%0AORDER+BY+%60c%60.%60customer_name%60+ASC+LIMIT+10&sql_signature=074616c55c87e9cd01484faac0eefe76f0b962ed4c04607e4893511c2184048f&session_max_rows=25&is_browse_distinct=0) | [**total\_sales**](http://localhost/phpMyAdmin5/index.php?route=/sql&db=superstore_db&table=customers&sql_query=SELECT%0D%0A%09c.customer_id%2C%0D%0A++++c.customer_name%2C%0D%0A%09ROUND%28SUM%28od.sales%29%2C+2%29+AS+total_sales%0D%0AFROM+customers+c%0D%0AJOIN+orders+o+ON+o.customer_id+%3D+c.customer_id%0D%0AJOIN+order_details+od+ON+od.order_id+%3D+o.order_id%0D%0AGROUP+BY+c.customer_id%2C+c.customer_name+%0AORDER+BY+%60total_sales%60++ASC+LIMIT+10&sql_signature=4595548c1f49db33db0951f27521bc66ec58f2e19397e79fc01cbfcc07127ef1&session_max_rows=25&is_browse_distinct=0) |
| SM-20320 | Sean Miller | 25043.07 |
| TC-20980 | Tamara Chand | 19052.22 |
| RB-19360 | Raymond Buch | 15117.35 |
| TA-21385 | Tom Ashbrook | 14595.62 |
| AB-10105 | Adrian Barton | 14473.57 |
| KL-16645 | Ken Lonsdale | 14175.23 |
| SC-20095 | Sanjit Chand | 14142.34 |
| HL-15040 | Hunter Lopez | 12873.30 |
| SE-20110 | Sanjit Engle | 12209.44 |
| CC-12370 | Christopher Conant | 12129.0 |

1. **Display top 10 Products by Quantity Sold.**

### **Query:**

SELECT

p.product\_id,

p.product\_name,

SUM(od.quantity) AS total\_quantity\_sold

FROM products p

JOIN order\_details od ON od.product\_id = p.product\_id

GROUP BY product\_id, product\_name

ORDER BY total\_quantity\_sold DESC

LIMIT 10;

|  |  |  |
| --- | --- | --- |
| [**product\_id**](http://localhost/phpMyAdmin5/index.php?route=/sql&db=superstore_db&table=products&sql_query=SELECT%0D%0A%09p.product_id%2C%0D%0A++++p.product_name%2C%0D%0A++++SUM%28od.quantity%29+AS+total_quantity_sold%0D%0AFROM+products+p%0D%0AJOIN+order_details+od+ON+od.product_id++%3D+p.product_id%0D%0AGROUP+BY+product_id%2C+product_name+%0AORDER+BY+%60p%60.%60product_id%60+ASC+LIMIT+10&sql_signature=4ef4bc663c553a428e95a73d202d772469735fb15b4ae61fb5f8a829ef0730f8&session_max_rows=25&is_browse_distinct=0) | [**product\_name**](http://localhost/phpMyAdmin5/index.php?route=/sql&db=superstore_db&table=products&sql_query=SELECT%0D%0A%09p.product_id%2C%0D%0A++++p.product_name%2C%0D%0A++++SUM%28od.quantity%29+AS+total_quantity_sold%0D%0AFROM+products+p%0D%0AJOIN+order_details+od+ON+od.product_id++%3D+p.product_id%0D%0AGROUP+BY+product_id%2C+product_name+%0AORDER+BY+%60p%60.%60product_name%60+ASC+LIMIT+10&sql_signature=55f9e2b04cf494dcb755963b04751438900fb2c5f7a9507fa68ffbc83b8d1973&session_max_rows=25&is_browse_distinct=0) | [**total\_quantity\_sold**](http://localhost/phpMyAdmin5/index.php?route=/sql&db=superstore_db&table=products&sql_query=SELECT%0D%0A%09p.product_id%2C%0D%0A++++p.product_name%2C%0D%0A++++SUM%28od.quantity%29+AS+total_quantity_sold%0D%0AFROM+products+p%0D%0AJOIN+order_details+od+ON+od.product_id++%3D+p.product_id%0D%0AGROUP+BY+product_id%2C+product_name+%0AORDER+BY+%60total_quantity_sold%60++ASC+LIMIT+10&sql_signature=100b1d0a5c688ac85c65b859e7d4b172dd9890c391ae1274e28f42948c18efb2&session_max_rows=25&is_browse_distinct=0) |
| TEC-AC-10003832 | Technology | 75 |
| OFF-PA-10001970 | Office Supplies | 70 |
| OFF-BI-10001524 | Office Supplies | 67 |
| FUR-CH-10002647 | Furniture | 64 |
| OFF-BI-10002026 | Office Supplies | 64 |
| FUR-TA-10001095 | Furniture | 61 |
| TEC-AC-10002049 | Technology | 60 |
| OFF-BI-10004728 | Office Supplies | 59 |
| FUR-CH-10003774 | Furniture | 59 |
| FUR-FU-10001473 | Furniture | 57 |

1. **Display top 10 products by Sales.**

### **Query:**

SELECT

p.product\_id,

p.product\_name,

SUM(od.sales) AS total\_sales

FROM products p

JOIN order\_details od ON od.product\_id = p.product\_id

GROUP BY product\_id, product\_name

ORDER BY total\_sales DESC

LIMIT 10;

|  |  |  |
| --- | --- | --- |
| [**product\_id**](http://localhost/phpMyAdmin5/index.php?route=/sql&db=superstore_db&table=products&sql_query=SELECT%0D%0A%09p.product_id%2C%0D%0A%09p.product_name%2C%0D%0ASUM%28od.sales%29+AS+total_sales%0D%0AFROM+products+p%0D%0AJOIN+order_details+od+ON+od.product_id++%3D+p.product_id%0D%0AGROUP+BY+product_id%2C+product_name+%0AORDER+BY+%60p%60.%60product_id%60+ASC+LIMIT+10&sql_signature=6b1ca9102c5b780f773eff0edaba0b1617f3a35f68aa2d404b6103d9814563a3&session_max_rows=25&is_browse_distinct=0) | [**product\_name**](http://localhost/phpMyAdmin5/index.php?route=/sql&db=superstore_db&table=products&sql_query=SELECT%0D%0A%09p.product_id%2C%0D%0A%09p.product_name%2C%0D%0ASUM%28od.sales%29+AS+total_sales%0D%0AFROM+products+p%0D%0AJOIN+order_details+od+ON+od.product_id++%3D+p.product_id%0D%0AGROUP+BY+product_id%2C+product_name+%0AORDER+BY+%60p%60.%60product_name%60+ASC+LIMIT+10&sql_signature=7c73e7b05475725bb85849d8807fd576fc81e3fcca5206fc567c14ce32f9569f&session_max_rows=25&is_browse_distinct=0) | [**total\_sales**](http://localhost/phpMyAdmin5/index.php?route=/sql&db=superstore_db&table=products&sql_query=SELECT%0D%0A%09p.product_id%2C%0D%0A%09p.product_name%2C%0D%0ASUM%28od.sales%29+AS+total_sales%0D%0AFROM+products+p%0D%0AJOIN+order_details+od+ON+od.product_id++%3D+p.product_id%0D%0AGROUP+BY+product_id%2C+product_name+%0AORDER+BY+%60total_sales%60++ASC+LIMIT+10&sql_signature=882389a2af1e005abb8758e7c6ff44547b2c5277b903bf636f8c22231af67a4c&session_max_rows=25&is_browse_distinct=0) |
| TEC-CO-10004722 | Technology | 61599.83 |
| OFF-BI-10003527 | Office Supplies | 27453.38 |
| TEC-MA-10002412 | Technology | 22638.48 |
| FUR-CH-10002024 | Furniture | 21870.57 |
| OFF-BI-10001359 | Office Supplies | 19823.48 |
| OFF-BI-10000545 | Office Supplies | 19024.50 |
| TEC-CO-10001449 | Technology | 18839.68 |
| TEC-MA-10001127 | Technology | 18374.90 |
| OFF-BI-10004995 | Office Supplies | 17965.07 |
| OFF-SU-10000151 | Office Supplies | 17030.3 |

1. **Which product categories are most profitable?**

### **Query:**

SELECT

p.category,

ROUND(SUM(od.profit),2) AS total\_profit

FROM products p

JOIN order\_details od ON od.product\_id = p.product\_id

GROUP BY p.category

ORDER BY total\_profit DESC;

|  |  |
| --- | --- |
| [**category**](http://localhost/phpMyAdmin5/index.php?route=/sql&db=superstore_db&table=products&sql_query=SELECT%0D%0A%09p.category%2C%0D%0A++++ROUND%28SUM%28od.profit%29%2C2%29+AS+total_profit%0D%0AFROM+products+p%0D%0AJOIN+order_details+od+ON+od.product_id%09+%3D+p.product_id%0D%0AGROUP+BY+p.category++%0AORDER+BY+%60p%60.%60category%60+ASC&sql_signature=48192c78520f3275cc047b129ffb5d194458e94211871b3be37e0c9c6acfada9&session_max_rows=25&is_browse_distinct=0) | [**total\_profit**](http://localhost/phpMyAdmin5/index.php?route=/sql&db=superstore_db&table=products&sql_query=SELECT%0D%0A%09p.category%2C%0D%0A++++ROUND%28SUM%28od.profit%29%2C2%29+AS+total_profit%0D%0AFROM+products+p%0D%0AJOIN+order_details+od+ON+od.product_id%09+%3D+p.product_id%0D%0AGROUP+BY+p.category++%0AORDER+BY+%60total_profit%60++ASC&sql_signature=7f0da98a70514fb6c8fade4a7478bbda726d89f937521fca7d7f666a3b976bfb&session_max_rows=25&is_browse_distinct=0) |
| Copiers | 55617.90 |
| Phones | 44516.25 |
| Accessories | 41936.78 |
| Paper | 34053.34 |
| Binders | 30221.64 |
| Chairs | 26602.21 |
| Storage | 21279.05 |
| Appliances | 18138.07 |
| Furnishings | 13059.25 |
| Envelopes | 6964.10 |
| Art | 6527.96 |
| Labels | 5546.18 |
| Machines | 3384.73 |
| Fasteners | 949.53 |
| Supplies | -1188.99 |
| Bookcases | -3472.56 |
| Tables | -17725.5 |

The total\_profit for Supplies, Bookcases and Tables are Negative. Investigating further.

Question: Check Average Discount on Category.

### **Query:**

SELECT

p.category,

ROUND(AVG(od.discount), 2) AS avg\_discount,

ROUND(SUM(od.profit), 2) AS total\_profit

FROM products p

JOIN order\_details od ON p.product\_id = od.product\_id

GROUP BY p.category

ORDER BY avg\_discount DESC;

|  |  |  |
| --- | --- | --- |
| [**avg\_discount**](http://localhost/phpMyAdmin5/index.php?route=/sql&db=superstore_db&table=products&sql_query=SELECT+%0D%0A++++p.category%2C%0D%0A++++ROUND%28AVG%28od.discount%29%2C+2%29+AS+avg_discount%2C%0D%0A++++ROUND%28SUM%28od.profit%29%2C+2%29+AS+total_profit%0D%0AFROM+products+p%0D%0AJOIN+order_details+od+ON+p.product_id+%3D+od.product_id%0D%0AGROUP+BY+p.category++%0AORDER+BY+%60avg_discount%60++ASC&sql_signature=295e351b90974a785f9b94e4af00573f5a7dbdd7b74fe06a337f1836060e9d27&session_max_rows=25&is_browse_distinct=0) | [**total\_profit**](http://localhost/phpMyAdmin5/index.php?route=/sql&db=superstore_db&table=products&sql_query=SELECT+%0D%0A++++p.category%2C%0D%0A++++ROUND%28AVG%28od.discount%29%2C+2%29+AS+avg_discount%2C%0D%0A++++ROUND%28SUM%28od.profit%29%2C+2%29+AS+total_profit%0D%0AFROM+products+p%0D%0AJOIN+order_details+od+ON+p.product_id+%3D+od.product_id%0D%0AGROUP+BY+p.category++%0AORDER+BY+%60total_profit%60+ASC&sql_signature=f4a8c7d79d6cbf3d1c01965edbb8a6f0306dc1d5d923f0c16a5005eee9c87b30&session_max_rows=25&is_browse_distinct=0) | [**category**](http://localhost/phpMyAdmin5/index.php?route=/sql&db=superstore_db&table=products&sql_query=SELECT+%0D%0A++++p.category%2C%0D%0A++++ROUND%28AVG%28od.discount%29%2C+2%29+AS+avg_discount%2C%0D%0A++++ROUND%28SUM%28od.profit%29%2C+2%29+AS+total_profit%0D%0AFROM+products+p%0D%0AJOIN+order_details+od+ON+p.product_id+%3D+od.product_id%0D%0AGROUP+BY+p.category++%0AORDER+BY+%60p%60.%60category%60+ASC&sql_signature=c793201cb405145c90e8e18fb1d09c5f80add26a9e0a3263ad70eb2d5f1076b9&session_max_rows=25&is_browse_distinct=0) |
| 0.37 | 30221.64 | Binders |
| 0.31 | 3384.73 | Machines |
| 0.26 | -17725.59 | Tables |
| 0.21 | -3472.56 | Bookcases |
| 0.17 | 26602.21 | Chairs |
| 0.17 | 18138.07 | Appliances |
| 0.16 | 55617.90 | Copiers |
| 0.15 | 44516.25 | Phones |
| 0.14 | 13059.25 | Furnishings |
| 0.08 | -1188.99 | Supplies |
| 0.08 | 6964.10 | Envelopes |
| 0.08 | 34053.34 | Paper |
| 0.08 | 41936.78 | Accessories |
| 0.08 | 949.53 | Fasteners |
| 0.07 | 5546.18 | Labels |
| 0.07 | 6527.96 | Art |
| 0.07 | 21279.05 | Storage |

The investigation reveals that the negative profit may be due to over-discounted products, lower margin or high shipping/handling costs. Investigating further at subcategory level.

### **Query:**

SELECT

p.sub\_category,

ROUND(AVG(od.discount), 2) AS avg\_discount,

ROUND(SUM(od.profit), 2) AS total\_profit

FROM products p

JOIN order\_details od ON p.product\_id = od.product\_id

GROUP BY p.sub\_category

ORDER BY avg\_discount DESC;

|  |  |  |
| --- | --- | --- |
| [**avg\_discount**](http://localhost/phpMyAdmin5/index.php?route=/sql&db=superstore_db&table=products&sql_query=SELECT+%0D%0A++++p.sub_category%2C%0D%0A++++ROUND%28AVG%28od.discount%29%2C+2%29+AS+avg_discount%2C%0D%0A++++ROUND%28SUM%28od.profit%29%2C+2%29+AS+total_profit%0D%0AFROM+products+p%0D%0AJOIN+order_details+od+ON+p.product_id+%3D+od.product_id%0D%0AGROUP+BY+p.sub_category++%0AORDER+BY+%60avg_discount%60++ASC&sql_signature=e538c966b8632cffe457b39d78fcce22d8ffe8347ab9c76d23d8af89d65431a5&session_max_rows=25&is_browse_distinct=0) | [**total\_profit**](http://localhost/phpMyAdmin5/index.php?route=/sql&db=superstore_db&table=products&sql_query=SELECT+%0D%0A++++p.sub_category%2C%0D%0A++++ROUND%28AVG%28od.discount%29%2C+2%29+AS+avg_discount%2C%0D%0A++++ROUND%28SUM%28od.profit%29%2C+2%29+AS+total_profit%0D%0AFROM+products+p%0D%0AJOIN+order_details+od+ON+p.product_id+%3D+od.product_id%0D%0AGROUP+BY+p.sub_category++%0AORDER+BY+%60total_profit%60+ASC&sql_signature=0b012c2404bd3d377ab968be041d6fd4822c61889efa433d719dc6c0407e5643&session_max_rows=25&is_browse_distinct=0) | [**sub\_category**](http://localhost/phpMyAdmin5/index.php?route=/sql&db=superstore_db&table=products&sql_query=SELECT+%0D%0A++++p.sub_category%2C%0D%0A++++ROUND%28AVG%28od.discount%29%2C+2%29+AS+avg_discount%2C%0D%0A++++ROUND%28SUM%28od.profit%29%2C+2%29+AS+total_profit%0D%0AFROM+products+p%0D%0AJOIN+order_details+od+ON+p.product_id+%3D+od.product_id%0D%0AGROUP+BY+p.sub_category++%0AORDER+BY+%60p%60.%60sub_category%60+ASC&sql_signature=6e4f0c8f9968d73a9f99fc4f8ae3a7091c90b960cc441a1f152711bc977807c3&session_max_rows=25&is_browse_distinct=0) |
| 0.80 | -4.47 | Eureka Disposable Bags for Sanitaire Vibra Groomer... |
| 0.73 | -68.43 | GBC Plasticlear Binding Covers |
| 0.73 | -411.33 | GBC VeloBinder Electric Binding Machine |
| 0.70 | -319.19 | Brother MFC-9340CDW LED All-In-One Printer, Copier... |
| 0.70 | -164.95 | Cisco 8961 IP Phone Charcoal |
| 0.70 | -938.28 | Zebra GK420t Direct Thermal/Thermal Transfer Print... |
| 0.70 | -21.95 | Plantronics Single Ear Headset |
| 0.70 | -172.49 | Epson Perfection V600 Photo Scanner |
| 0.70 | -71.40 | Hewlett-Packard Deskjet F4180 All-in-One Color Ink... |
| 0.70 | -251.99 | Okidata B401 Printer |
| 0.70 | -190.85 | Bush Westfield Collection Bookcases, Dark Cherry F... |
| 0.70 | -506.46 | Lexmark MarkNet N8150 Wireless Print Server |
| 0.64 | -201.51 | Premier Elliptical Ring Binder, Black |
| 0.63 | -19.69 | Wilson Jones Custom Binder Spines & Labels |
| 0.62 | -102.84 | Acco D-Ring Binder w/DublLock |
| 0.62 | -4.93 | Avery Triangle Shaped Sheet Lifters, Black, 2/Pack |
| 0.62 | -23.76 | Avery Durable Slant Ring Binders |

Products with average discounts over 60% are consistently generating negative profit. In particular, small electronics and office binding supplies show the steepest losses, suggesting either excessive markdowns or poor product cost/pricing configurations. These SKUs should be reviewed for discount eligibility, cost price accuracy, and relevance in inventory.

1. **Average Order Value per Customer (AOV)**

### **Query:**

SELECT

c.customer\_id,

c.customer\_name,

COUNT(DISTINCT o.order\_id) AS total\_orders,

ROUND(SUM(od.sales) / COUNT(DISTINCT o.order\_id), 2) AS avg\_order\_value

FROM customers c

JOIN orders o ON c.customer\_id = o.customer\_id

JOIN order\_details od ON o.order\_id = od.order\_id

GROUP BY c.customer\_id, c.customer\_name

ORDER BY avg\_order\_value DESC

LIMIT 10;

|  |  |  |  |
| --- | --- | --- | --- |
| [**customer\_id**](http://localhost/phpMyAdmin5/index.php?route=/sql&db=superstore_db&table=customers&sql_query=SELECT+%0D%0A++++c.customer_id%2C%0D%0A++++c.customer_name%2C%0D%0A++++COUNT%28DISTINCT+o.order_id%29+AS+total_orders%2C%0D%0A++++ROUND%28SUM%28od.sales%29+%2F+COUNT%28DISTINCT+o.order_id%29%2C+2%29+AS+avg_order_value%0D%0AFROM+customers+c%0D%0AJOIN+orders+o+ON+c.customer_id+%3D+o.customer_id%0D%0AJOIN+order_details+od+ON+o.order_id+%3D+od.order_id%0D%0AGROUP+BY+c.customer_id%2C+c.customer_name+%0AORDER+BY+%60c%60.%60customer_id%60+ASC+LIMIT+10&sql_signature=ce1058995eaa872c84fd6de81a4442d308f58758807bf725bb021b5cbb3334ab&session_max_rows=25&is_browse_distinct=0) | [**customer\_name**](http://localhost/phpMyAdmin5/index.php?route=/sql&db=superstore_db&table=customers&sql_query=SELECT+%0D%0A++++c.customer_id%2C%0D%0A++++c.customer_name%2C%0D%0A++++COUNT%28DISTINCT+o.order_id%29+AS+total_orders%2C%0D%0A++++ROUND%28SUM%28od.sales%29+%2F+COUNT%28DISTINCT+o.order_id%29%2C+2%29+AS+avg_order_value%0D%0AFROM+customers+c%0D%0AJOIN+orders+o+ON+c.customer_id+%3D+o.customer_id%0D%0AJOIN+order_details+od+ON+o.order_id+%3D+od.order_id%0D%0AGROUP+BY+c.customer_id%2C+c.customer_name+%0AORDER+BY+%60c%60.%60customer_name%60+ASC+LIMIT+10&sql_signature=92b5887df3b6f3e90da534757bbca9cebc09922ce7f13200727e238a34029717&session_max_rows=25&is_browse_distinct=0) | [**total\_orders**](http://localhost/phpMyAdmin5/index.php?route=/sql&db=superstore_db&table=customers&sql_query=SELECT+%0D%0A++++c.customer_id%2C%0D%0A++++c.customer_name%2C%0D%0A++++COUNT%28DISTINCT+o.order_id%29+AS+total_orders%2C%0D%0A++++ROUND%28SUM%28od.sales%29+%2F+COUNT%28DISTINCT+o.order_id%29%2C+2%29+AS+avg_order_value%0D%0AFROM+customers+c%0D%0AJOIN+orders+o+ON+c.customer_id+%3D+o.customer_id%0D%0AJOIN+order_details+od+ON+o.order_id+%3D+od.order_id%0D%0AGROUP+BY+c.customer_id%2C+c.customer_name+%0AORDER+BY+%60total_orders%60+ASC+LIMIT+10&sql_signature=5dda52041b86ce43ca0b930dfb4da529857c8b13ecb9a453d1067a61966135a6&session_max_rows=25&is_browse_distinct=0) | [**avg\_order\_value**](http://localhost/phpMyAdmin5/index.php?route=/sql&db=superstore_db&table=customers&sql_query=SELECT+%0D%0A++++c.customer_id%2C%0D%0A++++c.customer_name%2C%0D%0A++++COUNT%28DISTINCT+o.order_id%29+AS+total_orders%2C%0D%0A++++ROUND%28SUM%28od.sales%29+%2F+COUNT%28DISTINCT+o.order_id%29%2C+2%29+AS+avg_order_value%0D%0AFROM+customers+c%0D%0AJOIN+orders+o+ON+c.customer_id+%3D+o.customer_id%0D%0AJOIN+order_details+od+ON+o.order_id+%3D+od.order_id%0D%0AGROUP+BY+c.customer_id%2C+c.customer_name+%0AORDER+BY+%60avg_order_value%60++ASC+LIMIT+10&sql_signature=cfaa54301d7c4e1ee76dd1af74c3dfb1a1793daf19b9ad7bc8ab70cb5d4f9e23&session_max_rows=25&is_browse_distinct=0) |
| SM-20320 | Sean Miller | 5 | 5008.61 |
| TC-20980 | Tamara Chand | 5 | 3810.44 |
| TA-21385 | Tom Ashbrook | 4 | 3648.91 |
| GT-14635 | Grant Thornton | 3 | 3117.07 |
| BM-11140 | Becky Martin | 4 | 2947.41 |
| MW-18235 | Mitch Willingham | 2 | 2626.94 |
| RB-19360 | Raymond Buch | 6 | 2519.56 |
| CC-12370 | Christopher Conant | 5 | 2425.82 |
| PF-19120 | Peter Fuller | 4 | 2265.72 |
| CM-12385 | Christopher Martinez | 4 | 2238.50 |

1. **Customer Lifetime Value (CLTV)**

### **Query:**

SELECT

c.customer\_id,

c.customer\_name,

ROUND(SUM(od.sales), 2) AS lifetime\_value

FROM customers c

JOIN orders o ON c.customer\_id = o.customer\_id

JOIN order\_details od ON o.order\_id = od.order\_id

GROUP BY c.customer\_id, c.customer\_name

ORDER BY lifetime\_value DESC

LIMIT 10;

|  |  |  |
| --- | --- | --- |
| [**customer\_id**](http://localhost/phpMyAdmin5/index.php?route=/sql&db=superstore_db&table=customers&sql_query=SELECT+%0D%0A++++c.customer_id%2C%0D%0A++++c.customer_name%2C%0D%0A++++ROUND%28SUM%28od.sales%29%2C+2%29+AS+lifetime_value%0D%0AFROM+customers+c%0D%0AJOIN+orders+o+ON+c.customer_id+%3D+o.customer_id%0D%0AJOIN+order_details+od+ON+o.order_id+%3D+od.order_id%0D%0AGROUP+BY+c.customer_id%2C+c.customer_name+%0AORDER+BY+%60c%60.%60customer_id%60+ASC+LIMIT+10&sql_signature=a7d68304faee4915a87f941d3a6fdc4d6458594a8bede015d446445423dde6eb&session_max_rows=25&is_browse_distinct=0) | [**customer\_name**](http://localhost/phpMyAdmin5/index.php?route=/sql&db=superstore_db&table=customers&sql_query=SELECT+%0D%0A++++c.customer_id%2C%0D%0A++++c.customer_name%2C%0D%0A++++ROUND%28SUM%28od.sales%29%2C+2%29+AS+lifetime_value%0D%0AFROM+customers+c%0D%0AJOIN+orders+o+ON+c.customer_id+%3D+o.customer_id%0D%0AJOIN+order_details+od+ON+o.order_id+%3D+od.order_id%0D%0AGROUP+BY+c.customer_id%2C+c.customer_name+%0AORDER+BY+%60c%60.%60customer_name%60+ASC+LIMIT+10&sql_signature=03f7cf87bf9fa770910ff58aed9eaf909bfeeec3fa99460c33597ba8ec2ba282&session_max_rows=25&is_browse_distinct=0) | [**lifetime\_value**](http://localhost/phpMyAdmin5/index.php?route=/sql&db=superstore_db&table=customers&sql_query=SELECT+%0D%0A++++c.customer_id%2C%0D%0A++++c.customer_name%2C%0D%0A++++ROUND%28SUM%28od.sales%29%2C+2%29+AS+lifetime_value%0D%0AFROM+customers+c%0D%0AJOIN+orders+o+ON+c.customer_id+%3D+o.customer_id%0D%0AJOIN+order_details+od+ON+o.order_id+%3D+od.order_id%0D%0AGROUP+BY+c.customer_id%2C+c.customer_name+%0AORDER+BY+%60lifetime_value%60++ASC+LIMIT+10&sql_signature=e96a568aa47e41b29001e1c9cb245498e29c05e58630b6ab0f3ff46f22989063&session_max_rows=25&is_browse_distinct=0) |
| SM-20320 | Sean Miller | 25043.07 |
| TC-20980 | Tamara Chand | 19052.22 |
| RB-19360 | Raymond Buch | 15117.35 |
| TA-21385 | Tom Ashbrook | 14595.62 |
| AB-10105 | Adrian Barton | 14473.57 |
| KL-16645 | Ken Lonsdale | 14175.23 |
| SC-20095 | Sanjit Chand | 14142.34 |
| HL-15040 | Hunter Lopez | 12873.30 |
| SE-20110 | Sanjit Engle | 12209.44 |
| CC-12370 | Christopher Conant | 12129.08 |

1. **Display Profit Margin by Category.**

### **Query:**

SELECT

p.category,

ROUND(SUM(od.profit), 2) AS total\_profit,

ROUND(SUM(od.profit) / SUM(od.sales), 2) AS profit\_margin\_percent

FROM products p

JOIN order\_details od ON od.product\_id = p.product\_id

GROUP BY p.category

ORDER BY profit\_margin\_percent DESC;

|  |  |  |
| --- | --- | --- |
| [**total\_profit**](http://localhost/phpMyAdmin5/index.php?route=/sql&db=superstore_db&table=products&sql_query=SELECT%0D%0A%09p.category%2C%0D%0A++++ROUND%28SUM%28od.profit%29%2C+2%29+AS+total_profit%2C%0D%0A++++ROUND%28SUM%28od.profit%29+%2F+SUM%28od.sales%29%2C+2%29+AS+profit_margin_percent%0D%0AFROM+products+p%0D%0AJOIN+order_details+od+ON+od.product_id+%3D+p.product_id%0D%0AGROUP+BY+p.category++%0AORDER+BY+%60total_profit%60+ASC&sql_signature=51cd5e5e882fac39e001126d091e9139a418b37d8439dff91e9636c1b0b9979c&session_max_rows=50&is_browse_distinct=0) | [**profit\_margin\_percent**](http://localhost/phpMyAdmin5/index.php?route=/sql&db=superstore_db&table=products&sql_query=SELECT%0D%0A%09p.category%2C%0D%0A++++ROUND%28SUM%28od.profit%29%2C+2%29+AS+total_profit%2C%0D%0A++++ROUND%28SUM%28od.profit%29+%2F+SUM%28od.sales%29%2C+2%29+AS+profit_margin_percent%0D%0AFROM+products+p%0D%0AJOIN+order_details+od+ON+od.product_id+%3D+p.product_id%0D%0AGROUP+BY+p.category++%0AORDER+BY+%60profit_margin_percent%60++ASC&sql_signature=8bbc3e9cffe9b0ef85a6df05cfd0758760d0b77ceee2a246c6e211b5e1d2d9d3&session_max_rows=50&is_browse_distinct=0) | [**category**](http://localhost/phpMyAdmin5/index.php?route=/sql&db=superstore_db&table=products&sql_query=SELECT%0D%0A%09p.category%2C%0D%0A++++ROUND%28SUM%28od.profit%29%2C+2%29+AS+total_profit%2C%0D%0A++++ROUND%28SUM%28od.profit%29+%2F+SUM%28od.sales%29%2C+2%29+AS+profit_margin_percent%0D%0AFROM+products+p%0D%0AJOIN+order_details+od+ON+od.product_id+%3D+p.product_id%0D%0AGROUP+BY+p.category++%0AORDER+BY+%60p%60.%60category%60+ASC&sql_signature=64c7a8a26a61c9bd8ef45ddf44509df1070639488df736573e7478a709309d07&session_max_rows=50&is_browse_distinct=0) |
| 5546.18 | 0.44 | Labels |
| 34053.34 | 0.43 | Paper |
| 6964.10 | 0.42 | Envelopes |
| 55617.90 | 0.37 | Copiers |
| 949.53 | 0.31 | Fasteners |
| 41936.78 | 0.25 | Accessories |
| 6527.96 | 0.24 | Art |
| 18138.07 | 0.17 | Appliances |
| 30221.64 | 0.15 | Binders |
| 13059.25 | 0.14 | Furnishings |
| 44516.25 | 0.13 | Phones |
| 21279.05 | 0.10 | Storage |
| 26602.21 | 0.08 | Chairs |
| 3384.73 | 0.02 | Machines |
| -3472.56 | -0.03 | Bookcases |
| -1188.99 | -0.03 | Supplies |
| -17725.59 | -0.09 | Tables |

Moving to the next phase of the project.

# **Data Visualisation in Power BI:**

## **Objective:** To visualise key business metrics and trends using Power BI, enabling stakeholders to make data-driven decisions. The goal is to transform raw SQL outputs into interactive, insightful dashboards.

**Data Connection:**

**Approach:**  
The MySQL database was connected to Power BI to fetch live data.

**Options Considered:**

* **Import** mode: Good for performance and static snapshots
* **DirectQuery** mode: Best for real-time updates but slower

**Decision:**  
Used **Import** mode for better performance during initial dashboard creation.

**Tables Imported:**

From the superstore\_db MySQL database, the following 4 tables were imported:

* Customers
* Orders
* Order\_details
* Products

Power BI’s *Transform Data* (Power Query Editor) was used to inspect, rename columns, and ensure proper relationships.

## **Data Modelling in Power BI:**

**Relationships Defined:**

* Customers[CustomerID] → Orders[CustomerID]
* Orders[OrderID] → Order\_details[OrderID]
* Order\_details[ProductID] → Products[ProductID]

**Cardinality & Join Type:**

* All relationships were **One-to-Many**, with **Single Direction** filtering.

## **Measures Created Using DAX**

To support visualisations, the following key measures were created using DAX:

**DAX:**

Total Sales = SUM(order\_details[Sales])

Total Profit = SUM(order\_details[Profit])

Profit Margin % = DIVIDE(SUM(order\_details[Profit]), SUM(order\_details[Sales])) \* 100’

These measures were added to support performance KPIs and trend analysis.

## **Visuals Created (So Far)**

**1. KPI Cards**Purpose: Quickly show the high-level business metrics  
Visual Type: Card  
Cards Created:

* Total Sales
* Total Profit
* Profit Margin %

**2. Line Chart – Total Sales and Profit by Order Month**Visual Type: Line Chart  
Axes:

* X-axis: Order Month (converted to a Date Hierarchy or formatted Month-Year)
* Y-axis: Total Sales and Total Profit (two measures)  
  Purpose: To show the time series trend in sales and profit.

**3. Bar Chart – Customer Sales by Customer Name**Visual Type: Bar Chart  
Axes:

* X-axis: Customer Sales (measure: SUM(order\_details[Sales]))
* Y-axis: Customer Name  
  Purpose: Identify top customers by total sales.

**4. Bar Chart – Product Quantity Sold by Product Name**Visual Type: Bar Chart  
Axes:

* X-axis: Product Quantity Sold (measure: SUM(order\_details[Quantity]))
* Y-axis: Product Name  
  Purpose: Find top-selling products by quantity.

**5. Matrix Table – Product Profit by Category**Visual Type: Table  
Rows: Product Name  
Columns: Category  
Values: Total Profit  
Purpose: Drill-down into profit performance by product within each category.

**6. Slicers for Interactivity**Visual Type: Slicer  
Fields Added:

* Region
* Segment
* Order Month

These allow users to filter dashboard views based on business segment, region, or time frame.

**7. Bar Chart – Profit Margin % by Product Category**Visual Type: Bar Chart  
Axes:

* X-axis: Product Category
* Y-axis: Profit Margin % (DAX measure: Profit Margin % = DIVIDE(SUM(order\_details[Profit]), SUM(order\_details[Sales])) \* 100)  
  Purpose:  
  To compare how efficiently each product category converts sales into profit. This helps identify high-margin categories worth prioritising in promotions and marketing.

**8. Table – Product Name, Category, and Total Profit**Visual Type: Table  
Fields Added:

* Product Name
* Category
* Total Profit (DAX measure)  
  Purpose:  
  To list individual products under their respective categories and rank them by profitability. This detailed breakdown supports product-level profit optimisation decisions.

**9. Bar Chart – Average Shipping Delay by Ship Mode**Visual Type: Bar Chart  
DAX Measure Created:

DAX

CopyEdit

Average Shipping Delay = AVERAGE(DATEDIFF(orders[Order Date], orders[Ship Date], DAY))

Axes:

* X-axis: Ship Mode
* Y-axis: Average Shipping Delay (in days)  
  Purpose:  
  To evaluate the delivery performance of different shipping methods. A high average delay may indicate inefficiencies or customer dissatisfaction risk.

**10. Info Button (“i”) for Insights Panel**  
**Visual Type:** Shape (Icon – “Information” or custom image)  
**Feature Used:** Bookmark + Selection Pane + Tooltip Page

**Purpose:**  
To provide a clean, on-demand insights panel for users without cluttering the main dashboard. Clicking the “i” icon displays a summary of key insights and recommendations based on the data.

**Implementation Steps:**

* Created a separate **insight panel** with text boxes listing key findings.
* Used **Bookmarks** to toggle visibility of this panel.
* Used **Selection Pane** to show/hide visuals accordingly.
* Linked the icon to the bookmark using the **Action** property (Type: Bookmark → Target: “Insights View”).

**11. Back Button from Insights Panel**  
**Visual Type:** Shape (Back arrow icon)  
**Feature Used:** Bookmark + Action

**Purpose:**  
To allow users to return from the insights panel back to the main dashboard seamlessly.

**Implementation Steps:**

* Created a **Back** button shape on the insights panel.
* Set an **Action** on the button to navigate to the main dashboard view using a different **Bookmark** (e.g., “Main View”).

# **Exploratory Analysis in Jupyter Notebook**

## **Objective:**

To use Python and Jupyter Notebook for preliminary data validation, exploratory analysis, and visualisation before building the final Power BI dashboard. This phase helped validate SQL results, explore patterns, and guide dashboard design.

**Environment:**

* Tool: **Jupyter Notebook**
* Libraries used: pandas, matplotlib, seaborn

**Steps Taken:**

1. Loaded CSV files (orders.csv, order\_details.csv, products.csv, customers.csv) into pandas DataFrames.
2. Verified row counts, column names, and data types to match MySQL imports.
3. Cleaned data by checking/removing duplicates and handling null values. Ensured correct date formats.
4. Merged tables as needed for combined analysis.

**Exploratory Visualisations:**

* **Histogram**: Distribution of Sales and Profit.
* **Box Plot**: Discount vs Profit to spot negative-profit outliers.
* **Bar Chart**: Average Sales by Category.
* **Heatmap**: Correlation between numeric fields.

# **Customer Segmentation Using RFM Analysis and K-Means Clustering**

## **Objective:**

To segment customers based on their purchase behaviour using RFM (Recency, Frequency, Monetary) analysis. K-Means clustering was used to group similar customers, enabling targeted marketing and engagement strategies.

**Steps Taken:**

**1. Calculated RFM Metrics**

Customer data was analysed to compute three key behavioural indicators:

* **Recency**: How many days have passed since the customer's last order
* **Frequency**: How many orders a customer placed
* **Monetary**: Total value of purchases made by the customer

Each customer was assigned a Recency, Frequency, and Monetary value based on their purchase history.

**2. Normalised RFM Values**

Since Recency, Frequency, and Monetary values were on different scales, all values were normalised to bring them within a uniform range. This ensured better performance of clustering algorithms and fair comparison across features.

**3. Applied K-Means Clustering**

The Elbow Method was used to identify the optimal number of clusters for customer segmentation. Based on this, a K-Means clustering model was applied to the normalised RFM data. Each customer was assigned to one of the identified segments.

**4. Visualised and Interpreted Customer Segments**

Customer segments were visualised using scatter plots and pair plots. This helped assess how distinct the clusters were and how customers varied across Recency, Frequency, and Monetary dimensions. Each cluster showed clear patterns that reflected different customer behaviours.

**5. Labelled Customer Segments**

Based on the RFM characteristics of each cluster, segments were interpreted and labelled:

| **Segment** | **Characteristics** | **Label** |
| --- | --- | --- |
| 0 | Infrequent, older purchases, low spend | Lost Customers |
| 1 | Frequent, recent, high-value purchases | Loyal Customers |
| 2 | Moderate frequency and value | Potential Loyalists |
| 3 | New, recent customers with low frequency | New Customers |

**6. Exported Final Clustered Data**

The final dataset containing Customer IDs, RFM scores, assigned cluster numbers, and descriptive segment labels was exported. This data can be reused in Power BI dashboards or marketing campaigns to personalise customer targeting.

**Key Insights:**

* **Loyal Customers** formed a valuable core group with frequent and high-spending behaviour.
* **Lost Customers** had low engagement and were candidates for reactivation campaigns.
* **New Customers** showed potential for nurturing into loyal buyers.
* The segmentation provided clear opportunities for tailored engagement strategies across customer types.

# Conclusion

This project successfully demonstrated a complete data analytics workflow — from SQL-based data extraction and cleansing to Power BI dashboard creation and Python-based customer segmentation. Key business questions were answered using calculated metrics, visualisations, and statistical clustering techniques.  
  
**Key Achievements:**

* Identified top-performing products and customers
* Analysed profit margins and the impact of discounts
* Built an interactive Power BI dashboard with insights panel
* Segmented customers using RFM and K-Means clustering to inform targeted marketing  
    
  **Next Steps:**
* Integrate demographic or behavioural data (if available) for deeper segmentation
* Automate dashboard refresh with live data
* A/B test personalised campaigns based on customer clusters

This analysis equips decision-makers with data-driven insights to optimise sales strategies, manage inventory more effectively, and personalise customer engagement for increased profitability.