SALES

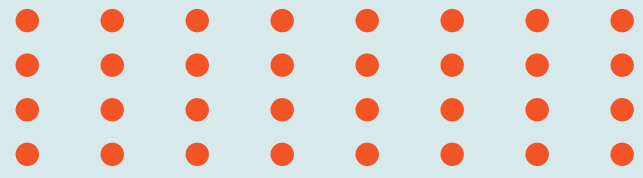## PERFORMANCE ANALYSIS USING SQL

About This Project

This project involves performing structured query language (SQL) analysis on the Superstore dataset, which is a popular retail dataset containing detailed transactional data including orders, products, customers, shipping, sales, and profit information. The goal of this project is to derive meaningful insights and answer business-related questions using SQL queries.

The analysis is divided into three levels of complexity — basic, intermediate, and advanced — allowing a gradual progression from simple data retrieval to more sophisticated analytical tasks. Basic queries focus on understanding data structure and summarizing key metrics such as total sales and order counts. Intermediate queries dive deeper into customer behavior, product performance, and regional sales trends. Advanced queries leverage concepts such as window functions, subqueries, and business logic to identify customer lifetime value, market basket associations, and profitability analysis.

Through this project, we demonstrate the power of SQL in transforming raw data into actionable insights for decision-making in a retail business context.

Queries:

Basic

1. Retrieve All Orders Data

2. Find Total Number of Orders

3. Find Total Sales and Profit

4. List Unique Product Categories and Sub-Categories

5. Find the Most Popular Shipping Mode
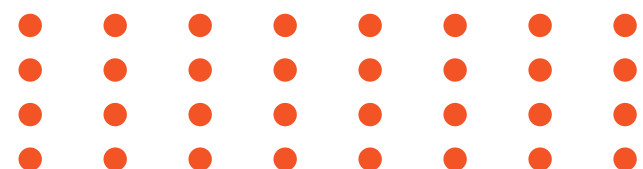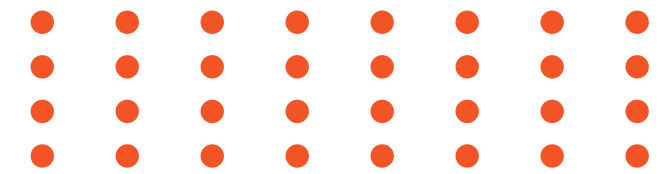
Queries:

Intermediate

1. Identify Top 5 Most Sold Products

 2. Average Discount per Category

3. Customers Who Made More Than 5 Purchases

4. Most Profitable Customer Segment

5. Compare Sales and Profit by Sub-Category
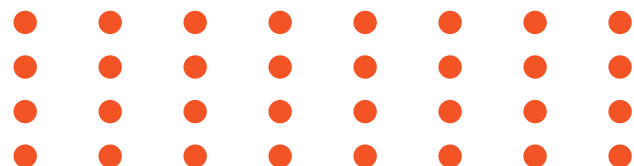
Queries:

Advanced

1. Customer with the Highest Lifetime Value

2. Impact of Discounts on Profit

3. High Sales but Low Profit Products

4. Most Profitable State in Each Region

5. Most Common Product Pairs (Market Basket Analysis)
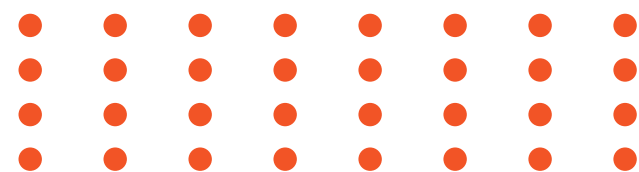
1. Retrieve All Orders Data

SELECT *

FROM

`superstore dataset`;

Row ID	Order ID	Order Date	Ship Date	Ship Mode	Customer ID	Customer Name	Segment	Country	City
1	CA-2013-152156	09-11-2013	12-11-2013	Second Class	CG-12520	Claire Gute	Consumer	United States	Henderson
2	CA-2013-152156	09-11-2013	12-11-2013	Second Class	CG-12520	Claire Gute	Consumer	United States	Henderson
3	CA-2013-138688	13-06-2013	17-06-2013	Second Class	DV-13045	Darrin Van Huff	Corporate	United States	Los Angeles
4	US-2012-108966	11-10-2012	18-10-2012	Standard Class	SO-20335	Sean O'Donnell	Consumer	United States	Fort Lauderdale
5	US-2012-108966	11-10-2012	18-10-2012	Standard Class	SO-20335	Sean O'Donnell	Consumer	United States	Fort Lauderdale
6	CA-2011-115812	09-06-2011	14-06-2011	Standard Class	BH-11710	Brosina Hoffman	Consumer	United States	Los Angeles
7	CA-2011-115812	09-06-2011	14-06-2011	Standard Class	BH-11710	Brosina Hoffman	Consumer	United States	Los Angeles
8	CA-2011-115812	09-06-2011	14-06-2011	Standard Class	BH-11710	Brosina Hoffman	Consumer	United States	Los Angeles
9	CA-2011-115812	09-06-2011	14-06-2011	Standard Class	BH-11710	Brosina Hoffman	Consumer	United States	Los Angeles
10	CA-2011-115812	09-06-2011	14-06-2011	Standard Class	BH-11710	Brosina Hoffman	Consumer	United States	Los Angeles

2. Find Total Number of Orders

SELECT
COUNT(DISTINCT
`Order ID`) AS
Total_Orders
FROM
`superstore dataset`;

Total_Orders
4931

3. Find Total Sales and Profit

SELECT

SUM(Sales) AS Total_Sales, SUM(Profit) AS Total_Profit

FROM

`superstore dataset`;

Total_Sales	Total_Profit
2272449.8562999545	282857.75420000001

4. List Unique Product Categories and Sub-Categories

SELECT DISTINCT
Category, `Sub-Category`
FROM
`superstore dataset`;

	Category	Sub-Category
▶	Furniture	Bookcases
	Furniture	Chairs
	Office Supplies	Labels
	Furniture	Tables
	Office Supplies	Storage
	Furniture	Furnishings
	Office Supplies	Art
	Technology	Phones
	Office Supplies	Binders
	Office Supplies	Appliances
	Office Supplies	Paper
	Technology	Accessories

5. Find the Most Popular Shipping Mode

SELECT

`Ship Mode`, COUNT(*) AS Count

FROM

`superstore dataset`

GROUP BY `Ship Mode`

ORDER BY Count DESC;

	Ship Mode	Count
▶	Standard Class	5780
	Second Class	1886
	First Class	1501
	Same Day	527

1. Identify Top 5 Most Sold Products

```
SELECT
`Product Name`, COUNT(*) AS
Sales_Count
FROM
`superstore dataset`
GROUP BY `Product Name`
ORDER BY Sales_Count DESC
LIMIT 5;
```

Product Name	Sales_Count
Staples	227
Avery Non-Stick Binders	20
KI Adjustable-Height Table	18
Storex Dura Pro Binders	17
Situations Contoured Folding Chairs, 4/Set	15

2. Average Discount per Category

```sql
SELECT
Category, AVG(Discount) AS
Avg_Discount
FROM
`superstore dataset`
GROUP BY Category
ORDER BY Avg_Discount DESC;
```

	Category	Avg_Discount
▶	Furniture	0.17439247830279556
	Office Supplies	0.15556132157066824
	Technology	0.13224578575312493

3. Customers Who Made More Than 5 Purchases

SELECT
`Customer ID`,`Customer Name`,
COUNT(`Order ID`) AS
Purchase_Count FROM `superstore
dataset`
GROUP BY `Customer ID` , `Customer
Name`
HAVING COUNT(`Order ID`) > 5
ORDER BY Purchase_Count DESC;

Customer ID	Customer Name	Purchase_Count
WB-21850	William Brown	34
MA-17560	Matt Abelman	33
JL-15835	John Lee	33
CK-12205	Chloris Kastensmidt	32
PP-18955	Paul Prost	32
EP-13915	Emily Phan	31
AP-10915	Arthur Prichep	31
EH-13765	Edward Hooks	31
JD-15895	Doherty	30
ZC-21910	Zuschuss Carroll	30
SV-20365	Seth Vernon	30

EH-13765

4. Most Profitable Customer Segment

```sql
SELECT
Segment, SUM(Profit) AS Total_Profit
FROM
`superstore dataset`
GROUP BY Segment
ORDER BY Total_Profit DESC;
```

Segment	Total_Profit
Consumer	132669.77519999977
Corporate	90366.3009000002
Home Office	59821.678100000085

5. Compare Sales and Profit by Sub-Category

SELECT
`Sub-Category`,
SUM(Sales) AS Total_Sales,
SUM(Profit) AS Total_Profit
FROM
`superstore dataset`
GROUP BY `Sub-Category`
ORDER BY Total_Profit DESC;

Sub-Category	Total_Sales	Total_Profit
Copiers	149528.0299999994	55617.8249000001
Phones	329753.0880000001	44447.8791000000006
Accessories	167380.3180000001	41936.6356999993
Paper	75356.11799999999	32712.16899999965
Binders	199905.7170000006	29983.02129999986
Chairs	328449.1030000076	26590.16630000026
Storage	216803.2120000012	21527.90829999996
Appliances	107532.161	18138.00539999995
Furnishings	82752.23000000004	11588.64199999987
Art	27118.79199999954	6527.786999999998
Envelopes	15339.48999999993	6460.8691000000035
Labels	12486.312	5546.25399999998

1. Customer with the Highest Lifetime Value

```sql
SELECT
`Customer ID`, `Customer Name`, SUM(Sales) AS Lifetime_Value
FROM
`superstore dataset`
GROUP BY `Customer ID` , `Customer Name`
ORDER BY Lifetime_Value DESC
LIMIT 1;
```

Customer ID	Customer Name	Lifetime_Value
SM-20320	Sean Miller	25043.05

2. Impact of Discounts on Profit

```
SELECT
Discount, ROUND(AVG(Profit), 2) AS Avg_Profit
FROM
`superstore dataset`
GROUP BY Discount
ORDER BY Discount ASC;
```

Discount	Avg_Profit
0	68.11
0.1	96.06
0.15	27.29
0.2	25.2
0.3	-45.68
0.32	-88.56
0.4	-112.51
0.45	-226.65
0.5	-310.7
0.6	-42.03
0.7	-101.39
0.8	-105.74

3. High Sales but Low Profit Products

```sql
SELECT
`Product Name`,
SUM(Sales) AS Total_Sales,
SUM(Profit) AS Total_Profit
FROM
`superstore dataset`
GROUP BY `Product Name`
HAVING SUM(Sales) > 5000 AND SUM(Profit) < 500
ORDER BY Total_Profit;
```

Product Name	Total_Sales	Total_Profit
Cubify CubeX 3D Printer Double Head Print	11099.963	-8879.9704
Lexmark MX611dhe Monochrome Laser Printer	16829.901	-4589.973
Cubify CubeX 3D Printer Triple Head Print	7999.98	-3839.9904
Chromcraft Bull-Nose Wood Oval Conference T...	9917.64	-2876.1156
Bush Advantage Collection Racetrack Conferen...	9544.725	-1934.3975999999998
GBC DocuBind P400 Electric Binding System	17965.068	-1878.1662000000003
Cisco TelePresence System EX90 Videoconferen...	22638.48	-1811.0784
Martin Yale Chadless Opener Electric Letter Ope...	16656.199999999997	-1299.1836
Balt Solid Wood Round Tables	6518.754	-1201.0581
Hon 2090 "Pillow Soft" Series Mid Back Swivel/Til...	5282.424	-989.0495999999999
Bretford "Just In Time" Height-Adjustable Multi-...	5634.9	-964.194
Bevis Oval Conference Table, Walnut	6942.068	-856.0144

4. Most Profitable State in Each Region

```
SELECT Region, State, SUM(Profit) AS Total_Profit
FROM `superstore dataset`
GROUP BY Region, State
HAVING SUM(Profit) = (
SELECT MAX(Total_Profit) FROM (
SELECT Region, State, SUM(Profit) AS Total_Profit
FROM `superstore dataset`
GROUP BY Region, Stat) AS Subquery)ORDER BY Total_Profit DESC;
```

Region	State	Total_Profit
West	California	74669.2036000014

5. Most Common Product Pairs (Market Basket Analysis)

```sql
SELECT
A.`Product Name` AS Product_1,
B.`Product Name` AS Product_2,
COUNT(*) AS Pair_Count
FROM
`superstore dataset` A
JOIN
`superstore dataset` B ON A.`Order ID` = B.`Order ID`
AND A.`Product Name` < B.`Product Name`
GROUP BY Product_1, Product_2
ORDER BY Pair_Count DESC
LIMIT 10;
```

Product_1	Product_2	Pair_Count
Newell 34	Staples	4
KI Adjustable-Height Table	Staples	3
#10 White Business Envelopes,4 1/8 x 9 1/2	Staples	3
Hon Olson Stacker Chairs	Staples	3
Staples	Xerox 1916	3
Adjustable Depth Letter/Legal Cart	Staples	3
Hoover Shoulder Vac Commercial Portable Vacuum	Staples	3
Fellowes Stor/Drawer Steel Plus Storage Drawers	Staples	3
Satellite Sectional Post Binders	Staples	3
Xerox 1894	Xerox 225	2

Thank You