“Superstore analysis isn't just about numbers; it's about understanding our customers and meeting their needs more effectively than ever.”

SUPERSTORE ANALYSIS

~ NISHA CHANDILA

INTRODUCTION

In this presentation, we will explore and analyze a comprehensive dataset containing critical sales information. This dataset provides insights into customer orders across various regions, helping us understand sales performance, profitability, and discount patterns. Key data points include:

- **Order Information**: Tracking orders through unique Order IDs, along with details such as the Order Date and Customer Name.
- **Product Categories**: Data is organized by product Category and Sub-Category, allowing for detailed analysis of different product lines.
- **Geographical Data**: The dataset captures customer locations, including City, State, and Region, which helps identify regional sales trends.
- **Sales Metrics**: Sales revenue, Discounts, and Profit margins provide insights into business performance and financial outcomes.

By leveraging this dataset, we will uncover valuable trends and patterns that can guide strategic business decisions, optimize sales strategies, and enhance profitability across regions and product categories.

Que1 What is the total sales amount?

SQL Query

```
SELECT
        SUM(sales) AS total_sales_amount
FROM `supermart grocery sales - retail analytics dataset`;
```

Output

total_sales_amount
14956982

Que2 What are the total sales, profit, and discount for each category and sub-category?

SQL Query

```
SELECT category,
       sub_category,
       SUM(sales) AS total_sales,
       SUM(profit) AS total_profit,
       SUM(discount) AS total_discount
FROM `supermart grocery sales - retail analytics dataset`
GROUP BY category, sub_category
ORDER BY total_sales DESC;
```

Output

category	sub_category	total_sales	total_profit	total_discount
Beverages	Health Drinks	1051439	267469.79	166.49
Beverages	Soft Drinks	1033874	258135.97	155.97
Snacks	Cookies	768213	190643.7	113.72
Bakery	Breads & Buns	742586	190764.98	113.7
Snacks	Noodles	735435	193685.81	109.86
Snacks	Chocolates	733898	183849.34	112.6
Oil & Masala	Masalas	697480	168999.11	103.65
Bakery	Cakes	685612	168398.46	101.54
Bakery	Biscuits	684083	169357.62	103.29
Oil & Masala	Spices	672876	160302.6	98.42

Que3 Who are the top 10 customers by total sales?

SQL Query

```
SELECT customer_name,
      SUM(sales) AS total_sales
FROM `supermart grocery sales - retail analytics dataset`
GROUP BY customer_name
ORDER BY total_sales
LIMIT 10;
```

Output

customer_name	total_sales
Hafiz	251965
Kumar	262623
Jackson	271793
Yadav	273162
Sudha	273493
Ganesh	278349
Esther	283656
Anu	283735
Rumaiza	284534
Vince	284536

Que4 What are the total sales, profit, and discount for each region?

SQL Query

```
SELECT region,
       SUM(sales) AS total_sales,
       SUM(profit) AS total_profit,
       SUM(discount) AS total_discount
FROM `supermart grocery sales - retail analytics dataset`
GROUP BY region
ORDER BY total_sales DESC;
```

Output

region	total_sales	total_profit	total_discount
West	4798743	1192004.61	719.8
East	4248368	1074345.58	648.41
Central	3468156	856806.84	531.33
South	2440461	623562.89	367.15
North	1254	401.28	0.12

Que5 What is the total sales, profit, and number of orders for a specific category (e.g., "Beverages") by city?

SQL Query

```sql
SELECT
	city,
	COUNT(order_id) AS no_of_orders,
	SUM(sales) AS total_sales,
	SUM(profit) AS total_profit,
	SUM(discount) AS total_discount
FROM `supermart grocery sales - retail analytics dataset`
WHERE category = 'Beverages'
GROUP BY city
ORDER BY total_sales DESC;
```

Output

city	no_of_orders	total_sales	total_profit	total_discount
Madurai	70	111161	27948.79	15.25
Perambalur	72	108348	27216.25	18.34
Ramanadhapuram	63	105203	25709.49	15
Chennai	68	101244	26607.38	16.33
Tirunelveli	71	101065	26211.62	15.72
Vellore	64	99174	27032.91	14.8
Coimbatore	61	98092	25488.05	13.05
Viluppuram	64	93541	22500.28	15.6
Ooty	62	90416	21333.25	14.39
Namakkal	55	89250	22083.86	11.97

Que6 What is the relationship between discount and profit?

SQL Query

```
SELECT discount,
       SUM(profit) AS total_profit
FROM `supermart grocery sales - retail analytics dataset`
GROUP BY discount
ORDER BY discount;
```

Output

discount	total_profit
0.1	132451.16
0.11	141644.47
0.12	136104.25
0.13	135511.87
0.14	142611.11
0.15	138646.69
0.16	147889.02
0.17	127199.25
0.18	141907.32
0.19	154141.1

Que7 What are the total sales and profit by city?

SQL Query

```
SELECT city,
       SUM(sales) AS total_sales,
       SUM(profit) AS total_profit
FROM `supermart grocery sales - retail analytics dataset`
GROUP BY city
ORDER BY total_sales DESC;
```

Output

city	total_sales	total_profit
Kanyakumari	706764	172217.74
Vellore	676550	174073.01
Bodi	667177	173655.13
Tirunelveli	659812	165169.01
Perambalur	659738	171132.19
Salem	657093	160899.3
Pudukottai	653179	164072.63
Tenkasi	643652	156230.72
Karur	642273	169305.94
Krishnagiri	637273	160477.48

Que8 What are the details of the most and least profitable orders?

SQL Query

```
SELECT order_id, customer_name, category, sub_category, sales, profit
FROM `supermart grocery sales - retail analytics dataset`
ORDER BY profit DESC
LIMIT 1;
```

```
SELECT order_id, customer_name, category, sub_category, sales, profit
FROM `supermart grocery sales - retail analytics dataset`
ORDER BY profit ASC
LIMIT 1;
```

Output

order_id	customer_name	category	sub_category	sales	profit
OD3160	Haseena	Bakery	Cakes	2491	1121

order_id	customer_name	category	sub_category	sales	profit
OD9552	Vidya	Beverages	Soft Drinks	505	25.3

Que9 Segment customers based on their total sales into High, Medium, and Low volume category.

SQL Query

```sql
SELECT customer_name,
       SUM(sales) AS total_sales,
       CASE WHEN SUM(sales) > 320000 THEN 'High'
            WHEN SUM(sales) BETWEEN 280000 AND 320000 THEN 'Medium'
            ELSE 'Low'
       END AS sales_categories
FROM `supermart grocery sales - retail analytics dataset`
GROUP BY customer_name
ORDER BY total_sales DESC;
```

Output

customer_name	total_sales	sales_categories
Krithika	334361	High
Amrish	333351	High
Verma	331665	High
Arutra	325720	High
Vidya	321798	High
Vinne	319565	Medium
Shah	318588	Medium
Suresh	315973	Medium
Adavan	315341	Medium
Kumar	262623	Low
Hafiz	251965	Low

Que10 What is the average order value (AOV) for each category?

SQL Query

```
SELECT category,
       AVG(sales) AS avg_order_value
FROM `supermart grocery sales - retail analytics dataset`
GROUP BY category
ORDER BY avg_order_value DESC;
```

Output

category	avg_order_value
Eggs, Meat & Fish	1521.7456
Food Grains	1513.0701
Oil & Masala	1497.7531
Bakery	1494.891
Beverages	1489.5093
Fruits & Veggies	1481.4718
Snacks	1477.9036

Que11 How many orders were placed on each date?

SQL Query

```
SELECT order_date,
       COUNT(order_id) AS number_of_orders
FROM `supermart grocery sales - retail analytics dataset`
GROUP BY order_date
ORDER BY order_date;
```

Output

order_date	number_of_orders
01-01-2018	10
01-02-2016	8
01-02-2017	2
01-02-2018	8
01-03-2015	1
01-03-2016	4
01-03-2017	6
01-03-2018	3
01-04-2015	3
01-04-2016	3

Que12 Which dates had the highest and lowest sales, and what were the sales amounts?

SQL Query

```sql
SELECT order_date,
       SUM(sales) AS total_sales
FROM `supermart grocery sales - retail analytics dataset`
GROUP BY order_date
ORDER BY total_sales DESC
LIMIT 1;

SELECT order_date,
       SUM(sales) AS total_sales
FROM `supermart grocery sales - retail analytics dataset`
GROUP BY order_date
ORDER BY total_sales ASC
LIMIT 1;
```

Output

order_date	total_sales
5/16/2018	541

order_date	total_sales
09-02-2018	55976

Que13 What is the profit margin for each sub-category?

SQL Query

```
SELECT sub_category,
       SUM(profit) / SUM(sales) * 100 AS profit_margin
FROM `supermart grocery sales - retail analytics dataset`
GROUP BY sub_category
ORDER BY profit_margin;
```

Output

sub_category	profit_margin
Spices	23.823498
Chicken	23.832969
Atta & Flour	23.914961
Masalas	24.229958
Fresh Fruits	24.431317
Cakes	24.561773
Biscuits	24.756882
Mutton	24.769208
Cookies	24.816516
Dals & Pulses	24.88336

Que14 How many repeat customers are there, and what is their total sales contribution?

SQL Query

```
SELECT customer_name,
       COUNT(order_id) AS number_of_orders,
       SUM(sales) AS total_sales
FROM `supermart grocery sales - retail analytics dataset`
GROUP BY customer_name
HAVING number_of_orders > 1
ORDER BY total_sales DESC;
```

Output

customer_name	number_of_orders	total_sales
Krithika	224	334361
Amrish	227	333351
Verma	218	331665
Arutra	218	325720
Vidya	215	321798
Vinne	203	319565
Shah	215	318588
Suresh	212	315973
Adavan	205	315341
Surya	209	312645

Que15 Which regions offer the highest and lowest average discounts?

SQL Query

```
SELECT region,
       AVG(discount) AS avg_discount
FROM `supermart grocery sales - retail analytics dataset`
GROUP BY region
ORDER BY avg_discount DESC;
```

Output

region	avg_discount
Central	0.228726
East	0.227672
South	0.226776
West	0.224727
North	0.12

Que16 How many orders were placed in each state, broken down by category?

SQL Query

```
SELECT category,
       state,
       COUNT(order_id) AS number_of_orders
FROM `supermart grocery sales - retail analytics dataset`
GROUP BY state, category
ORDER BY state, category;
```

Output

category	state	number_of_orders
Bakery	Tamil Nadu	1413
Beverages	Tamil Nadu	1400
Eggs, Meat & Fish	Tamil Nadu	1490
Food Grains	Tamil Nadu	1398
Fruits & Veggies	Tamil Nadu	1418
Oil & Masala	Tamil Nadu	1361
Snacks	Tamil Nadu	1514

Que17 What percentage of the total profit does each category contribute?

SQL Query

```
SELECT category,
       SUM(profit) AS total_profit,
       (SUM(profit) / (SELECT SUM(profit) FROM `supermart grocery sales - retail analytics dataset`) * 100) AS profit_contribution_per
FROM `supermart grocery sales - retail analytics dataset`
GROUP BY category
ORDER BY profit_contribution_per;
```

Output

category	total_profit	profit_contribution_per
Oil & Masala	497895.29	13.287408
Beverages	525605.76	14.026922
Bakery	528521.06	14.104723
Food Grains	529162.64	14.121845
Fruits & Veggies	530400.38	14.154877
Eggs, Meat & Fish	567357.22	15.141149
Snacks	568178.85	15.163076

Que18 What are the total sales and profit by city in a specific region (e.g., "West")?

SQL Query

```sql
SELECT customer_name,
       SUM(profit) AS total_profit,
       CASE WHEN SUM(profit) > 86000 THEN 'High'
            WHEN SUM(profit) BETWEEN 68000 AND 86000 THEN 'Medium'
            ELSE 'Low'
       END AS profit_segment
FROM `supermart grocery sales - retail analytics dataset`
GROUP BY customer_name
ORDER BY total_profit DESC;
```

Output

city	total_sales	total_profit
Bodi	247873	65524.36
Tenkasi	225701	55160.21
Vellore	225617	55517.4
Dharmapuri	220077	55299.1
Tirunelveli	219492	54063.63
Perambalur	215899	56035.2
Chennai	207339	51681.22
Karur	206125	52420.11
Virudhunagar	206119	51188.65
Kanyakumari	204365	50399.89

Que19 Segment customers based on their total profit contribution into High, Medium, and Low profit segments.

SQL Query

```sql
SELECT customer_name,
       SUM(profit) AS total_profit,
       CASE WHEN SUM(profit) > 86000 THEN 'High'
            WHEN SUM(profit) BETWEEN 68000 AND 86000 THEN 'Medium'
            ELSE 'Low'
       END AS profit_segment
FROM `supermart grocery sales - retail analytics dataset`
GROUP BY customer_name
ORDER BY total_profit DESC;
```

Output

customer_name	total_profit	profit_segment
Arutra	87572.4	High
Vidya	86725.64	High
Krithika	85633.03	Medium
Akash	82121.26	Medium
Surya	80996.85	Medium
Ganesh	67094.89	Low
Ram	66970.6	Low
Hafiz	64037.35	Low

Que20 What is the distribution of sales by sub-category in a specific state (e.g., "Tamil Nadu")?

SQL Query

```
SELECT sub_category,
       SUM(sales) AS total_sales
FROM `supermart grocery sales - retail analytics dataset`
WHERE state = 'Tamil Nadu'
GROUP BY sub_category
ORDER BY total_sales DESC;
```

Output

sub_category	total_sales
Health Drinks	1051439
Soft Drinks	1033874
Cookies	768213
Breads & Buns	742586
Noodles	735435
Chocolates	733898
Masalas	697480
Cakes	685612
Biscuits	684083
Spices	672876

Que21 What percentage of the total sales comes from the top 20% of customers?

SQL Query

```
WITH customer_sales AS (
    SELECT customer_name,
           SUM(sales) AS total_sales
    FROM `supermart grocery sales - retail analytics dataset`
    GROUP BY customer_name
),
customer_rank AS (
    SELECT customer_name,
           total_sales,
           NTILE(5) OVER (ORDER BY total_sales DESC) AS sales_rank
    FROM customer_sales
)
SELECT SUM(total_sales) / (SELECT SUM(sales) FROM `supermart grocery sales - retail analytics dataset`) * 100 AS sales_contribution_pct
FROM customer_rank
WHERE sales_rank = 1;
```

Output

sales_contribution_pct
21.5886

Que22 Which sub-categories are the best performers in each state?

SQL Query

```
SELECT state,
       sub_category,
       SUM(sales) AS total_sales
FROM `supermart grocery sales - retail analytics dataset`
GROUP BY state, sub_category
ORDER BY state, total_sales DESC;
```

Output

state	sub_category	total_sales
Tamil Nadu	Health Drinks	1051439
Tamil Nadu	Soft Drinks	1033874
Tamil Nadu	Cookies	768213
Tamil Nadu	Breads & Buns	742586
Tamil Nadu	Noodles	735435
Tamil Nadu	Chocolates	733898
Tamil Nadu	Masalas	697480
Tamil Nadu	Cakes	685612
Tamil Nadu	Biscuits	684083
Tamil Nadu	Spices	672876


THANK
YOU