Answers

1. Which product categories generate the highest Profit? Also, display top 10 sub-categories generating maximum profit.

```
SELECT
 p.Category,
 ROUND(
   SUM(t.Line_Total)                -- total revenue per category
   - SUM(p.Production_Cost * t.Quantity),  -- total cost per category
   2
 ) AS Profit
FROM transaction t
JOIN Product p
 ON t.Product_ID = p.Product_ID
WHERE t.Transac_Type = 'Sale'
GROUP BY p.Category
ORDER BY Profit DESC;

SELECT
 p.Category,
 p.Sub_Category,
 ROUND(
   SUM(t.Line_Total)
   - SUM(p.Production_Cost * t.Quantity),
   2
 ) AS Profit
FROM `transaction` t
JOIN Product p
 ON t.Product_ID = p.Product_ID
WHERE t.Transac_Type = 'Sale'
GROUP BY
 p.Category,
 p.Sub_Category
ORDER BY
 Profit DESC LIMIT 10;
```

Answer-
Categories ordered by their profit.

	Category	Profit
1	Feminine	55095.28
2	Masculine	54324.19
3	Children	11194.54

Top Sub-Categories with their profit-

	Category	Sub_Category	Profit
1	Masculine	Suits and Blazers	13488.39
2	Feminine	Suits and Sets	9946.98
3	Masculine	Pants and Jeans	8758.78
4	Masculine	Coats and Blazers	7959.08
5	Feminine	Pants and Jeans	7707.59
6	Feminine	Coats and Blazers	7560.39
7	Feminine	Dresses and Jumpsuits	7425.33
8	Masculine	Sportswear	5983.64
9	Masculine	Sweaters and Sweatshirts	5333.48
10	Masculine	T-shirts and Polos	5039.01

From results of this query, we can see that both Masculine and Feminine clothes generate approximately same revenue, while among sub-categories, revenue of Masculine Suits and Blazers is way higher than others.

2. Compare average profit rom discounted v/s non-discounted products? Also display the percentage share of discounted products in total sales.

```
SELECT
 SUM(CASE WHEN t.Discount > 0 THEN t.Quantity ELSE 0 END) AS
Discounted_Units_Sold,
 SUM(CASE WHEN t.Discount = 0 THEN t.Quantity ELSE 0 END) AS
NonDiscounted_Units_Sold,
 ROUND(
   SUM(CASE WHEN t.Discount > 0 THEN t.Quantity ELSE 0 END) * 100.0
   / SUM(t.Quantity),
   2
 ) AS Discounted_Share_Percent
```

```
FROM `transaction` t
WHERE t.Transac_Type = 'Sale';


SELECT
 CASE
   WHEN t.Discount > 0 THEN 'Discounted'
   ELSE 'Non-Discounted'
 END AS Price_Type,
 ROUND(
   AVG(
     t.Line_Total
     - (p.Production_Cost * t.Quantity)
   ),
   2
 ) AS Avg_Profit_Per_Transaction
FROM `transaction` t
JOIN Product p
 ON t.Product_ID = p.Product_ID
WHERE t.Transac_Type = 'Sale'
GROUP BY Price_Type;
```

Answer-

	Discounted_Units_Sold	NonDiscounted_Units_Sold	Discounted_Share_Percent
1	3502	5059	40.91

	Price_Type	Avg_Profit_Per_Transaction
1	Discounted	6.12
2	Non-Discounted	21.89

Above results imply that there is a major issue with discount strategy. Not only the no. of discounted products sold is less than No. of non-discounted units, even, the average profit with discounted products is way lesser than profit from Non-Discounted products.

3. What are top 5 most purchased products in the store?

```
SELECT
 p.Product_ID,
 p.Description,
```

```
 SUM(t.Quantity) AS Total_Units_Sold
FROM transaction t
JOIN Product p
 ON t.Product_ID = p.Product_ID
WHERE t.Transac_Type = 'Sale'
GROUP BY
 p.Product_ID,
 p.Description
ORDER BY
 Total_Units_Sold DESC
LIMIT 5;
```

Answer-

	Product_ID	Description	Total_Units_Sold
1	261	Executivo Linho Mostarda Com Laço	16
2	689	Streetwear Algodão Vermelho Com Capuz	15
3	1406	Clássico Denim Vermelho Estampado	14
4	54	Despojado Seda Branco Com Zíper	14
5	299	High-Tech Lã Preto Estampado	14

Top 5 most purchased styles are- Executive, Streetwear, Classic , Laid-Back and Knitted.

4. What is the percentage of revenue coming from each payment method?

```
SELECT
 Payment_Method,
 ROUND(SUM(Invoice_Total), 2) AS Sum_Revenue,
 ROUND(SUM(Invoice_Total) * 100 /
       (SELECT SUM(Invoice_Total)
        FROM `transaction`
        WHERE Transac_Type = 'Sale'), 2) AS Revenue_Percentage
FROM `transaction`
WHERE Transac_Type = 'Sale'
GROUP BY Payment_Method
ORDER BY Revenue_Percentage DESC;
```

Answer-

	Payment_Method	Sum_Revenue	Revenue_Percentage
1	Credit Card	309550.3	79.75
2	Cash	78601.8	20.25

A huge amount of revenue comes from Credit Cards, therefore, store can announce partnerships with Credit Card companies providing cashbacks and other exclusive deals on different Credit Cards, thus boosting customer engagement.

5. What is the average spending of customers of each Gender respectively. Also, what are the top 3 most frequently bought products for each gender?

```
SELECT
 c.Gender,
 ROUND(SUM(t.Line_Total) / COUNT(DISTINCT t.Invoice_ID), 2) AS
Avg_Spending_Per_Order
FROM `transaction` t
JOIN Customer c ON t.Customer_ID = c.Customer_ID
WHERE t.Transac_Type = 'Sale'
GROUP BY c.Gender;

SELECT
 c.Gender,
 p.Product_ID,
 p.Description,
 SUM(t.Quantity) AS Total_Quantity
FROM `transaction` t
JOIN Customer c ON t.Customer_ID = c.Customer_ID
JOIN Product p ON t.Product_ID = p.Product_ID
WHERE t.Transac_Type = 'Sale' AND c.Gender = 'M'
GROUP BY c.Gender, p.Product_ID, p.Description
ORDER BY Total_Quantity DESC
LIMIT 3;

SELECT
 c.Gender,
 p.Product_ID,
 p.Description,
 SUM(t.Quantity) AS Total_Quantity
FROM `transaction` t
JOIN Customer c ON t.Customer_ID = c.Customer_ID
JOIN Product p ON t.Product_ID = p.Product_ID
WHERE t.Transac_Type = 'Sale' AND c.Gender = 'F'
GROUP BY c.Gender, p.Product_ID, p.Description
ORDER BY Total_Quantity DESC
LIMIT 3;

SELECT
 c.Gender,
 p.Product_ID,
 p.Description,
 SUM(t.Quantity) AS Total_Quantity
```

```
FROM `transaction` t
JOIN Customer c ON t.Customer_ID = c.Customer_ID
JOIN Product p ON t.Product_ID = p.Product_ID
WHERE t.Transac_Type = 'Sale' AND c.Gender = 'D'
GROUP BY c.Gender, p.Product_ID, p.Description
ORDER BY Total_Quantity DESC
LIMIT 3;
```

Answer-
Dataset has three genders - M (Male), F(Female) and D(Other Gender).
Average Spending per Gender-

	Gender	Avg_Spending_Per_Order
1	M	34.82
2	F	35.29
3	D	23.88

Top 3 Products bought by Male-

	Gender	Product_ID	Description	Total_Quantity
1	M	2067	Retrô Poliéster Prateado Acolchoado	10
2	M	195	Casual Cetim Preto Com Bolsos	10
3	M	314	Vintage Veludo Verde Com Bolsos	9

Top 3 Products bought by Female-

	Gender	Product_ID	Description	Total_Quantity
1	F	54	Despojado Seda Branco Com Ziper	12
2	F	261	Executivo Linho Mostarda Com Laço	11
3	F	689	Streetwear Algodão Vermelho Com Capuz	11

Top 3 Products bought by Other Gender people-

	Gender	Product_ID	Description	Total_Quantity
1	D	1000	High-Tech Tricot Rosa Com Botões	1
2	D	1380	Punk Algodão Vermelho Bordado	1
3	D	845	Retrô Camurça Turquesa Com Botões	1

From the results we can see that both Male and Female spend approximately the same amount on each order. While people from other Gender spend significantly less. Therefore, Store can include more inclusive clothing design to cater to everyone.
Product names are present in Portugese. On translating to English, we can see that Retro, Casual and Vintage style are most popular among Male. Among individuals identifying as Female, Executive, Laid-Back, and Streetwear styles are the most popular. For those identifying as Other, Knitted, Embroidered, and Retro designs are more favored.

6. Find out RFM metrics for each customer.

```sql
WITH customer_metrics AS (
 SELECT
   Customer_ID,
   DATEDIFF(CURDATE(), MAX(Date)) AS Recency,
   COUNT(*) AS Frequency,
   SUM(Line_Total) AS Monetary
 FROM `transaction`
 WHERE Transac_Type = 'Sale'
 GROUP BY Customer_ID
)
SELECT
 Customer_ID,
 Recency,
 Frequency,
 Monetary,
 CONCAT(
   NTILE(5) OVER (ORDER BY Recency ASC),
   NTILE(5) OVER (ORDER BY Frequency DESC),
   NTILE(5) OVER (ORDER BY Monetary DESC)
 ) AS RFM_Score
FROM customer_metrics;
```

Answer-

	Customer_ID	Recency	Frequency	Monetary	RFM_Score
1	8329	840	2	324	511
2	71632	833	1	288	241
3	48914	840	1	286.5	521
4	2220	830	1	280.5	241
5	22103	832	1	265.5	241
6	34301	840	2	262	511
7	7616	826	2	239	111
8	48315	826	2	223.5	111
9	44352	827	1	223.5	131
10	69233	835	1	223.5	351
11	74631	826	1	208.5	131
12	26504	825	2	208	111
13	42649	838	1	205.5	411
14	440	828	2	201.5	111
15	36316	831	1	196	241

RFM analysis is a marketing technique used to quantitatively rank and group customers based on the recency, frequency and monetary total of their recent transactions to identify the best customers and perform targeted marketing campaigns.

7. Obtain the top 3 most popular sizes for each product category.

```
SELECT
 css.Category,
 css.Size,
 css.units_sold
FROM (
 SELECT
   p.Category,
   t.Size,
   SUM(t.Quantity) AS units_sold
 FROM `transaction` t
 JOIN Product p
   ON t.Product_ID = p.Product_ID
 WHERE t.Transac_Type = 'Sale'
 GROUP BY
   p.Category,
   t.Size
) AS css
WHERE (
 SELECT COUNT(*)
 FROM (
   SELECT
     p2.Category,
```

```
    t2.Size,
    SUM(t2.Quantity) AS units_sold2
  FROM `transaction` t2
  JOIN Product p2
    ON t2.Product_ID = p2.Product_ID
  WHERE
    t2.Transac_Type = 'Sale'
    AND p2.Category = css.Category
  GROUP BY
    p2.Category,
    t2.Size
) AS css2
WHERE css2.units_sold2 > css.units_sold
) < 3
ORDER BY
 css.Category,
 css.units_sold DESC;
```

Answer-

	Category	Size	units_sold
1	Children	P	588
2	Children	M	541
3	Children	G	239
4	Feminine	M	1312
5	Feminine	S	1232
6	Feminine	L	461
7	Masculine	M	1904
8	Masculine	L	687
9	Masculine	XL	244

Most popular sizes among Children are P,M,G . In Feminine it is M,S,L and M,L,XL for Masculine. Therefore, store needs to keep maximum stock of these products.