**The entire Assignment has been completed on the dummy dataset created by me :**

**Task 1: Data Exploration and Cleaning**

**1 Load the dataset into a SQL database and examine its structure.**

SELECT \* FROM Customers;

SELECT \* FROM Products;

SELECT \* FROM Sales;

SELECT \* FROM Inventory;

1. **Identify and handle missing values appropriately (e.g., remove rows, impute values).**

REMOVE ROWS: SELECT \* FROM Customers where name IS NOT NULL;

IMPUTE VALUES: NVL AND COALESCE are used to handle null values in sql.

NVL takes two arguments whereas COALESCE takes more than two values

and both returns first non-null values

**3.Perform any necessary data cleaning operations to ensure data integrity and consistency**

Data integrity refers to the overall accuracy, completeness, and reliability of data.

It can be specified by the lack of variation between two instances or consecutive updates of a record,

indicating that your information is error-free.

**Task 2: Data Analysis**

1. **Calculate the total revenue generated by the company for each product category.**

SELECT

    p.category,

    SUM(p.price\*s.quantity) AS total\_revenue

from Products p

join Sales s

on p.product\_id = s.product\_id

GROUP BY p.category;

**SOLUTION:**

"**womenswear" 4750**

**"menswear" 32100**

1. **Determine the top 5 customers who have made the highest total purchases, considering the customer's age and gender.**

SELECT

    c.name,

    c.customer\_id,

    c.age,

    c.gender ,

   sum(s.amount) as total\_amount

from Customers c

join Sales s

on c.customer\_id = s.customer\_id

group by c.customer\_id,c.age,c.gender,c.name

order by total\_amount desc

limit 5

**SOLUTION**

**"Nichole"**

**"Sam"**

**"Anusha"**

**"Kriti"**

**"Yashi"**

1. **Identify the most profitable product category by calculating the average revenue per unit sold.**

with cte AS

(SELECT

    p.category,

    SUM(p.price \* s.quantity) AS total\_revenue, SUM(s.quantity) AS total\_quantity

FROM Products p

JOIN Sales s

ON p.product\_id = s.product\_id

GROUP BY p.category

)

SELECT category, total\_revenue/total\_quantity AS  avg\_revenue\_unit FROM cte

ORDER BY avg\_revenue\_unit DESC

LIMIT 1

**SOLUTION: "menswear" 1783**

**4.Analyze the inventory data and identify products that need restocking (stock count less than a specified threshold).**

SELECT

    p.name as product\_name,

     i.stock\_count

from Products p

JOIN Inventory i

on p.product\_id = i.product\_id

Where i.stock\_count <=50;

**SOLUTION:**

**"dress" 20**

**"shirt" 10**

**"t-shirt" 17**

**"bags" 25**

**"earrings"30**

**"shoes" 22**

**"boots" 10**

**Task 3: Advanced Analysis and Reporting**

1. **Write a SQL query to calculate the average age of customers for each product category.**
2. SELECT
3. p.category,
4. ROUND(AVG(C.age::numeric),0)AS average\_age
5. FROM Products p
6. JOIN Sales s
7. on p.product\_id = s.product\_id
8. join Customers c
9. on c.customer\_id = s.customer\_id
10. group by p.category;

**SOLUTION:**

**"womenswear" 45**

**"menswear" 23**

**2. Write a SQL query to retrieve the top 3 product categories that have the highest average transaction amount.**

 SELECT

    p.category, p.name,

    ROUND(AVG(s.amount),2) AS transaction\_amount

FROM Sales s JOIN products p

ON s.product\_id= p.product\_id

GROUP BY p.category, p.name

ORDER BY transaction\_amount DESC

LIMIT 3;

**SOLUTION:**

**"womenswear" "watch" 6000.00**

**"menswear" "shirt" 2800.0**

**"menswear" "shoes" 2700.00**

1. **Create a comprehensive report summarizing your findings from Task 2 and Task 3. Include relevant tables, charts, and explanations to present your analysis clearly.**

* **Total revenue generated from each category:**
  + "womenswear" = 4750
  + "menswear" = 32100
* **Top 5 customers with highest purchasing amount:**
  + "Nichole"
  + "Sam"
  + "Anusha"
  + "Kriti"
  + "Yashi"
* **Most profitable product category based on average revenue per unit:**

"menswear" 1783

* **Stock units less than 50 units are:**

"dress" 20,

"shirt" 10

"t-shirt" 17

"bags" 25

"earrings"30

"shoes" 22

"boots" 10

* **Average age based on each category:**

"womenswear" 45

"menswear" 23

* **Top 3 category based on highest average transaction amount:**

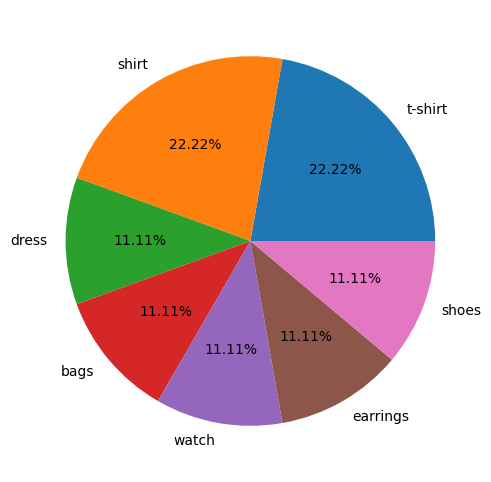
"womenswear" "watch" 6000.00

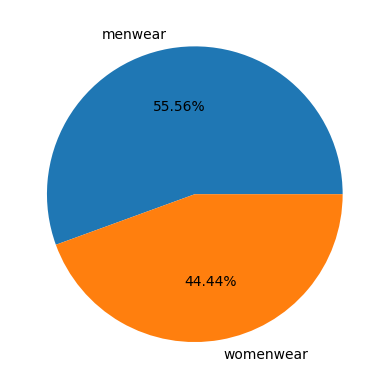
"menswear" "shirt" 2800.0

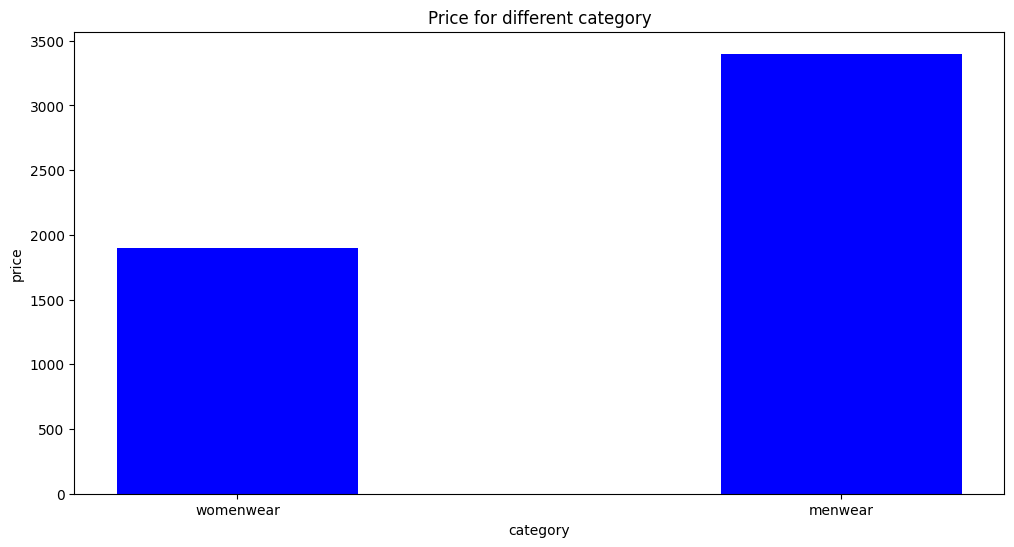
"menswear" "shoes" 2700.00

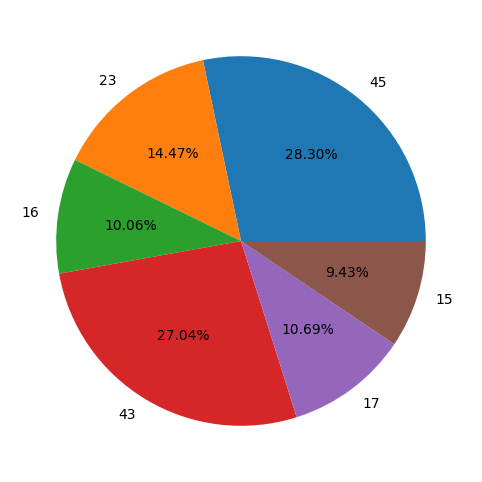
**Bonus Task (Optional): Perform a customer segmentation analysis to identify different customer groups based on their purchasing behaviour, age, and gender. Provide insights on each customer segment and suggest personalized marketing strategies for each segment.**

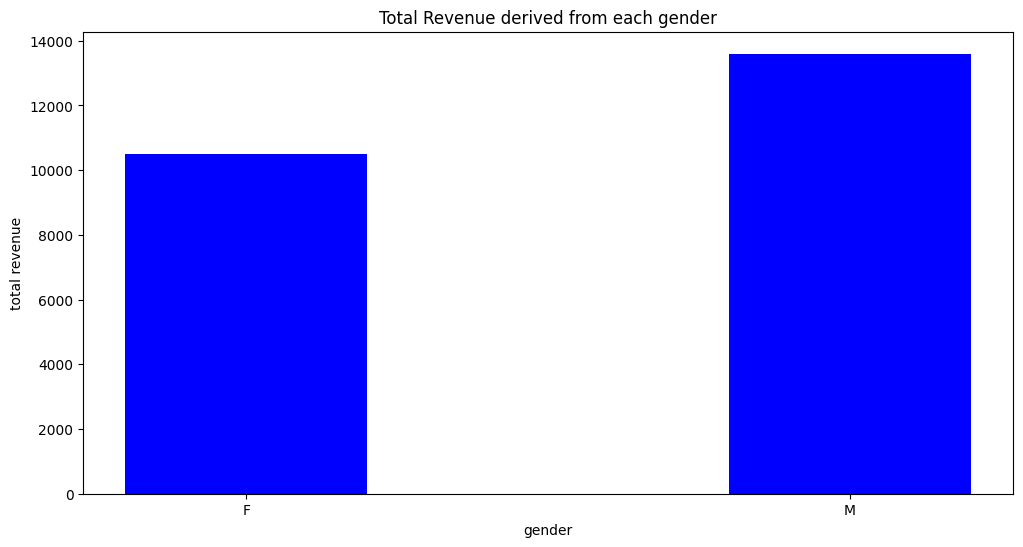
**Click on the link below to see a detailed analysis of the project based on purchasing behaviour, age, and gender**

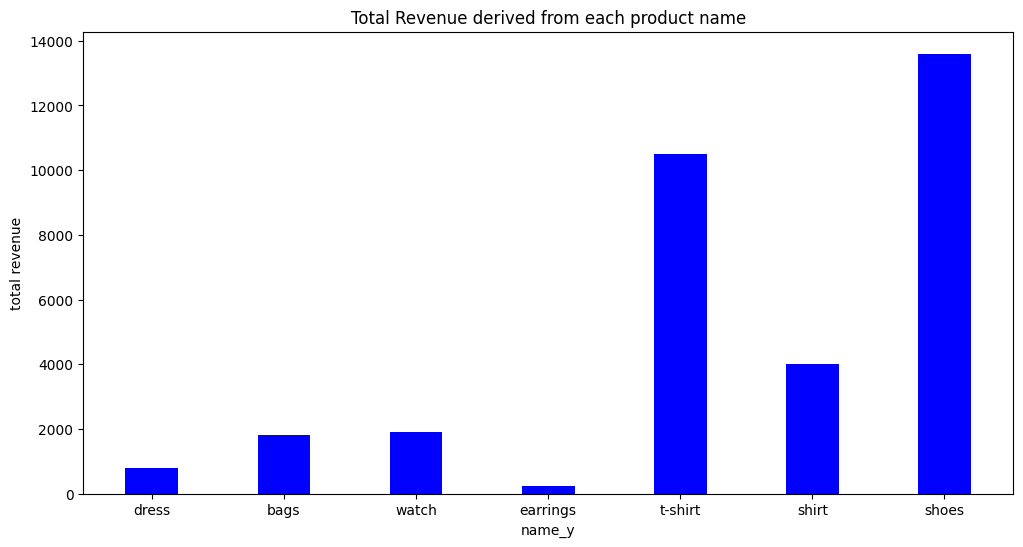


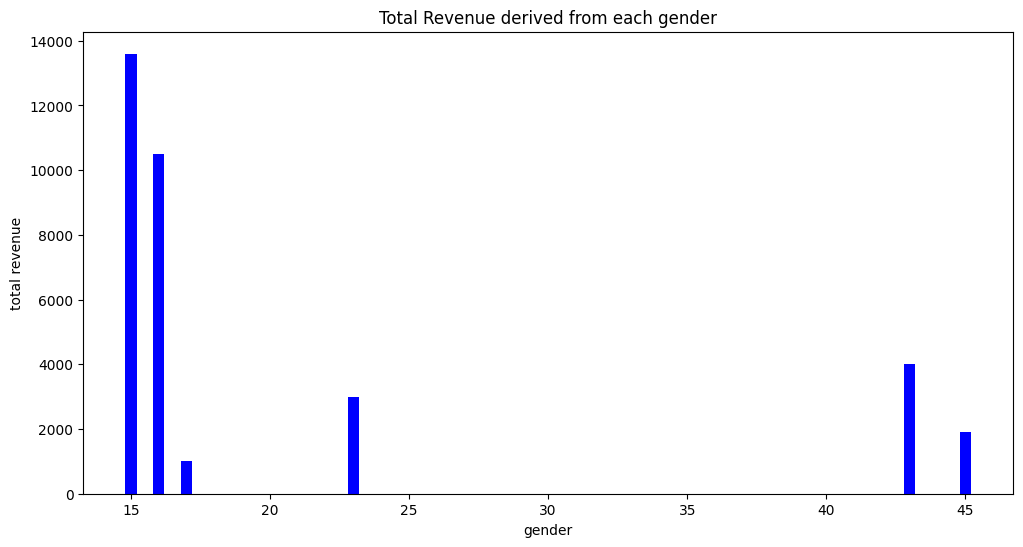












* **most sold product name is t-shirt and shirt**
* **menswear includes 55.56 percent of the product category and 44.46 is covered under womenswear**
* **Prices are high for the category that belongs to menswear**
* **maximum revenue is generated from the product shoes**
* **28.30% of people fall under the age bracket of 45, who are using our product**