```
CREATE DATABASE p1_retail_db;

use p1_retail_db;

CREATE TABLE retail_sales

(

   transactions_id INT PRIMARY KEY,

   sale_date DATE,

   sale_time TIME,

   customer_id INT,

   gender VARCHAR(10),

   age INT,

   category VARCHAR(35),

   quantity INT,

   price_per_unit FLOAT,

   cogs FLOAT,

   total_sale FLOAT

);

select*from retail_sales;

insert into retail_sales (transactions_id, sale_date, sale_time, customer_id, gender, age, category,
quantity, price_per_unit, cogs, total_sale)

VALUES

(1, '2022-11-05', '10:47:00', 117, 'Male', 41, 'Clothing', 3, 300, 129, 900),

(2, '2022-07-09', '11:00:00', 52, 'Male', 46, 'Beauty', 3, 500, 145, 1500),

(3, '2022-12-12', '10:48:00', 5, 'Female', 40, 'Clothing', 4, 300, 84, 1200),

(4, '2022-01-06', '08:53:00', 85, 'Male', 31, 'Clothing', 3, 300, 129, 900),

(5, '2022-11-14', '08:35:00', 48, 'Male', 46, 'Beauty', 3, 500, 235, 1500),

(6, '2023-11-10', '06:46:00', 13, 'Male', 28, 'Clothing', 4, 300, 144, 1200),

(7, '2023-07-01', '06:58:00', 104, 'Male', 64, 'Beauty', 2, 500, 195, 1000),

(8, '2023-03-20', '06:17:00', 106, 'Male', 24, 'Clothing', 4, 300, 81, 1200),

(9, '2023-01-29', '08:46:00', 42, 'Male', 24, 'Clothing', 4, 300, 99, 1200),
```

(10, '2023-08-26', '10:04:00', 71, 'Male', 63, 'Electronics', 4, 300, 132, 1200),

(11, '2023-02-13', '06:28:00', 61, 'Female', 39, 'Beauty', 4, 300, 120, 1200),

(12, '2023-11-22', '07:18:00', 30, 'Male', 33, 'Beauty', 4, 500, 195, 2000),

(13, '2023-11-13', '10:42:00', 124, 'Male', 63, 'Electronics', 4, 300, 114, 1200),

(14, '2023-09-29', '21:00:00', 4, 'Female', 60, 'Beauty', 3, 30, 10.8, 90),

(15, '2023-07-21', '22:42:00', 1, 'Male', 32, 'Clothing', 2, 25, 12, 50),

(16, '2022-02-08', '19:34:00', 21, 'Male', 41, 'Electronics', 1, 30, 9.3, 30),

(17, '2023-05-09', '18:39:00', 85, 'Male', 42, 'Beauty', 4, 50, 24.5, 200),

(18, '2023-01-20', '18:15:00', 136, 'Male', 41, 'Electronics', 1, 30, 13.2, 30),

(19, '2022-01-30', '17:20:00', 56, 'Male', 42, 'Beauty', 4, 50, 18, 200),

(20, '2023-06-01', '18:41:00', 50, 'Female', 20, 'Beauty', 1, 50, 24, 50),

##DATA EXPLORATION AND CLEANING :

1.Determine the total number of records in the dataset.

```
select count(*) from retail_sales;
```

2.Find out how many unique customers are in the dataset.

```
select customer_id from retail_sales;
select count(distinct customer_id) from retail_sales;
```

#.3. Check for any null values in the dataset and delete records with missing data.

DATA ANALYSIS & FINDINGS :

1. Write a SQL query to retrieve all columns for sales made on '2022-11-05:

```
select * from retail_sales where sale_date='2022-11-05';
```

#2. Write a SQL query to retrieve all transactions where the category is 'Clothing' and the quantity sold is more than 4 in the month of Nov-2022:

```
select * from  retail_sales where category='Clothing' and quantity >= 4 and month(sale_date) = 11 and year(sale_date)=2022;
```

#3. Write a SQL query to calculate the total sales (total_sale) for each category.:

```
select category,sum(total_sale)
from retail_sales
group by category;
```

#4.Write a SQL query to find the average age of customers who purchased items from the 'Beauty' category.:

```
select avg(age) from retail_sales where category= 'Beauty';
```

#5.Write a SQL query to find all transactions where the total_sale is greater than 1000.:

```
select * from retail_sales
where total_sale > 1000;
```

#6. Write a SQL query to find the total number of transactions (transaction_id) made by each gender in each category.:

```
select gender, category, count(transactions_id ) from retail_sales
```

```
group by gender, category;


#7. **Write a SQL query to calculate the average sale for each month.
# Find out best selling month in each year**:

select avg(total_sale), year(sale_date), month(sale_date) from retail_sales
group by year(sale_date), month(sale_date);


 select max(temp.average_sale) from
(select avg(r1.total_sale) average_sale, year(r1.sale_date), month(r1.sale_date) from retail_sales as r1
where average_sale >
(
select avg(s1.total_sale) from retail_sales s1
where year(s1.sale_date)=year(r1.sale_date)
group by year(s1.sale_date),month(s1.sale_date)
)
 group by year(r1.sale_date)) as temp;


#8. *Write a SQL query to find the top 5 customers based on the highest total sales *:

select sum(total_sale) , customer_id
from retail_sales
group by customer_id order by sum(total_sale) desc limit 5;

select sum(total_sale) , customer_id
from retail_sales where customer_id=3;
```

```
select * from retail_sales;
```

#9.Write a SQL query to find the number of unique customers who purchased items from each category.:

```
select category, count(distinct customer_id)
from  retail_sales
group by category;
```

#10. Write a SQL query to create each shift and number of orders (Example Morning <12, Afternoon Between 12 & 17, Evening >17):

```
select count(transactions_id), sale_time
from retail_sales group by sale_time;
```