SQL Code

```sql
use my_db;

-- Verify data insertion
SELECT COUNT(*) AS total_records FROM retail_sales;
SELECT * FROM retail_sales LIMIT 10;


-- Checked for NULL values across all columns.
SELECT * FROM retail_sales
WHERE transactions_id IS NULL;

SELECT * FROM retail_sales
WHERE sale_time IS NULL;

SELECT * FROM retail_sales
WHERE sale_date IS NULL;

SELECT * FROM retail_sales
WHERE customer_id IS NULL;

SELECT * FROM retail_sales
WHERE gender IS NULL;

SELECT * FROM retail_sales
WHERE age IS NULL;

SELECT * FROM retail_sales
WHERE category IS NULL;

SELECT * FROM retail_sales
WHERE quantity IS NULL;

SELECT * FROM retail_sales
WHERE cogs IS NULL;

SELECT * FROM retail_sales
WHERE total_sale IS NULL;

-- Removed incomplete records to ensure data integrity.
SELECT * FROM retail_sales
WHERE transactions_id IS NULL
    OR sale_date IS NULL
    OR sale_time IS NULL
    OR customer_id IS NULL
    OR gender IS NULL
    OR age IS NULL
    OR category IS NULL
    OR quantity IS NULL
    OR price_per_unit IS NULL
    OR cogs IS NULL
    OR total_sale IS NULL;

```

```sql
set sql_safe_updates = 0;

DELETE FROM retail_sales
WHERE transactions_id IS NULL
    OR sale_date IS NULL
    OR sale_time IS NULL
    OR customer_id IS NULL
    OR gender IS NULL
    OR age IS NULL
    OR category IS NULL
    OR quantity IS NULL
    OR price_per_unit IS NULL
    OR cogs IS NULL
    OR total_sale IS NULL;


-- Exploratory Data Analysis (EDA)
/*
Q1. Total Records Count
*/
SELECT COUNT(*) AS TotalRecords
FROM retail_sales;

/*
Q2. Count of unique customers
*/
SELECT COUNT(DISTINCT customer_id) AS TotalCustomers
FROM retail_sales;


/*
Q3. HOW MANY CATEGORIES DO WE HAVE ?
*/
SELECT COUNT(DISTINCT Category) AS TotalCategories
FROM retail_sales;


-- DATA ANALYSIS OVER BUSINESS KEY PROBLEM
/*
 Q1.Write a SQL query to retrieve all columns for sales made on '202211-05:
*/
SELECT
    *
FROM retail_sales
WHERE sale_date = '2022-11-05';

/*
Q2.Write a SQL query to retrieve all transactions where the category is 'Clothing' a
nd the quantity sold is more than 3 in the month of Nov-2022
*/

SELECT
    *
FROM retail_sales
WHERE category = 'Clothing'
```

```sql
    AND
        quantity > 3
    AND
        sale_date BETWEEN '2022-11-01' AND '2022-11-30';

/*
Q3.Write a SQL query to calculate the total sales (total_sale) for each category.
*/
SELECT
    category,
    SUM(total_sale) AS TotalSales
FROM retail_sales
GROUP BY
    category
ORDER BY
    TotalSales DESC;

SELECT
    category,
    count(category) AS totalquantity
FROM retail_sales
GROUP BY category;

/*
 Q4.Write a SQL query to find the average age of customers who purchased items from the 'Beauty' category.
*/

SELECT
    category,
    ROUND(AVG(age)) AS AverageAge
FROM retail_sales
WHERE category = 'Beauty';

/*
 Q5.Write a SQL query to find all transactions where the total_sale is greater than 1000.
*/
SELECT
    transactions_id,
    total_sale
FROM retail_sales
WHERE total_sale > 1000;


/*
Q6.Write a SQL query to find the total number of transactions (transaction_id) made by each gender in each category.
*/
SELECT
    gender,
    category,
    COUNT(*) AS TotalTransactions
FROM retail_sales
GROUP BY 1,2
```

```sql
ORDER BY 2;


/*
Q7.Write a SQL query to calculate the average sale for each month. Find out best sel
ling month in each year:
*/
SELECT
    Year,
    Month,
    AverageSale
FROM (
        SELECT
            YEAR(sale_date) AS Year,
            MONTH(sale_date) AS Month,
            ROUND(AVG(total_sale),2) AS AverageSale,
            DENSE_RANK() OVER(PARTITION BY YEAR(sale_date) ORDER BY AVG(total_sale)
DESC) AS AverageRank
        FROM retail_sales
        GROUP BY 1,2
    )t
WHERE AverageRank = 1;


/*
Q8.Write a SQL query to find the top 5 customers based on the highest total sales:
*/
SELECT
    customer_id,
    COUNT(*) AS Purchase,
    SUM(total_sale) AS TotalSales
FROM retail_sales
GROUP BY
    customer_id
ORDER BY
    TotalSales DESC
LIMIT 5;

SELECT
    *
FROM retail_sales
WHERE customer_id = 1003;


-- Beauty category attracts younger customers
SELECT category,
       ROUND(AVG(age), 1) as avg_age,
       MIN(age) as min_age,
       MAX(age) as max_age
FROM retail_sales
GROUP BY category
ORDER BY avg_age;


/*
```

```sql
Q9.Write a SQL query to find the number of unique customers who purchased items from each category:
*/
SELECT
    category,
    COUNT(DISTINCT(customer_id)) AS uique_customers
FROM retail_sales
GROUP BY category;


/*
Q10.Write a SQL query to create each shift and number of orders
(Example Morning <12, Afternoon Between 12 & 17, Evening >17):
*/

WITH hourly_sale
AS
(
    SELECT
        *,
        CASE
            WHEN HOUR(sale_time) < 12 THEN "Morning"
            WHEN HOUR(sale_time) BETWEEN 12 AND 17 THEN "Afternoon"
            ELSE "Evening"
        END AS Shift
    FROM retail_sales
)
SELECT
    Shift,
    COUNT(*) AS TotalOrders
FROM hourly_sale
GROUP BY Shift;
```