**Task 1: Data Exploration and Cleaning**

**Load the dataset in SQL**

CREATE TABLE Customers (

customer\_id INT PRIMARY KEY,

name VARCHAR(100),

age INT,

gender VARCHAR(10));

CREATE TABLE Products (

product\_id INT PRIMARY KEY,

name VARCHAR(100),

category VARCHAR(50),

price DECIMAL(10, 2));

CREATE TABLE Sales (

transaction\_id INT PRIMARY KEY,

customer\_id INT,

product\_id INT,

date DATE,

quantity INT,

amount DECIMAL(10, 2),

FOREIGN KEY (customer\_id) REFERENCES Customers(customer\_id),

FOREIGN KEY (product\_id) REFERENCES Products(product\_id));

CREATE TABLE Inventory (

product\_id INT PRIMARY KEY,

stock\_count INT,

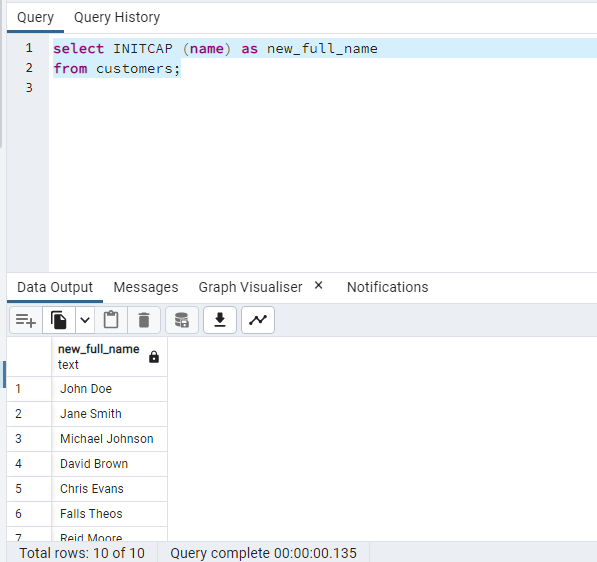
FOREIGN KEY (product\_id) REFERENCES Products(product\_id));

* **Cleaning**

**To capitalise the initial letter of the names of customers.**

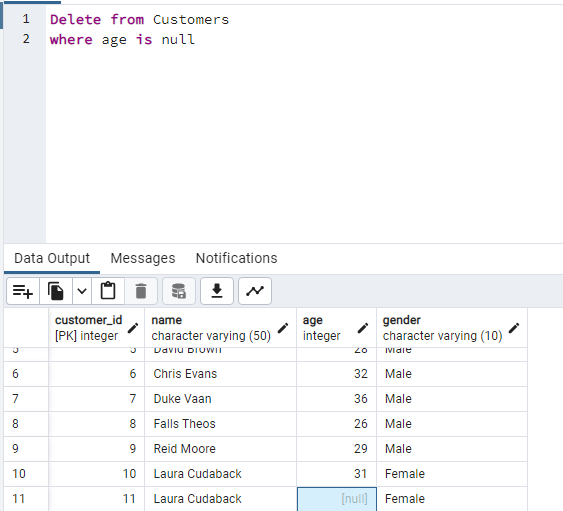
*select INITCAP (name) as new\_full\_name*

*from customers;*

****

*Delete from customer*

*where age is null*



**Task 2: Data Analysis**

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

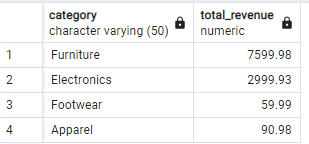
*select p.category, sum(s.amount) as total\_revenue*

*from Products p*

*join sales s*

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

*group by category*

**

1. **Determine the top 5 customers who have made the highest total purchases, considering**

**the customer's age and gender.**

*select c.name, c.age, c.gender, s.amount*

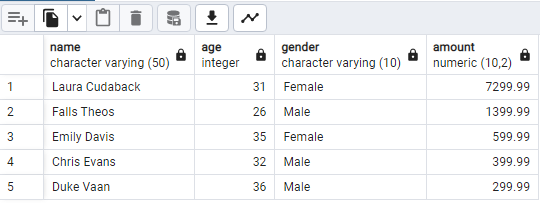
*from Customers c*

*join sales s*

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

*order by amount desc*

*limit 5*

**

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

**sold.**

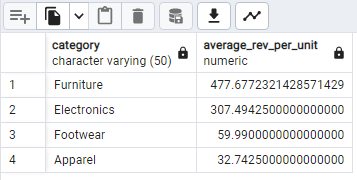
*select category, AVG(amount/quantity) as Average\_rev\_per\_unit*

*from products p*

*join sales s*

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

*group by category*

**

* *Furniture is the highest revenue category.*

**4. Analyse the inventory data and identify products that need restocking (stock count less**

**than a specified threshold).**

*select p.name, i.product\_id, i.stock\_count*

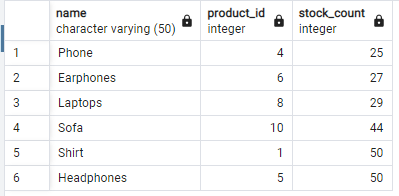
*from inventory i*

*join products p*

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

*where stock\_count < 55*

*order by stock\_count asc*

**

* *Threshold was below than 55 need restocking*

**Task 3: Advanced Analysis and Reporting**

**1. Write a SQL query to calculate the average age of customers for each product category**

*select p.category product\_category, Avg(c.age) as Ave\_Age\_of\_Customers*

*from customers c*

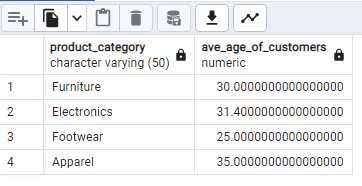
*join sales s*

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

*join products p*

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

*group by p.category*

**

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

**transaction amount.**

***select p.category, AVG(s.amount)***

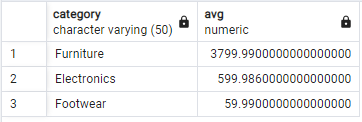
***from products p***

***join sales s***

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

***Group by p.category***

***limit 3***

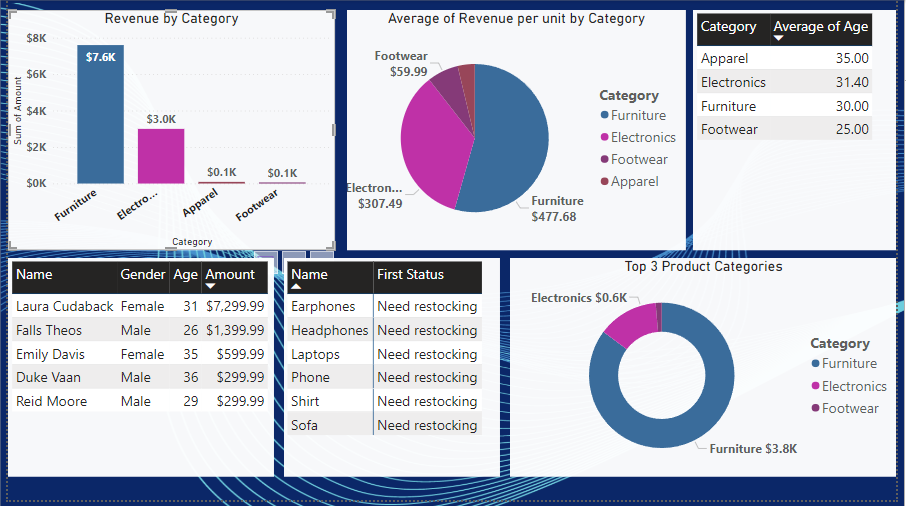
******

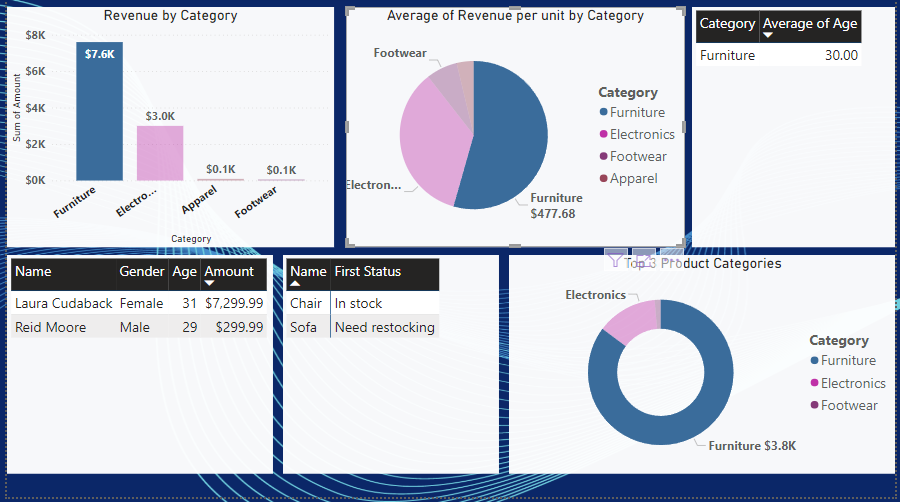
**Analysis Report**

I made an analysis of an ecommerce dataset, by studying the KPIs and using measures based on the dimensions of the data.

Calculated ***Total revenue*** generated by the company i.e. ***$10750.88*** with the ***Furniture*** as the highest profitable category. Along with this I made an analysis on the customers. The highest purchase was done by the ***Laura Cudaback*** i.e. ***$7299.99***

*Created the dynamic dashboard using PowerBI:*

**

**