Data Analyst Assignment

Dataset Description: For this assignment, you will be working with a dataset containing information about a retail company's sales and inventory transactions. The dataset includes the following tables:

- Customers:
  - customer_id: Unique identifier for each customer.
  - name: Customer's name.
  - age: Customer's age.
  - gender: Customer's gender.
- Products:
  - product_id: Unique identifier for each product.
  - name: Product name.
  - category: Product category.
  - price: Price of the product.
- Sales:
  - transaction_id: Unique identifier for each transaction.
  - customer_id: Unique identifier for each customer.
  - product_id: Unique identifier for each product.
  - date: The date of the transaction.
  - quantity: The quantity of products purchased in that transaction.
  - amount: The total amount spent in that transaction.
- Inventory:
  - product_id: Unique identifier for each product.
  - stock_count: The current stock count for each product.

Task 1: Data Exploration and Cleaning

1. Load the dataset into a SQL database and examine its structure.
2. Identify and handle missing values appropriately (e.g., remove rows, impute values).
3. Perform any necessary data cleaning operations to ensure data integrity and consistency.

Task 2: Data Analysis

1. Calculate the total revenue generated by the company for each product category.
2. Determine the top 5 customers who have made the highest total purchases, considering the customer's age and gender.
3. Identify the most profitable product category by calculating the average revenue per unit sold.
4. Analyze the inventory data and identify products that need restocking (stock count less than a specified threshold).

Task 3: Advanced Analysis and Reporting

1. Write a SQL query to calculate the average age of customers for each product category.
2. Write a SQL query to retrieve the top 3 product categories that have the highest average transaction amount.
3. Create a comprehensive report summarizing your findings from Task 2 and Task 3. Include relevant tables, charts, and explanations to present your analysis clearly.

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

Submission Guidelines:

1. Write the SQL queries and perform the analysis in your preferred SQL environment (e.g., MySQL, PostgreSQL).
2. Include comments in your SQL code to explain your approach and assumptions.
3. Provide the SQL code, along with the query results, visualizations, and your report, in a well-organized document (e.g., Word, PDF).
4. Submit your completed assignment document by the specified deadline.