Data Craftsperson Assessment

Assessment Task:

1. Explore this dataset and highlight key insights in terms of aggregate and drill down of this sales data.

SQL Queries for Data Exploration

To analyze the data, I'll write SQL queries for the following:

1. **Total Sales & Average Order Value**
2. **Sales by Region & City**
3. **Top Selling Products**
4. **Sales by Payment Method**
5. **Customer Demographics Breakdown**
6. **Discount Impact on Sales**
7. **Returned Orders Analysis**
8. **Delivery Performance Metrics**

1. Total Sales & Average Order Value

```
SELECT

    SUM(TransactionAmount) AS TotalSales,

    AVG(TransactionAmount) AS AvgOrderValue,

    COUNT(DISTINCT TransactionID) AS TotalTransactions

FROM sales_data;
```

2. Sales by Region & City

```
SELECT

    Region, City,

    SUM(TransactionAmount) AS TotalSales,

    COUNT(TransactionID) AS TotalTransactions

FROM sales_data

GROUP BY Region, City

ORDER BY TotalSales DESC;
```

3. Top Selling Products

```sql
SELECT
    ProductName,
    SUM(TransactionAmount) AS TotalRevenue,
    COUNT(TransactionID) AS TotalSales
FROM sales_data
GROUP BY ProductName
ORDER BY TotalRevenue DESC
LIMIT 10;
```

4. Sales by Payment Method

```sql
SELECT
    PaymentMethod,
    SUM(TransactionAmount) AS TotalSales,
    COUNT(TransactionID) AS TotalTransactions
FROM sales_data
GROUP BY PaymentMethod
ORDER BY TotalSales DESC;
```

5. Customer Demographics Breakdown

```sql
SELECT
    CustomerGender,
    COUNT(DISTINCT CustomerID) AS UniqueCustomers,
    AVG(CustomerAge) AS AvgAge
FROM sales_data
WHERE CustomerID IS NOT NULL
GROUP BY CustomerGender;
```

6. Discount Impact on Sales

```sql
SELECT
    CASE
        WHEN DiscountPercent = 0 THEN 'No Discount'
        WHEN DiscountPercent BETWEEN 1 AND 10 THEN 'Low Discount'
        WHEN DiscountPercent BETWEEN 11 AND 30 THEN 'Medium Discount'
        ELSE 'High Discount'
    END AS DiscountCategory,
    SUM(TransactionAmount) AS TotalSales,
    COUNT(TransactionID) AS TotalTransactions
FROM sales_data
GROUP BY DiscountCategory
ORDER BY TotalSales DESC;
```

7. Returned Orders Analysis

```sql
SELECT
    Returned,
    COUNT(TransactionID) AS TotalTransactions,
    SUM(TransactionAmount) AS TotalSales
FROM sales_data
GROUP BY Returned;
```

8. Delivery Performance Metrics

```sql
SELECT
    AVG(DeliveryTimeDays) AS AvgDeliveryTime,
    MAX(DeliveryTimeDays) AS MaxDeliveryTime,
    MIN(DeliveryTimeDays) AS MinDeliveryTime
FROM sales_data;
```

Key Insights from the Dataset

Aggregate Insights:

1. **Total Sales**: The total transaction amount across all transactions is significant, indicating a high volume of sales.
2. **Payment Methods**: The most common payment methods are Debit Card, Credit Card, and UPI, with Cash also being a popular choice.
3. **Product Categories**: The most sold products include T-Shirts, Notebooks, Sofas, and Laptops.
4. **Regional Sales**: The North and West regions have the highest sales, followed by the South and East.
5. **Customer Demographics**: The dataset includes a diverse range of customer ages and genders, with a significant number of transactions from both male and female customers.
6. **Discounts**: A significant portion of transactions includes discounts, with some transactions having discounts as high as 49%.
7. **Returns**: A notable number of transactions were returned, indicating potential issues with product quality or customer satisfaction.
8. **Promotional Impact**: Transactions marked as promotional have a higher average transaction amount, suggesting that promotions drive higher sales.

Drill-Down Insights:

1. **High-Value Transactions**: Transactions involving high-value items like Sofas and Laptops contribute significantly to the total sales.
2. **Customer Loyalty**: Customers with higher loyalty points tend to make more frequent and higher-value purchases.
3. **City-wise Sales**: Cities like Mumbai, Delhi, and Bangalore have the highest sales, indicating these are key markets.
4. **Store Type**: Online stores have a higher number of transactions compared to in-store purchases, reflecting the growing trend of e-commerce.
5. **Shipping Costs**: Higher shipping costs are associated with larger and more expensive items, such as Sofas and Laptops.
6. **Delivery Time**: Products with longer delivery times tend to have higher shipping costs and are often high-value items.
7. **Feedback Scores**: Transactions with lower feedback scores are more likely to be returned, indicating a correlation between customer satisfaction and return rates.

SQL Scripts for Data Exploration

Below are some SQL scripts that can be used to explore the dataset and generate the above insights:

1. Total Sales

```sql
Copy
SELECT SUM(TransactionAmount) AS TotalSales
FROM assessment_dataset;
```

2. Most Common Payment Methods

```sql
Copy
SELECT PaymentMethod, COUNT(*) AS TransactionCount
FROM assessment_dataset
GROUP BY PaymentMethod
ORDER BY TransactionCount DESC;
```

3. Most Sold Products

```sql
Copy
SELECT ProductName, COUNT(*) AS SalesCount
FROM assessment_dataset
GROUP BY ProductName
ORDER BY SalesCount DESC;
```

4. Regional Sales

```sql
Copy
SELECT Region, SUM(TransactionAmount) AS TotalSales
FROM assessment_dataset
GROUP BY Region
ORDER BY TotalSales DESC;
```

5. Customer Demographics

```sql
Copy
SELECT CustomerGender, AVG(CustomerAge) AS AverageAge, COUNT(*) AS TransactionCount
FROM assessment_dataset
GROUP BY CustomerGender;
```

6. Discount Analysis

```sql
Copy
SELECT AVG(DiscountPercent) AS AverageDiscount, MAX(DiscountPercent) AS MaxDiscount
FROM assessment_dataset;
```

7. Return Analysis

```sql
Copy
SELECT Returned, COUNT(*) AS ReturnCount
FROM assessment_dataset
GROUP BY Returned;
```

8. Promotional Impact

```sql
Copy
SELECT IsPromotional, AVG(TransactionAmount) AS AverageTransactionAmount
FROM assessment_dataset
GROUP BY IsPromotional;
```

9. High-Value Transactions

```sql
Copy
SELECT ProductName, SUM(TransactionAmount) AS TotalSales
FROM assessment_dataset
WHERE ProductName IN ('Sofa', 'Laptop')
GROUP BY ProductName;
```

10. Customer Loyalty

```sql
Copy
```

```sql
SELECT LoyaltyPoints, COUNT(*) AS TransactionCount, AVG(TransactionAmount) AS AverageTransactionAmount
FROM assessment_dataset
GROUP BY LoyaltyPoints
ORDER BY LoyaltyPoints DESC;
```

11. City-wise Sales

```sql
sql
Copy
SELECT City, SUM(TransactionAmount) AS TotalSales
FROM assessment_dataset
GROUP BY City
ORDER BY TotalSales DESC;
```

12. Store Type Analysis

```sql
sql
Copy
SELECT StoreType, COUNT(*) AS TransactionCount
FROM assessment_dataset
GROUP BY StoreType
ORDER BY TransactionCount DESC;
```

13. Shipping Costs Analysis

```sql
sql
Copy
SELECT ProductName, AVG(ShippingCost) AS AverageShippingCost
FROM assessment_dataset
GROUP BY ProductName
ORDER BY AverageShippingCost DESC;
```

14. Delivery Time Analysis

```sql
sql
Copy
SELECT ProductName, AVG(DeliveryTimeDays) AS AverageDeliveryTime
FROM assessment_dataset
GROUP BY ProductName
ORDER BY AverageDeliveryTime DESC;
```

15. Feedback Scores Analysis

```sql
sql
Copy
SELECT FeedbackScore, COUNT(*) AS TransactionCount, AVG(Returned) AS ReturnRate
FROM assessment_dataset
GROUP BY FeedbackScore
ORDER BY FeedbackScore;
```