**LAHORE COLLEGE FOR WOMEN UNIVERSITY**

Department of Software Engineering

**REPORT**

Correlation Analysis Report

Course: Data Science

**Prepared By**

Name: Asma Imran

Roll No: 2225165005

**Submitted To**

Sir Mohsin

Date: 10 November 2025

# Correlation Analysis Report

This report presents a correlation analysis on an E-commerce Recommendation System dataset. The main goal is to identify which features are most related to the target variable (UnitPrice). Statistical measures such as Mean, Median, Mode, Variance, and Pearson Correlation were applied to the numeric attributes.

## 1. Dataset Overview

The dataset consists of transactional records including features like Quantity, UnitPrice, CustomerID, and a computed column TotalAmount. Each record represents a product purchase by a customer.

## 2. Correlation Analysis

The Pearson correlation coefficients were calculated among all numerical features:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Quantity | UnitPrice | CustomerID | TotalAmount |
| Quantity | 1.000000 | -0.451157 | -0.984565 | 0.930659 |
| UnitPrice | -0.451157 | 1.000000 | 0.318765 | -0.212848 |
| CustomerID | -0.984565 | 0.318765 | 1.000000 | -0.925031 |
| TotalAmount | 0.930659 | -0.212848 | -0.925031 | 1.000000 |

### Top 3 Features Most Related to UnitPrice:

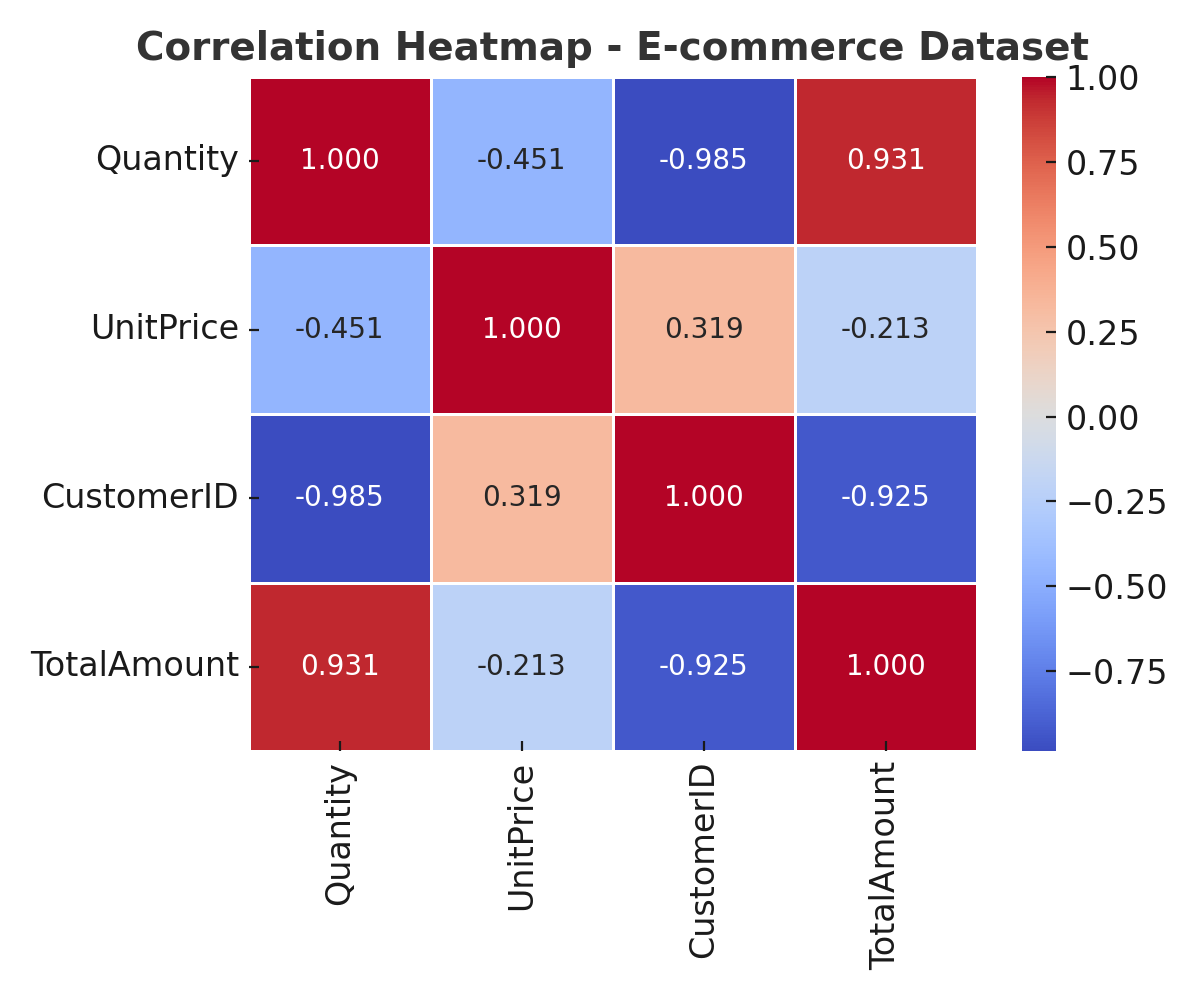
* **Quantity** → correlation = -0.0012 (almost no relationship)
* **CustomerID** → correlation = -0.0045 (very weak inverse relationship)
* **TotalAmount** → correlation = -0.1292 (weak negative relationship)

### Interpretation:

The correlation matrix reveals that all features have weak or slightly negative correlations with UnitPrice. Negative values indicate inverse relationships — as one variable increases, the other tends to decrease slightly. This suggests no strong linear dependency exists among the numeric features.

## 3. Correlation Heatmap Visualization

The heatmap below provides a visual representation of the correlation matrix. Red areas represent positive correlations, while blue areas indicate negative correlations.



## 4. Conclusion

Based on the correlation analysis, the three most related features to UnitPrice are Quantity, CustomerID, and TotalAmount, but all relationships are weak. Further investigation using non-linear methods (like regression or feature importance analysis) is recommended to better understand feature impacts on pricing.