Summary of Key Insights – Data Cleaning & Preprocessing

1. **Missing Values Handled**
    - Critical fields such as order_id and user_id were checked for missing values.
    - Rows with missing IDs were removed to maintain data integrity.
    - Missing numeric values (e.g., price, quantity) were filled using the **median** to reduce the effect of outliers.
2. **Duplicate Records Removed**
    - Duplicate transactions were identified using order_id and user_id.
    - **X duplicate rows** (replace with actual number if known) were removed to avoid double-counting in sales and revenue metrics.
3. **Data Types Standardized**
    - Date columns such as order_date were successfully converted to datetime format.
    - Columns like price and quantity were converted to numeric types for proper calculations.
    - Categorical values like category were lowercased and stripped of whitespace for consistency.
4. **Inconsistencies Resolved**
    - Negative values in price and quantity were found and removed as they are logically incorrect.
    - Variations in product category naming (e.g., "Electronics", "electronics") were standardized.
5. **Outliers Detected and Removed**
    - Outliers in price and quantity were identified using the Interquartile Range (IQR) method.
    - Rows with extremely high or low values were removed to ensure realistic insights in analysis.
6. **New Columns Created**
    - A new column total_revenue was added to compute overall transaction value (Price × Quantity).
    - Date-based features (year, month, day_of_week) were extracted to support time-series analysis.
7. **Final Dataset Ready**
    - The cleaned dataset contains **N rows and M columns** (replace with actual numbers).
    - It was saved as ecommerce_data_final_cleaned.csv for further analysis and modeling.