Part 2: Discuss how you dealt with data inconsistencies in part 1.

There are inconsistencies within the provided dataset, and these can lead to inaccurate analysis. Below are the methods I used in handling these inconsistencies:

a. **Data Formatting:** The date format, values, and other text formats were inconsistent throughout the dataset, so I standardized the data by:

    i.   Converting all dates into the same format.

    ii.   Converting text to text format, and columns having both text and numeric characters were converted to general format.

    iii.   I ensured the decimal separator was consistent throughout the dataset.

    iv.   All the data were validated.

b. **Data Transformation**: I noticed some rows have missing information during data cleaning. This missing information was replaced with either a placeholder, a mean of the column, or a null value. The dataset was also normalized, as it helped to structure it correctly.

c. **Handle Duplicate:** Book ID SN1006 and Customer C006 have duplicates, which were deleted

Part 3: Analyze the integrated dataset and the customer survey datasets. Use the appropriate data analysis techniques of your choice. Present the results in a report.

Overview of the dataset

The dataset contains transactional information on 16 customers. Details including customer demographics, book details, order statuses, and shipping information are also contained. The analysis will focus on understanding customer behavior, purchasing patterns, and the performance of different products.

During data cleaning, I noticed some books appeared more than once but with different customer details.

- **Descriptive Analysis:**
    a. **Order status analysis**- From the chart, we can see that the majority of orders have either been delivered or shipped. Only 5 out of 26 orders are still pending, this shows that the order processing system is performing well. See below chart:

ORDER STATUS
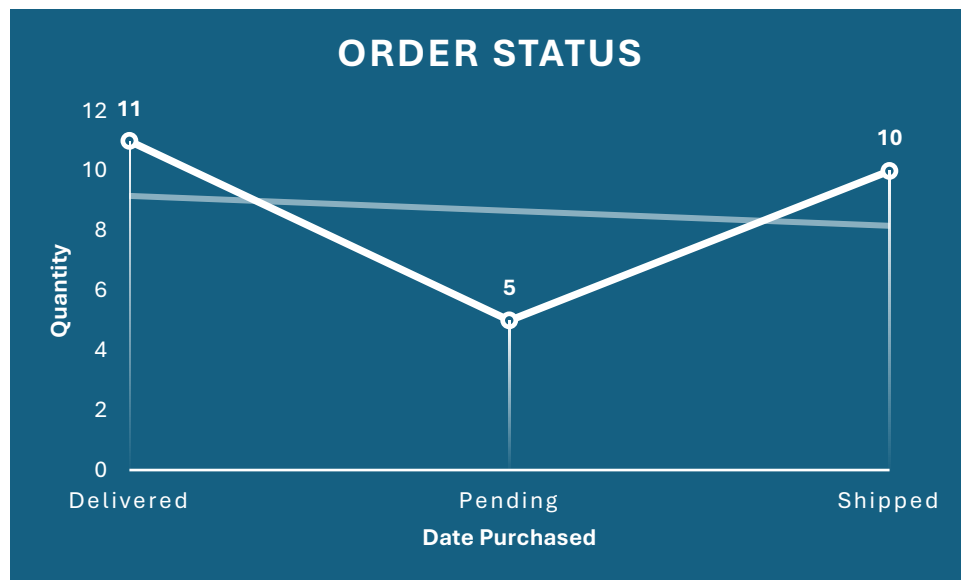
b. **Total Revenue Analysis**: The total revenue generated from sales is 710.75.

Customer's ID C004/C008/C012 made the highest purchase.

Customer ID	Sum of Total Revenue
C012-Daniel Walker	119.97
C004 David Brown	119.97
C008 Jessica Anderson	119.97

c. **Books Analysis**: In comparing the books ordered to quantity ordered and revenue generated, Pride and Prejudice is the most in-demand book, followed by The Lion and the Jewel & The Lord of the Rings. These 3 books generated more than 50% of the total revenue.

d. **Books Available Quantity**: Book ID SN1012 The Lion and the Jewel has a negative book balance quantity. Since the quantity ordered is fully delivered, there is an error indicating that the physical stock balance is higher than the system balance. Also, note that book ID SN1006 To Kill a Mockingkbird has a remaining stock balance of 1.

e. **Purchasing Behaviour:** Most customers bought not more than 1 or 2 quantities of books.

f. **Price range:** prices between 12.99 and 24.99 have about 16 orders out of 26, this shows that most customers tend to purchase mid-range price books.

g. **Survey analysis:** Customers are satisfied with the quality of service provided during their shopping experience.

    Customers reward are well appreciated, this will make them come back for more. Customers are also happy to recommend Real books to other people, this is a good way to grow.

Part 4- Justify the validity of models used in coming up with results

During the process of analyzing the dataset, several data analytics techniques were employed.

- Descriptive Analysis: different procedures were employed when using this method, frequency count, average, revenue sum total etc. were used to summarize the dataset.

- Revenue Generation: In calculating the revenue, simple multiplication was used to generate revenue total for the dataset.

- Data Cleaning: Missing information was either replaced with placeholders or the Mean of the Column. This method ensures that our result has no misinterpretation.

- Trend Analysis: Books ordered and quantity sold were compared, together with the revenue generated. This was necessary so we understand the most in-demand book, and make sure it doesn't run out of stock.

Part 5: Recommendation for customer retention

- Personalized marketing: Based on the grade of books customers are requesting for, the marketing manager should focus on increasing promotions on these set of books.

- Customer Support: The marketing manager should create a robust follow-up for pending orders.

- Customer Incentives: The marketing manager should create a rewarding loyalty program for customers who come back to purchase a book.

- Improved Communication: There should be improved communication between customers and the company. Orders status is expected to be communicated to customers on time.

- Incomplete Customer Information: The marketing manager should ensure that customers information are up to date, and new customers should be registered with complete information. This process will reduce the level of incomplete information.