

Customer Dataset:

- a) In customer dataset most the customers are from the south America with approx of 59.
This shows that the maximum transaction are from the south America region.
- b) Most of the customers sell products in 2024.
- c) CustomerId is the primary key for customer table and it should not be null or duplicate.
- d) The transaction is carried out in 4 regions i.e. South America, Asia, North America, Europe.
- e) The least region of customer is Asia with almost of 43 customers.

Products Dataset:

- a) In product dataset the categories books and electronics have same values. Which means that they are equally sell.
- b) The Maximum price of the product is approx 400.
- c) The price of the product are lies between the range of 3 to 499 approx with no outliers.
All the values are at the same range.
- d) The price of the books are 292.21 and electronics are 275.35 but the count of these two categories are same.
- e) There are only four categories books, Electronics , clothing and Home décor with least selling of home décor things.

Transaction Dataset:

- a) Total value of the transaction are between 250-500 with the count of more then 150.
- b) Maximum quantity of the transaction is between 3.5 and 4 with the count of 250.
- c) The range of the total-value features is between 2 to less then 2000.
- d) There is high correlation between the total-value and quantity with the value of 60% approx.
- e) After joining all the tables the graph of quantity distribution by category shows that the maximum customer purchase the books.
- f) The maximum revenue by product is from ComfortLiving Mystery Book.