Cleaning and processing data from Superstore data

1. Importing customers-superstore.csv file.
2. Activate view- column quality and column distribution. Change column profiling based on entire data set from top 1000 rows. General outline of data shows 9 columns x 733 rows
3. Track of activities
    - Customer ID and customer name are unique. Each customer is mentioned only once.
    - Customer name first two names initials are used in forming the customer ID
    - Segment is with 3 distinct values - consumer, corporate and home office.
    - Age column numbers range from 70 to 18. Converting to an age brackett using a command. Use a conditional column to define age bracket in 4 bins- Young Adult, Adult, Mid and Senior. Change the datatype to Text. Re-order column next to age column.
    - Country (text) unique value 1
    - City (text) 148 unique and 252 distinct.
    - State (text) 41 distinct values.
    - Postal code is in number datatype. Converted to text data type.
    - Region is a category based text datatype.
4. Importing Sales-superstore.csv file.
5. Activate view- column quality and column distribution. Change column profiling based on entire data set from top 1000 rows. General outline of data shows 13 columns x 9994 rows
6. Track of activities
    - Ensure that there are no duplicate rows in the entire table using top left icon.
    - Orderline is a number based identifier.
    - OrderID is a text based identifier. Each orderID can have multiple line items.
    - OrderDate is shown as a text. Changed to date format. Led to 59% error. On checking the Error, it is seen that month and day is notproperly recognized.
        - Split the order date by delimiter
        - Add column by example to rejoin and change datatype to date. This cleaning was successful.
        - Create separate columns for month and year for Order date.
    - Do similar transformation for the ship date which is a text data type.
    - Create ordertoshipdays as difference between orderdate and ship_date. Convert to numerals.
    - Shipmode is a categorical text with 4 dsitinct values.
    - CustomerID is a primary key whcih can be used to link customers and sales csv files.
    - ProductID is a culmination of category, subcategory and productcode.
    - Sales is mentioned in USD but in text. Converted to a usable form. Delete text sales column.
    - Quantity is already in numeric format.
    - Discount is converted to percentage.
    - Profit is converted from decimal to fixed decimal.
7. Now merge both the tables into a new file thathas customer ID and customerID as common column with RIGHT OUTER JOIN so that all orders are taken into consideration. Open the table with only required columns.

Apply and save the changes for visualization. Processing complete as file is ready to visualize.