Case Study 1:

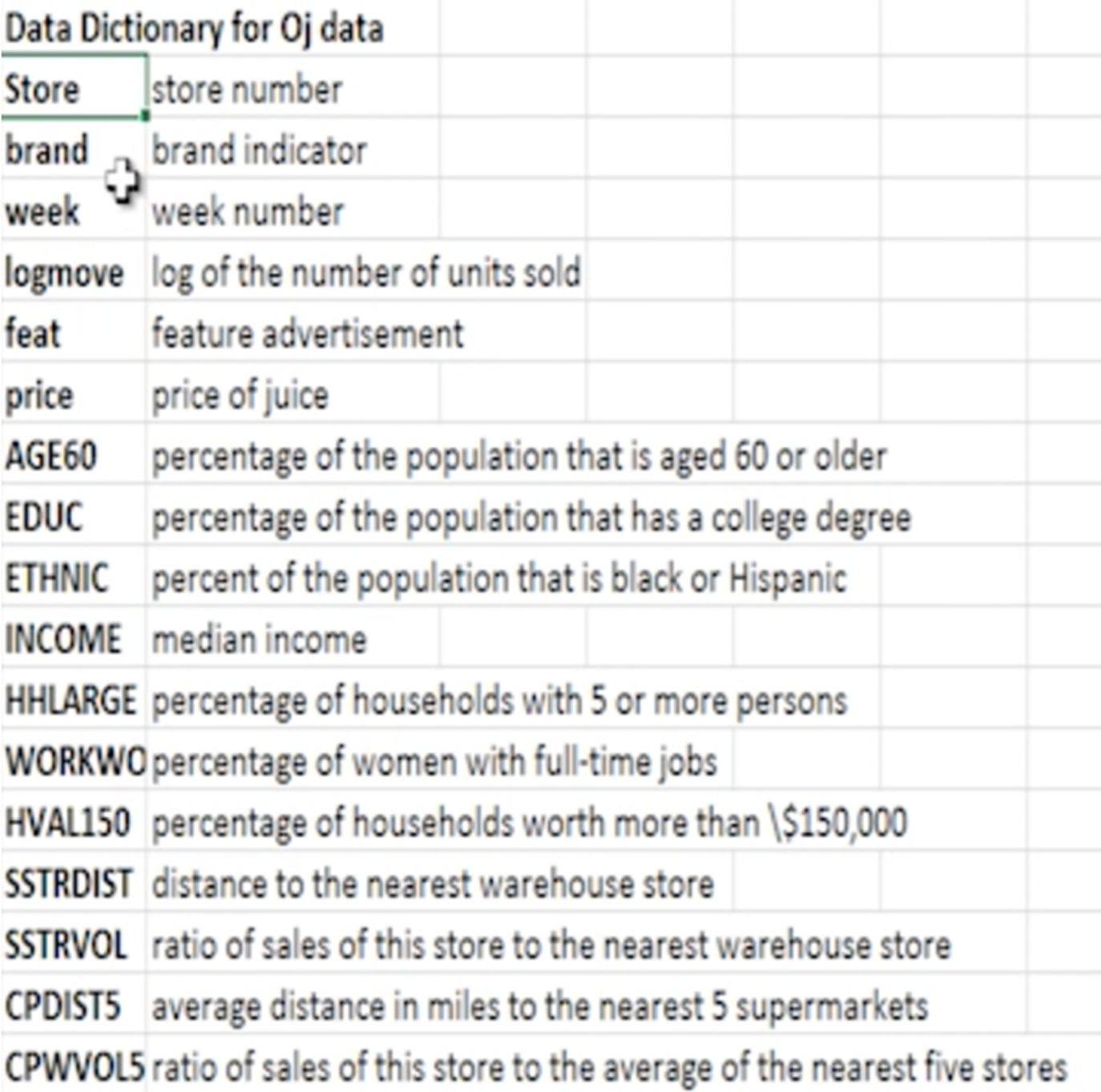
Identify the anomalies in the given dataset “Ecommerce\_data” and impute/treat them with appropriate measures.

Case study 2 :

OJ Case Study

The data contains 28947 purchases where the customer either purchased either Tropicana or Minute Maid or dominicks Orange Juice. A number of characteristics of the customer and product details are recorded.

Here is the dictionary of the data



Case Study

Please perform EDA on the dataset given and answer the below questions

1. Find the dimensions of the oj data set
2. Find the structure of the data set
3. Find out the column names in the data set
4. Describe the data set
5. Update the column names of the data set, such that every column name is in lowercase
6. Fetch the first-row 3rd column from the data set
7. Fetch the first, second and Third columns of the oj data frame
8. Fetch the first, second, eighth and the 456th rows of the 1st, third and the sixth columns of the data frame
9. Fetch the top 5 rows of the brand column
10. Fetch top 5 rows of the brand, week and feat details
11. Fetch the details of all distinct stores
12. Fetch all the observations for Tropicana brand using query function
13. Fetch the columns store, brand and price for observations that belong to the brand tropicana
14. Fetch bottom 5 observations for those who have bought Tropicana or dominics
15. Fetch the income, brand, price observations with Tropicana brand without the feature advertisement (feat ==0)
16. Remove the column normInc from the data frame
17. Fetch the observations for the tropicana brand excluding the store, price and income fields
18. Sort the Data in the increasing order of the week
19. Sort the data in the decreasing order of Income and increasing order of week
20. Find the mean of the juice price for each brand
21. Find the mean, sum, std of price for every brand
22. For every brand find the mean price and sum of income
23. Find the average income for each brand and at each store
24. Find:
    1. Mean and std deviation of the income
    2. For income greater than or equal to 10.5, find the mean income
    3. For each brand having price >=2.5 find the mean, median, sd of the log of income
25. Find the Cross tabulation of brands and feature advertisement