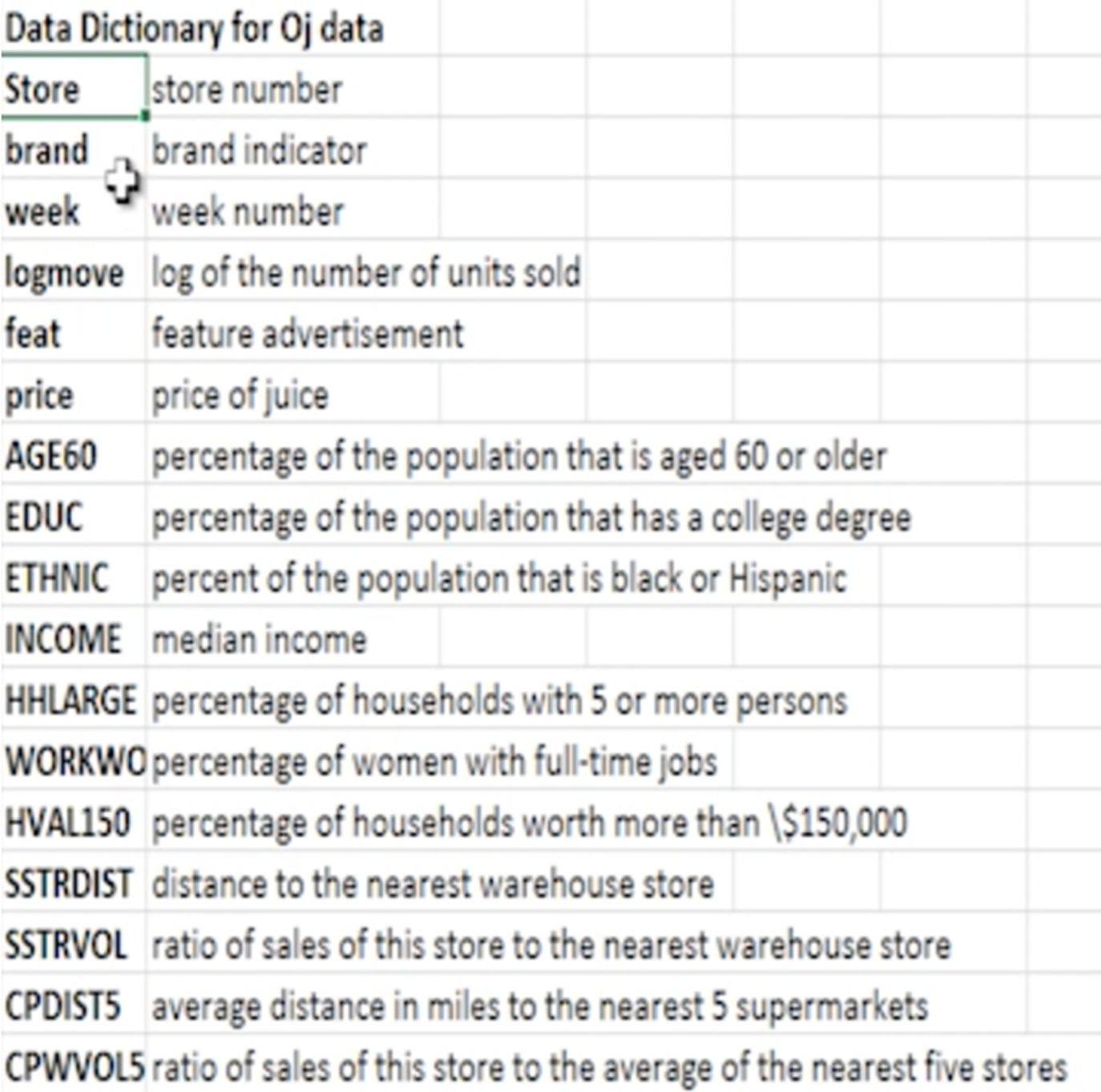
Case Study

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

Using this dataset answer the below queries to understand and clean data set

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
13. Fetch all the observations for Tropicana brand using query function
14. Fetch store, brand and price observations for the brand tropicana
15. Fetch bottom 5 observations for those who have bought Tropicana or dominics
16. Fetch the income, brand, price observations with Tropicana brand without feature advertisement
17. Add a new column to the dataframe normInc which is the normalized value of income
18. Remove the column normInc from the data frame
19. Fetch the observations for the tropicana brand excluding the store, price and income fields
20. Sort the Data in the increasing order of the week
21. Sort the data in the decreasing order of Income and increasing order of week
22. Find the mean of the juice price for each brand
23. Find the mean, sum, std of price for every brand
24. For every brand find the mean price and sum of income
25. Find the average income for each brand and at each store
26. 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
27. Find the Cross tabulation of brands and feature advertisement