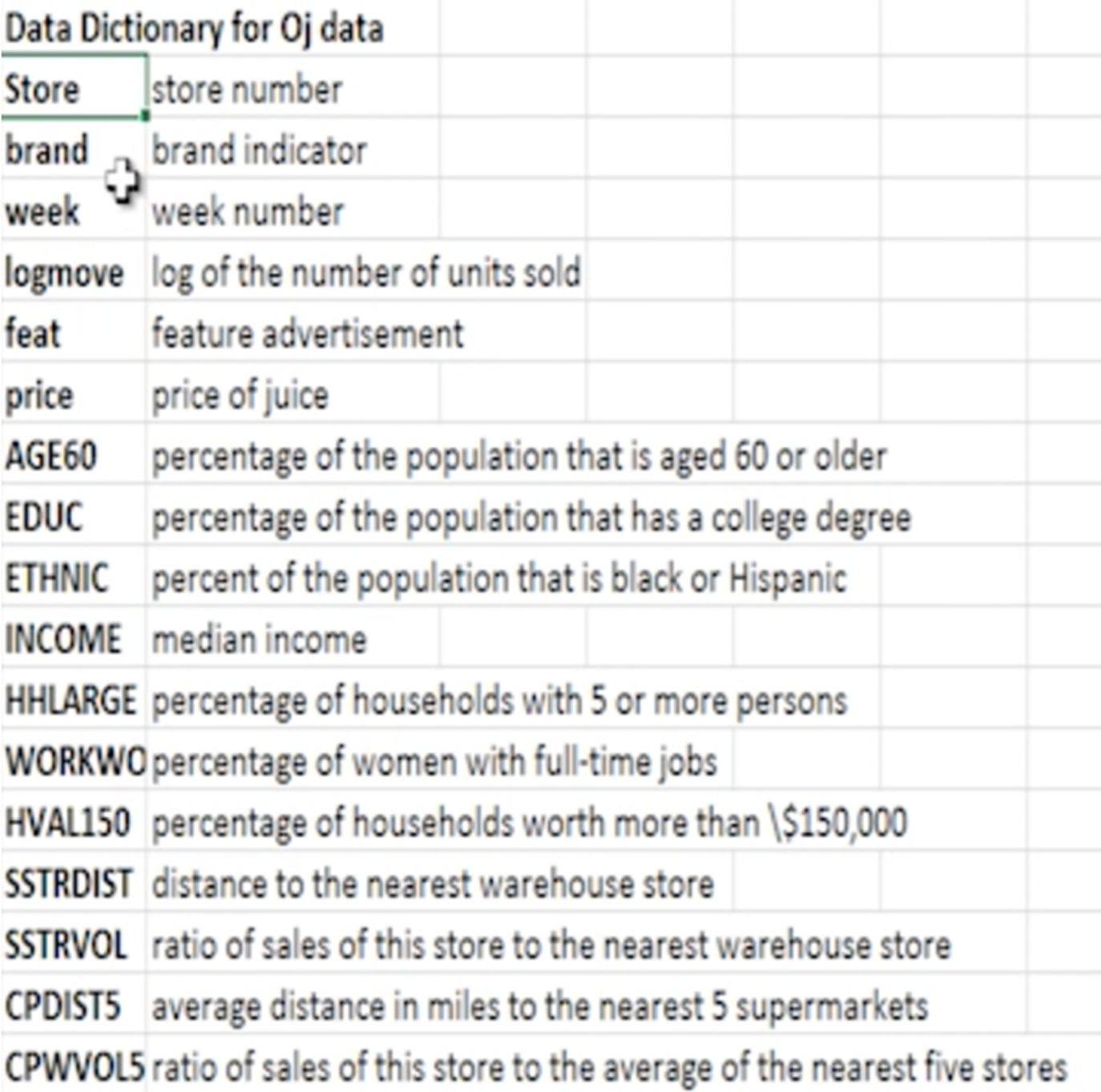
Case Study

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. Fetch the first row 3rd column from the data set
6. Fetch the first, second and Third columns of the oj data frame
7. Fetch the first, second, eighth and the 456th rows of the 1st, third and the sixth columns of the data frame
8. Fetch the top 5 rows of the brand column
9. Fetch top 5 rows of the brand, week and feat details
10. Fetch the details of all distinct stores
11. Fetch all the observations for Tropicana brand
12. Fetch all the observations for Tropicana brand using query function
13. Fetch bottom 5 observations for those who have bought Tropicana or dominics
14. Fetch the income, brand, price observations with Tropicana brand without feature advertisement
15. Add a new column in the dataset: logInc which is the logarithm of the income
16. Sort the Data in the increasing order of the week
17. Sort the data in the decreasing order of Income
18. Find the mean of the juice price for each brand
19. Find the average income for each brand and at each store
20. 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
21. Find the Cross tabulation of brands and feature advertisement