Problem Set 3

Bhavna Phogaat
ECON833: Computational Methods for Economists

Fall 2021

Objective

Create three visualization tools on a dataset.

Dataset

This is a panel dataset about **customer visit to stores** in 5 states spanning 6 months. There are 500 unique customer IDs with demographic information. The important variables are:

1. CUSTOMER_ID
2. DAY_NUMBER
3. DAY_FREQUENCY
4. LOC_NUM
5. AGE
6. GENDER_INPUT
7. NBR_OF_ADULTS
8. STATE
9. DISTANCE
10. ADD_MEMBERS

11. TOTAL_MEMBERS
12. DROPPED_MEMBERS

Figure

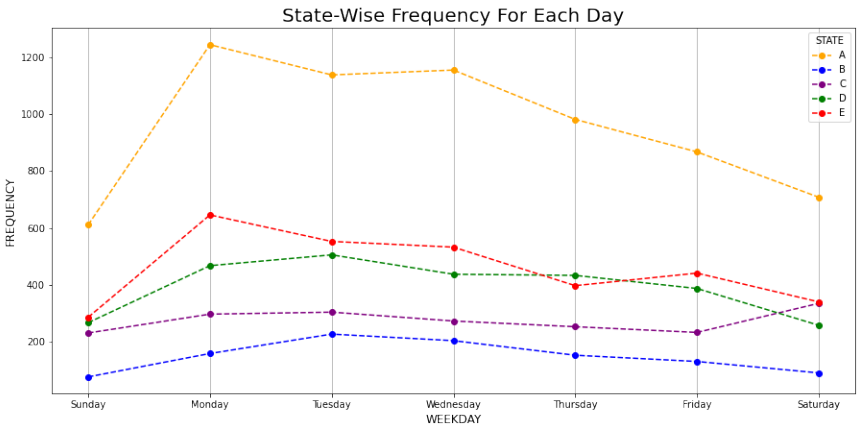

Figure 1: Shows the number of times customers visit the stores in all states. We can see the customer visit is highest for **State A** and on **Mondays**.

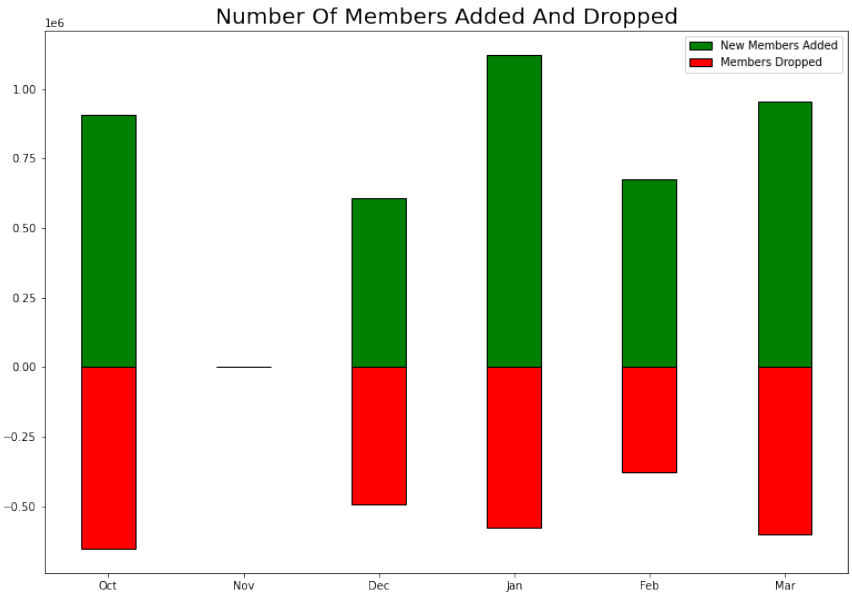

Figure 2: Shows the number of customers added and dropped in all months. No data is provided for November but we still keep this here for future references. We can see the addition of customers increase in the month of **January** and the least customer drop out is noticed in the month of **February**.

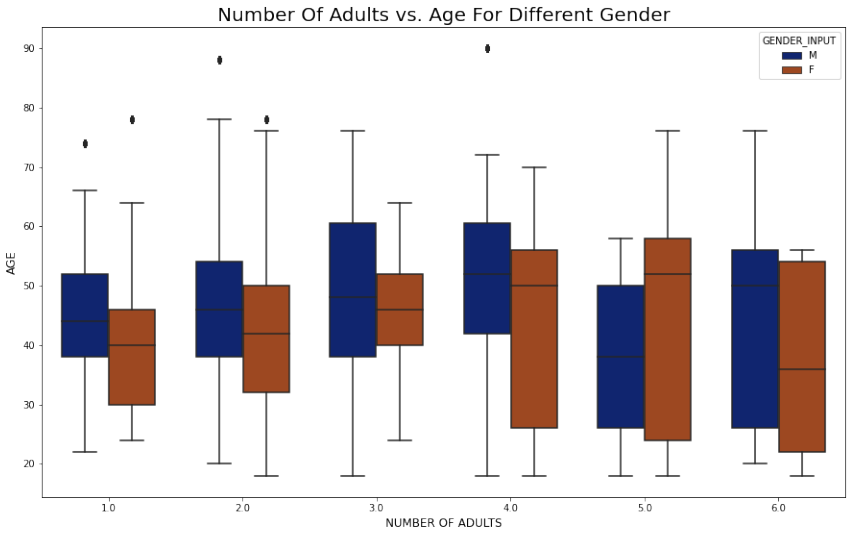

Figure 3: Is a boxplot highlighting 3 variables together. We can understand the customer demographic using this plot. For customers who have 3 or less adults in their household we see that their average age is 45. However, for customers with 4 or more adults in their household the average age is different for both genders.