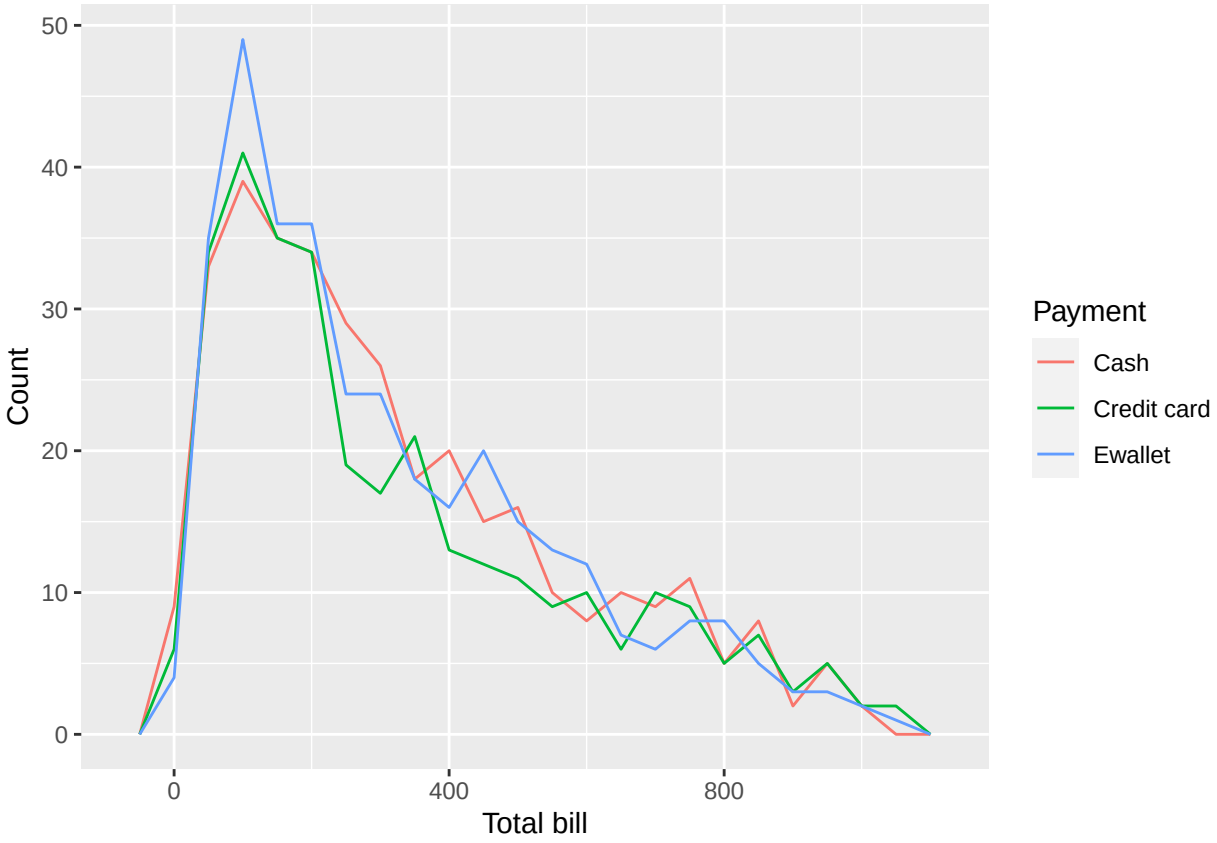
Supermarket Data Analysis

Preet Inder Pall Singh

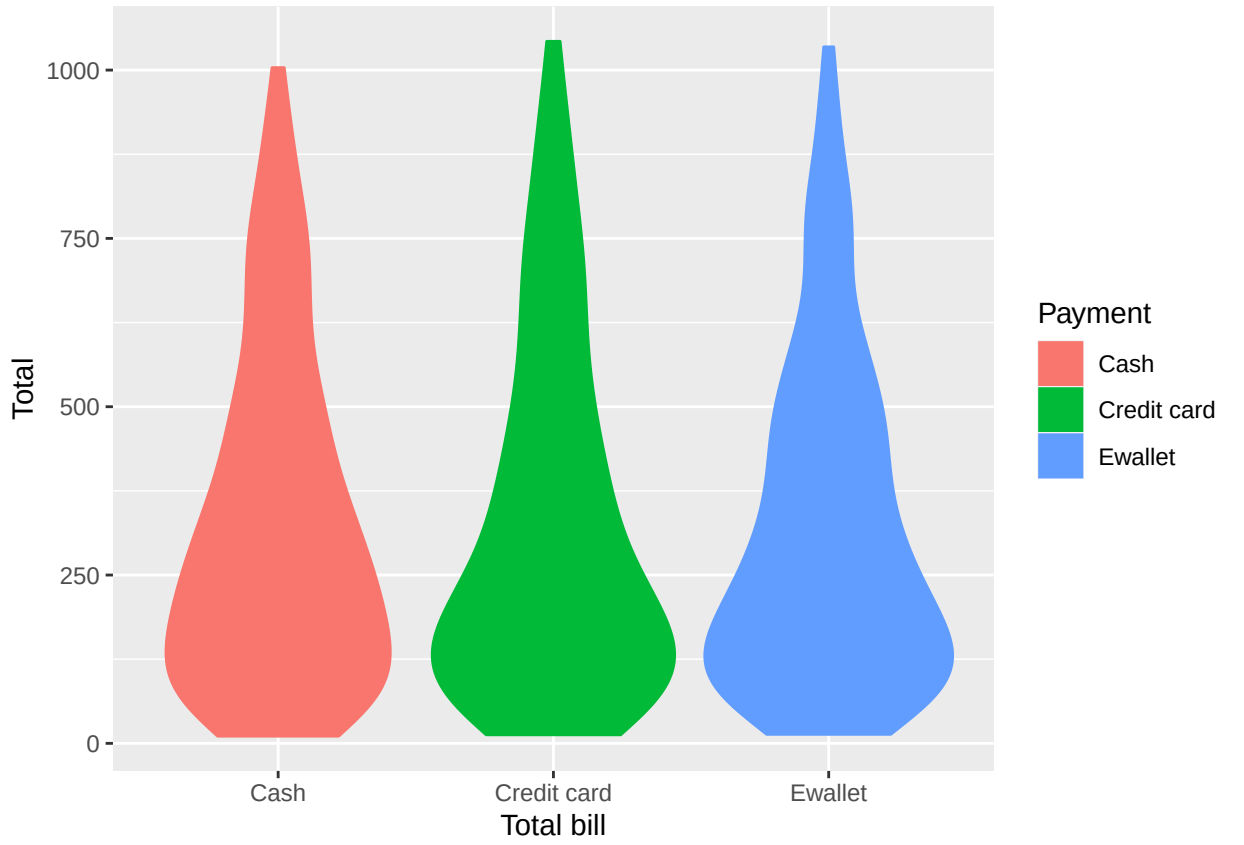
15/06/2021

Introduction

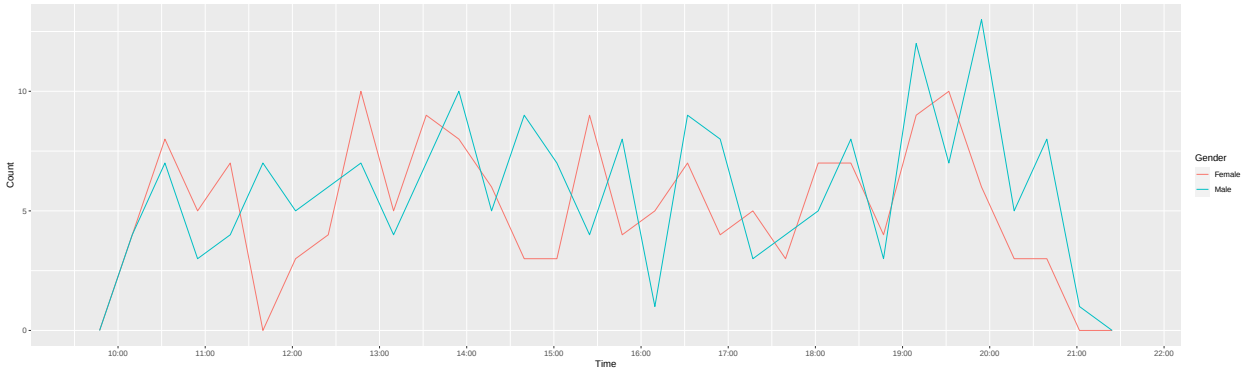
We will clear the console and import the library. The data is available from this link. The data set contains transition details of a supermarket chains in Myanmar, from three branches in three cities Yangon, Naypyitaw and Mandalay for 85 days. The data set contains information about the customer type, gender, mode of payment, city, branch, Product line, total, payment etc.
Our first analysis is how the payment techniques is impacted by cost.



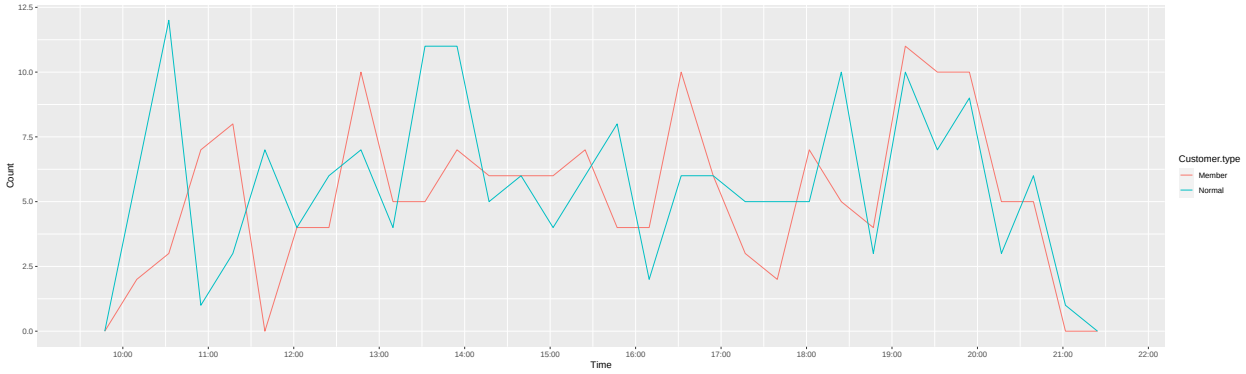
From the above two graph graph the mode of payment is fairly insensitive to cost, and we can validate our conclusion from the below graph. As a result we can conclude that mode of payment is not a suitable feature to better understand data set.



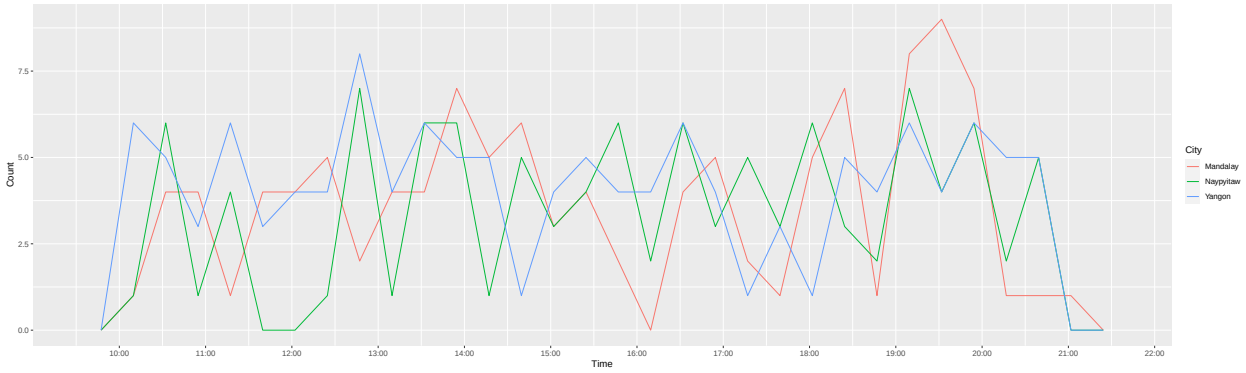
Time analysis for one month



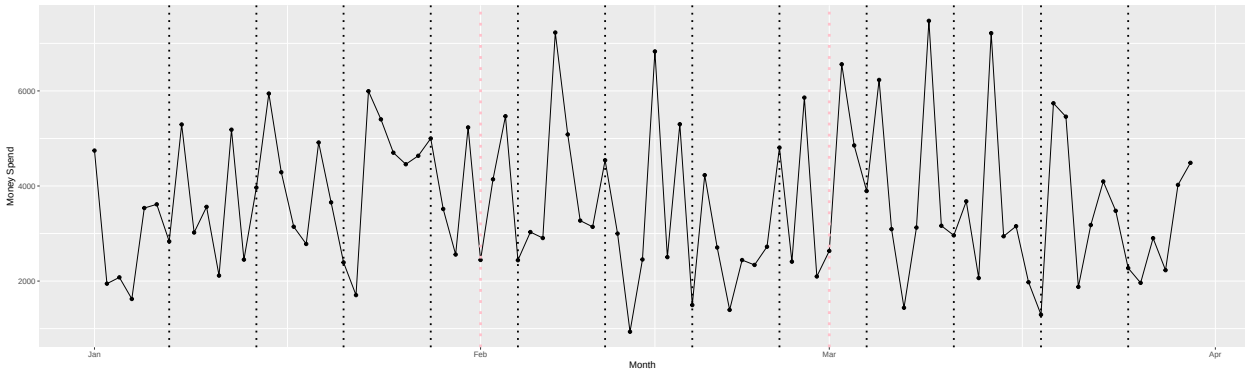
We now see how for the month of March, except for the time between 11:30 AM 12:30 PM when number of female decline, both male and female visit the super market at the same time. We now can clearly say that there is no difference in visting time between male and female customers, both gender visit the supermarkets at the same time except for 11:30 AM to 12:30 PM.

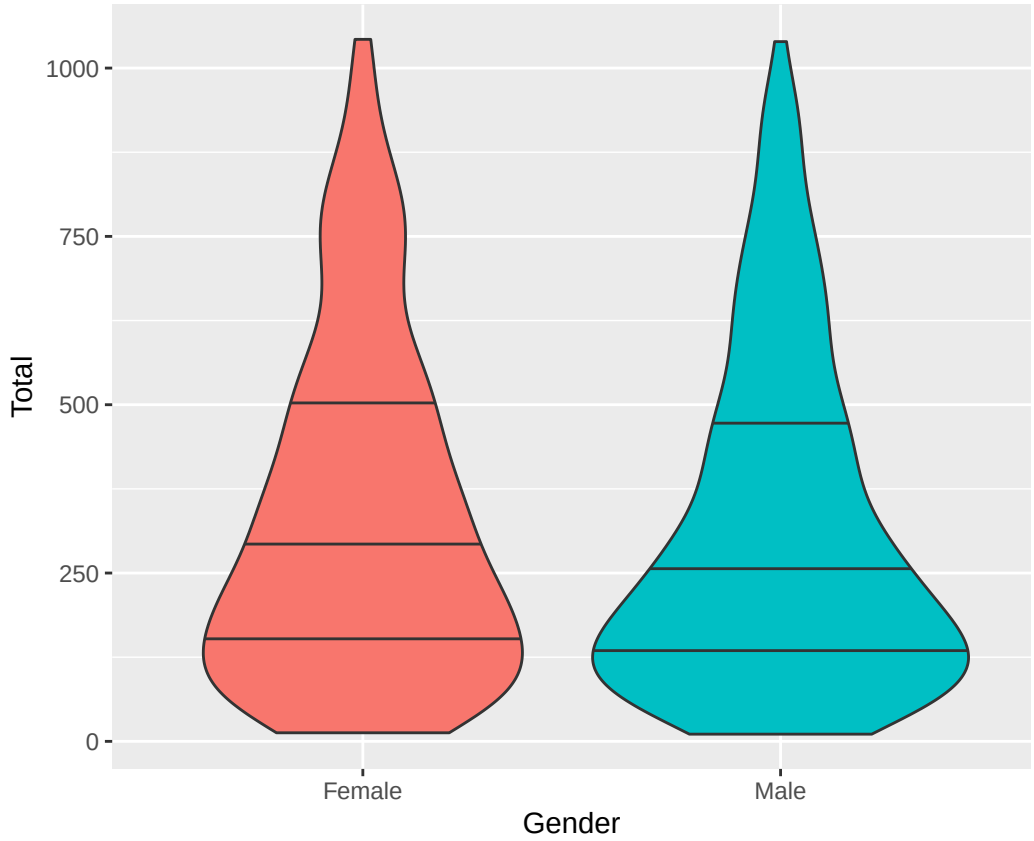


Similarly for members and normal customers, we see that no is no correlations between time and membership. We can clearly say that customer type no effect on visiting time of the customer.



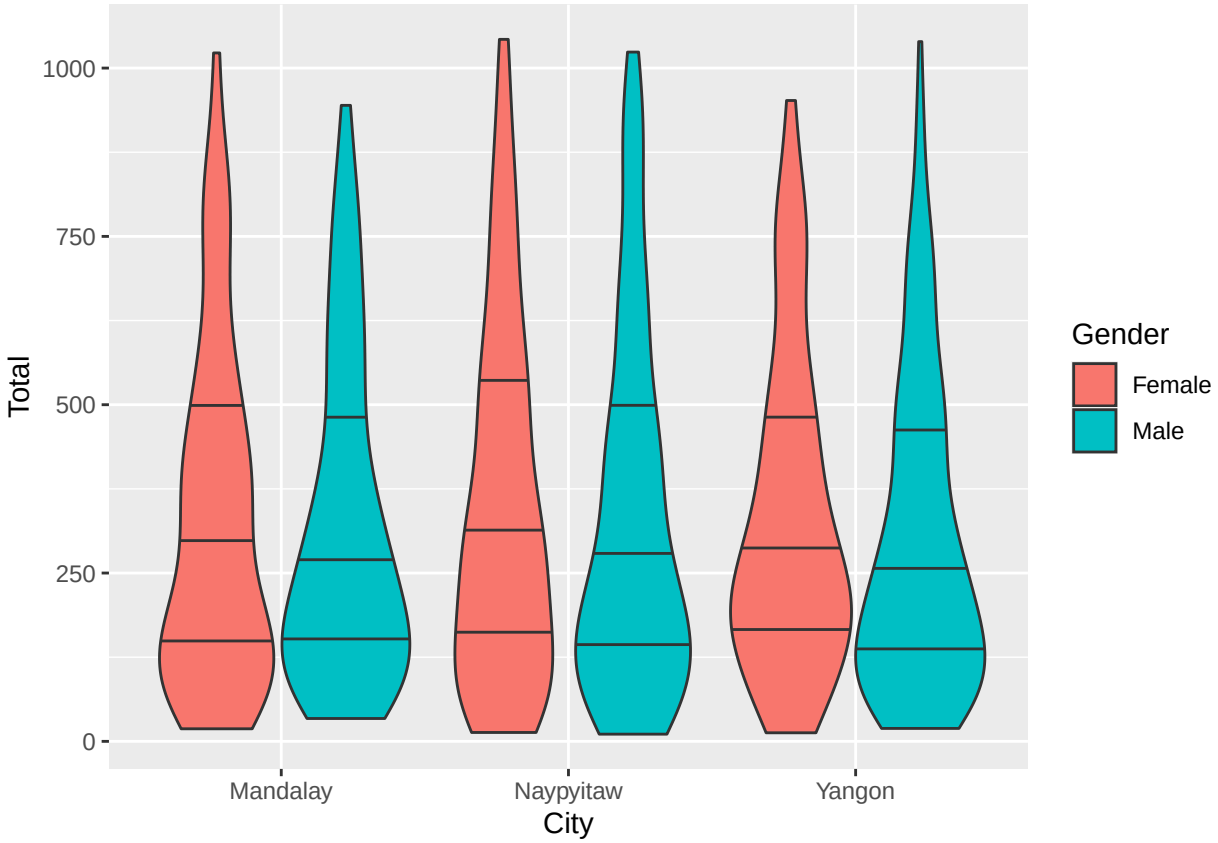
As we can see from the above graph, there is no correlation between the city and timing of customers visiting the super market. We will try to look at the total money spend depends on the day of the month.As we can wee from the below graph spending increase on the first week of a new month and a decrease on the last week of the month.



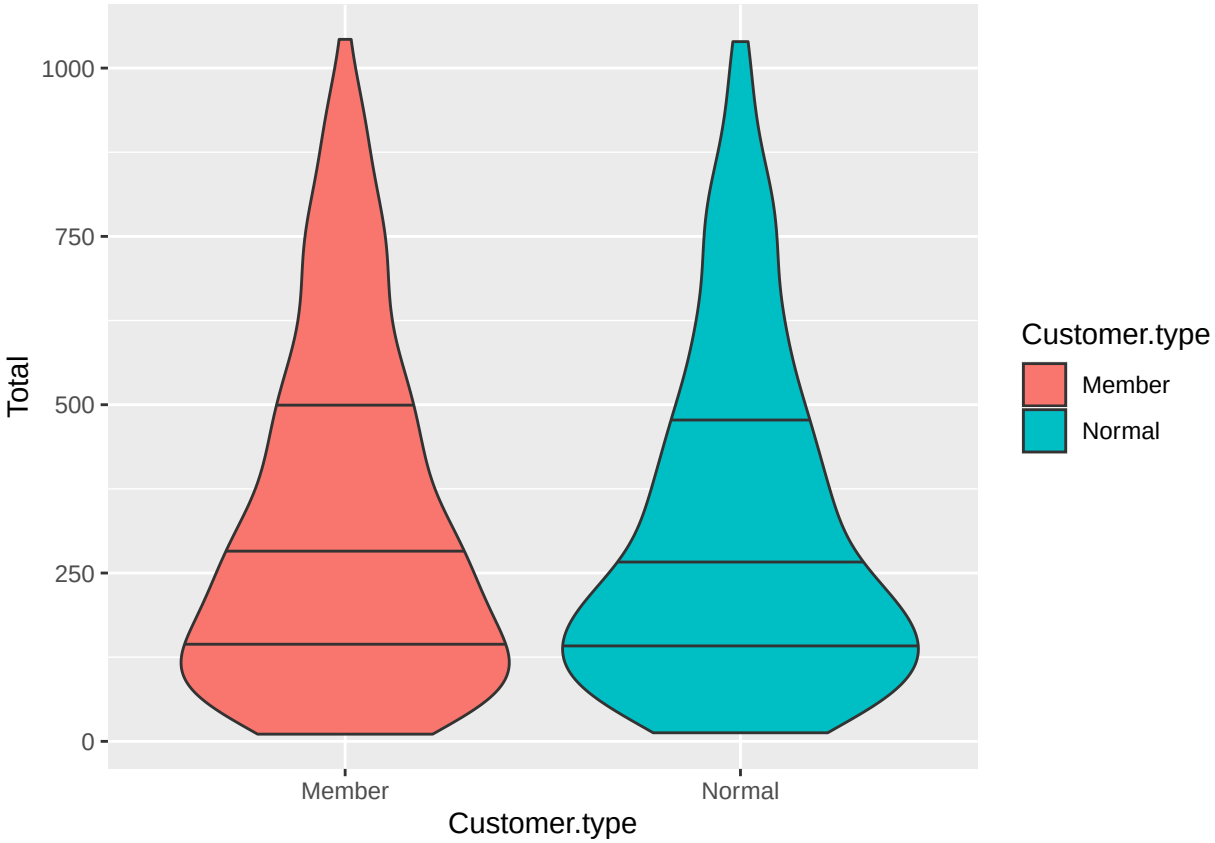


Catogory Analysis of dataset

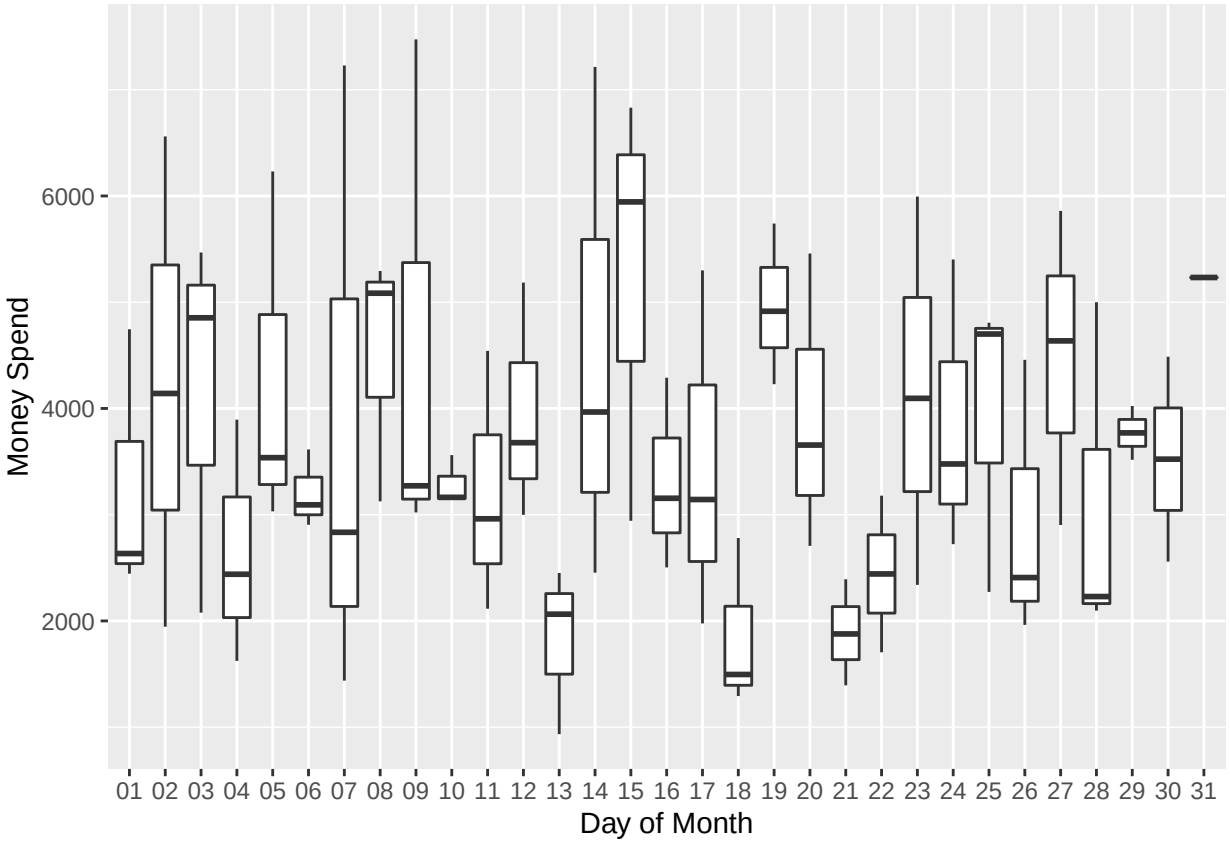
As we can see from the above and below graph, female have a higher 75% quantile and longer tail except for the city of Yagon where male have a longer tail but females still have higher 75% quantile. Otherwise there is negligible difference on spending on the bases of gender even across different cities.



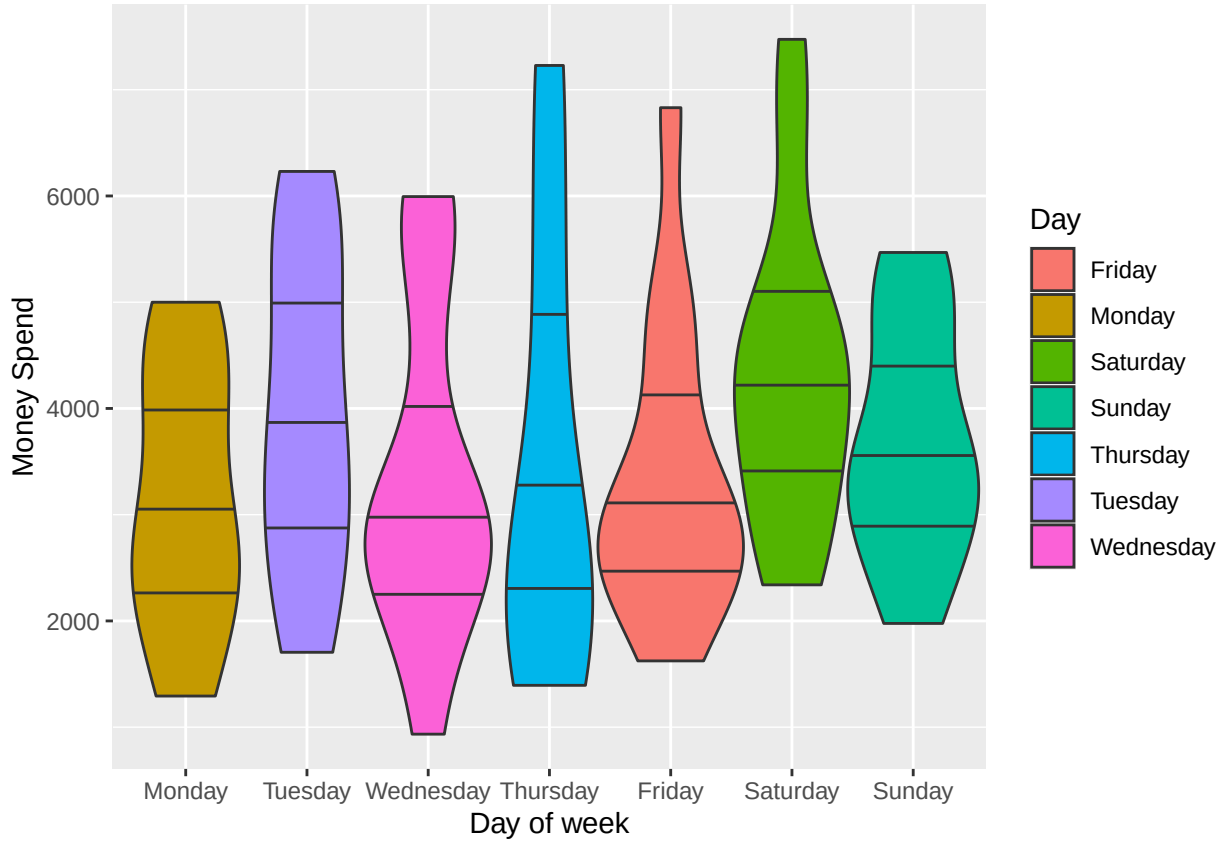
As we can see from the below graph, we can say that visually customer type have no effect on the money the customer. We can clearly say that both gender and customer type not useful to under customer behavior and segmentation

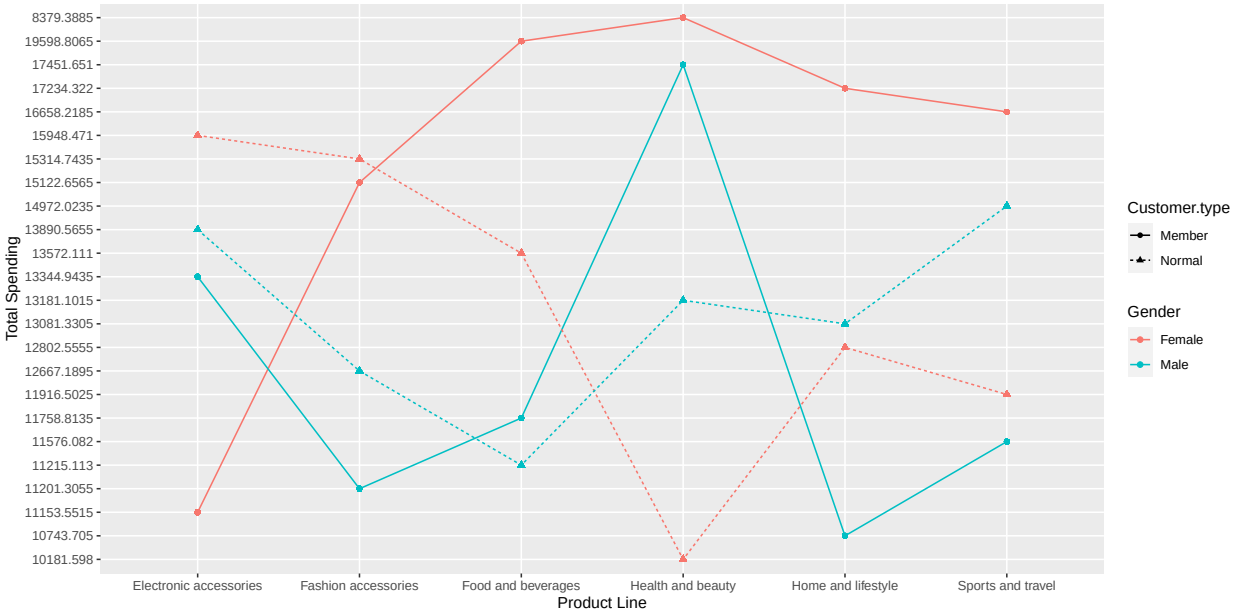


As we can see from the below graph, we see that for the first 15 days we generally have higher spending then last 15 days. This can be attribute as most people get thier salary on the first week of the month and hence they have more spending power.

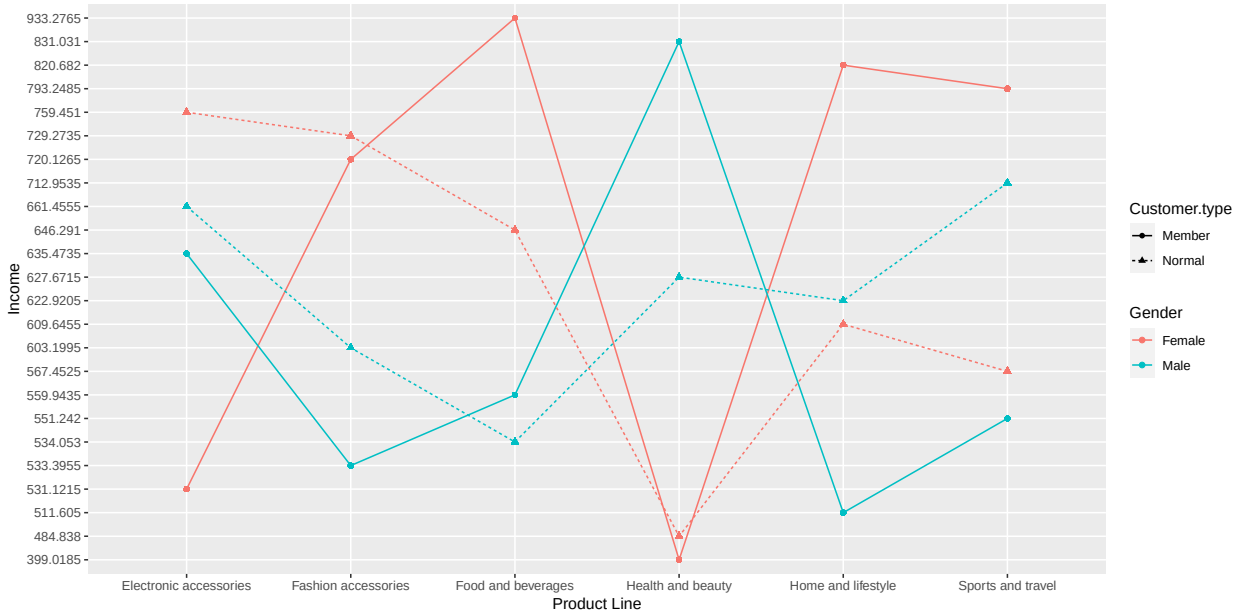


As we can see from the below graph, Saturday on average have the highest money spend but Thursday have a similar value for 75% percentile. Thursday have a very long tail, whereas tuseday and monday have close to a uniform distribution.

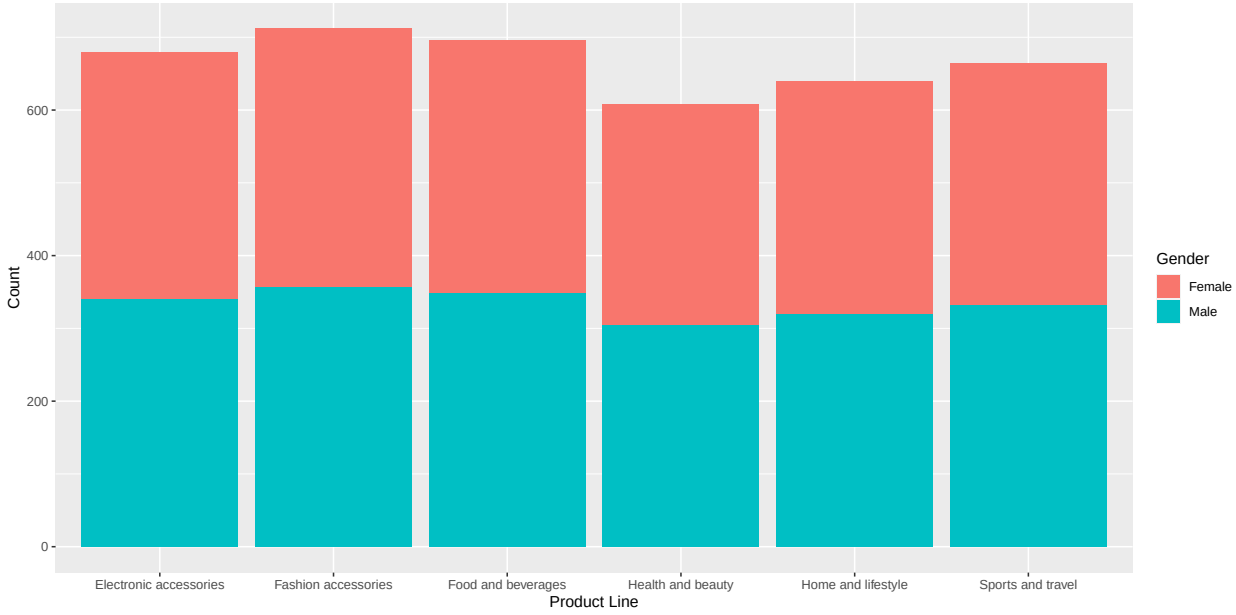


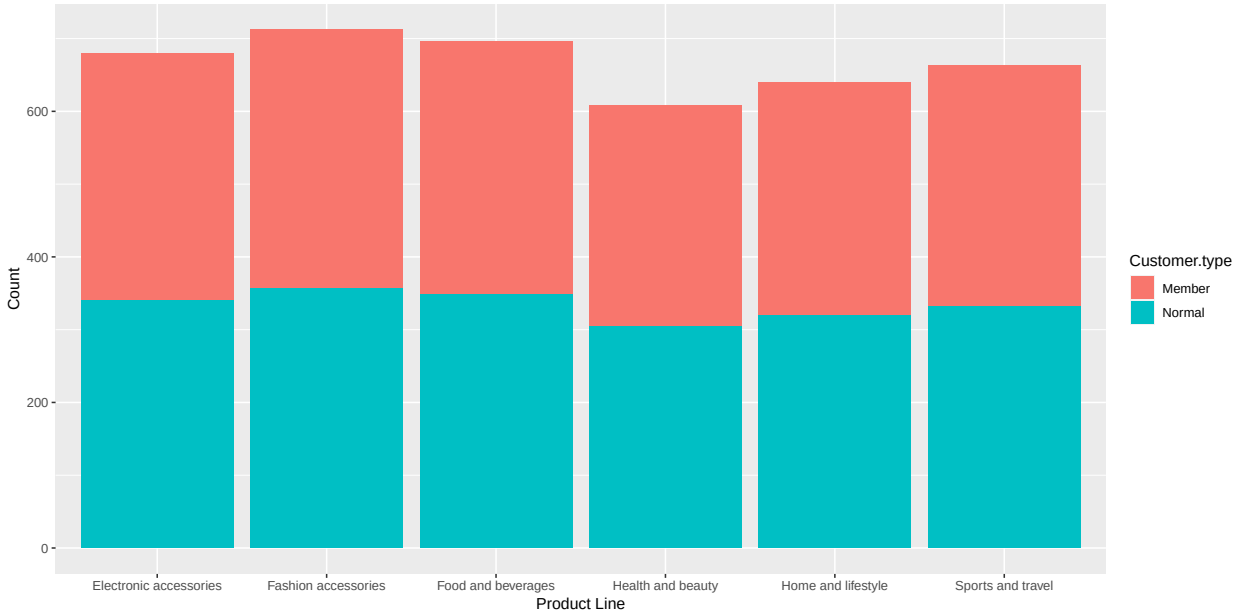


As we can see from the above graph female members, spend more on every product line except for electrical and fashion accessories. We can clearly see that they are actually 4 customer category- female member, female non member, male member and male non member.



We can more clearly see the difference in behavior in all four category form the above graph.





As we can see from the above two graph, using only gender or customer type alone will not allow us to understand customer behavior but using both feature together help us to better understand user behavior.