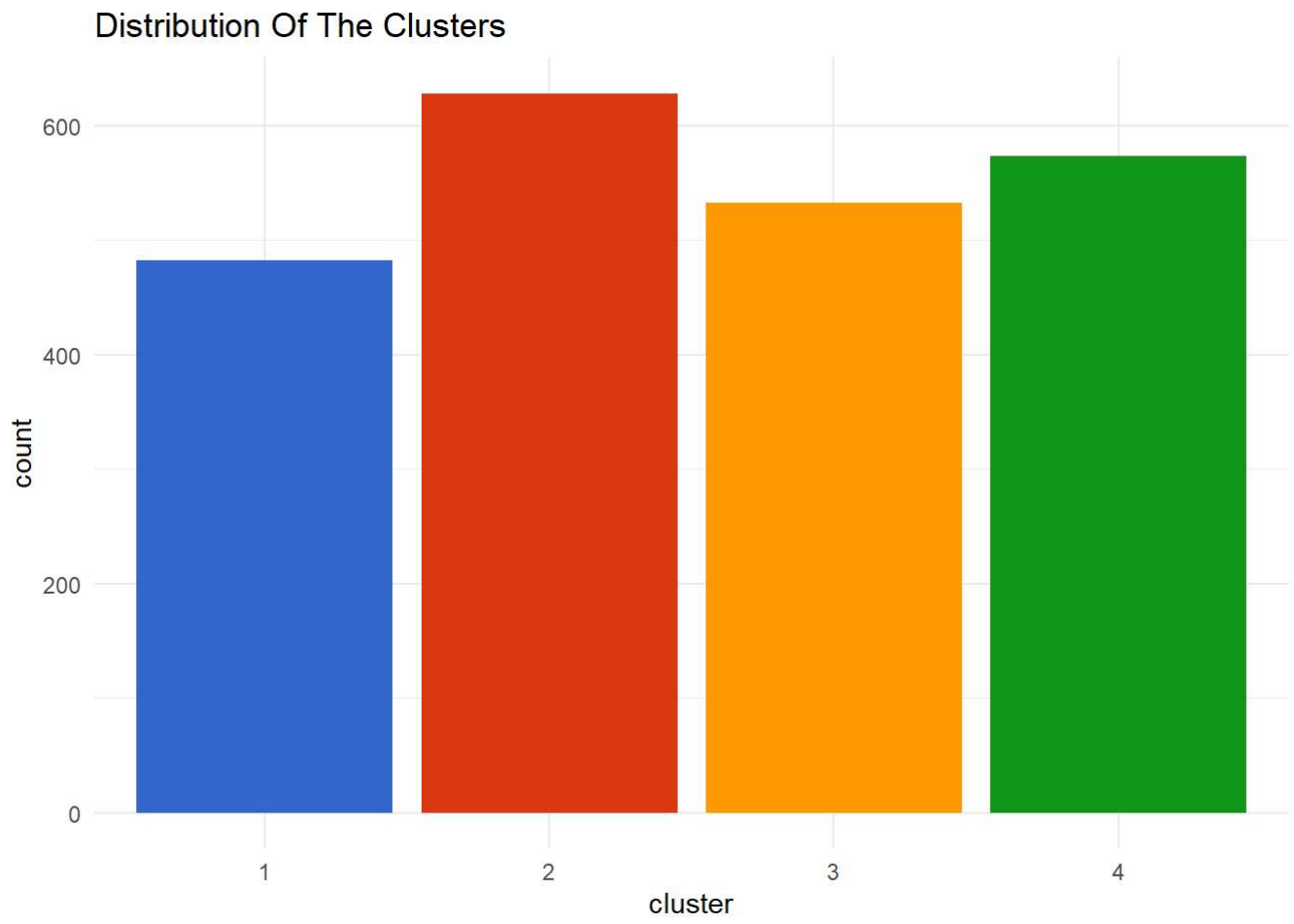
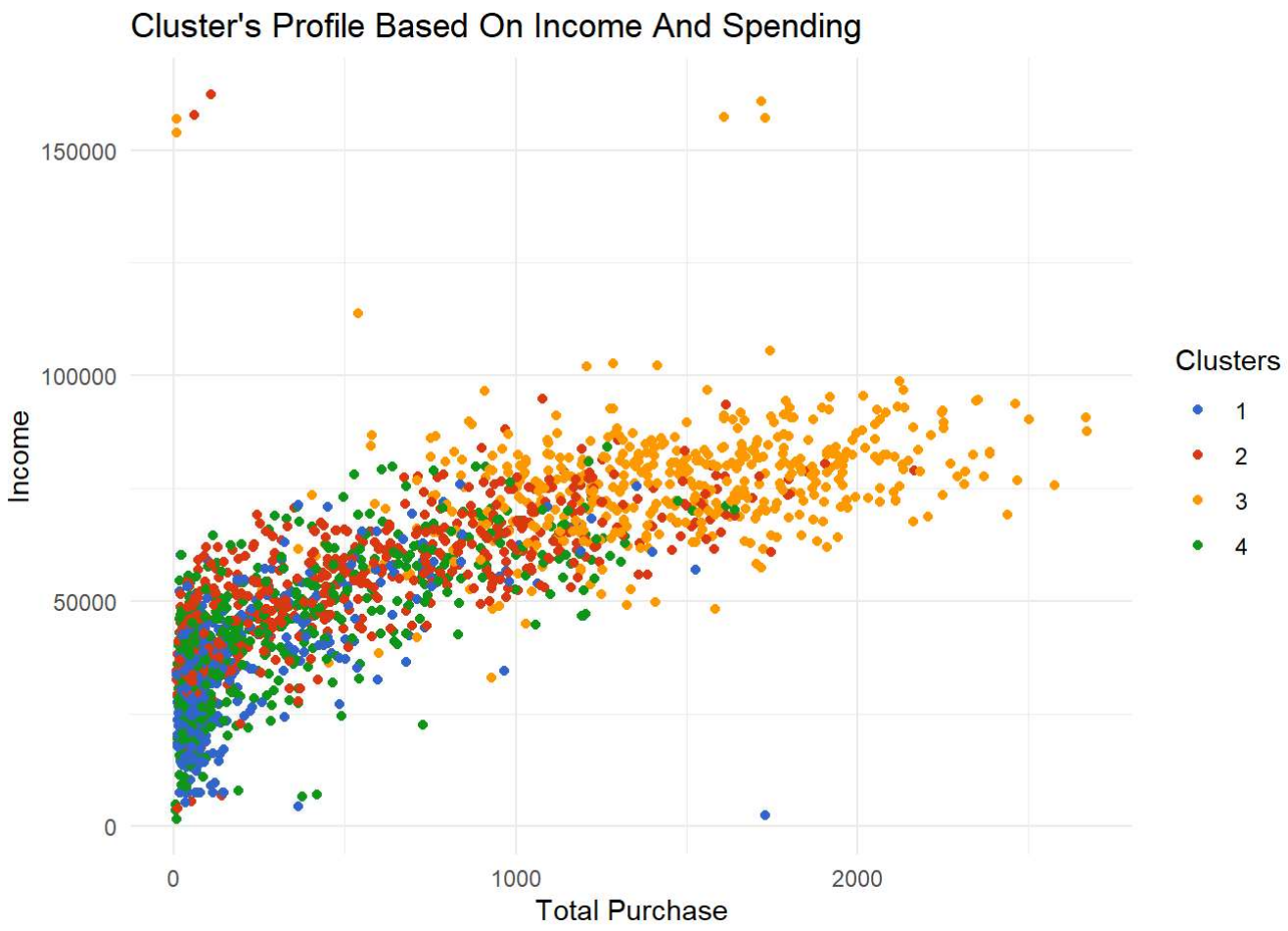


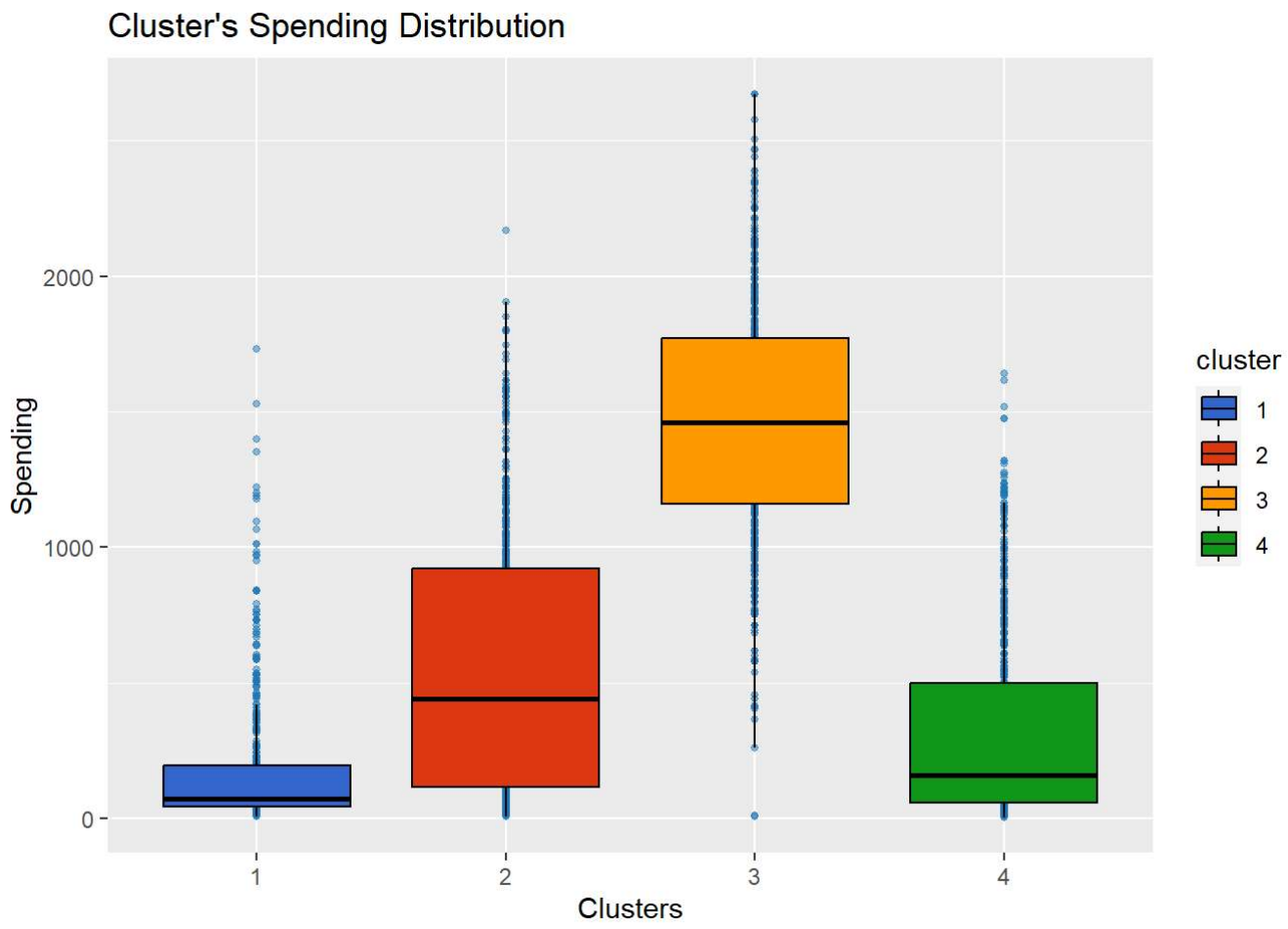
Evenly distributed size of each cluster

```
ggplot(df, aes(x = Total_Purchaase, y = Income, color = cluster)) +
  geom_point() +
  scale_color_manual(values = c("#3366CC", "#DC3912", "#FF9900", "#109618")) +
  ggtitle("Cluster's Profile Based On Income And Spending") +
  xlab("Total Purchase") +
  ylab("Income")+
  guides(color = guide_legend(title = "Clusters"))+
  theme_minimal()
```



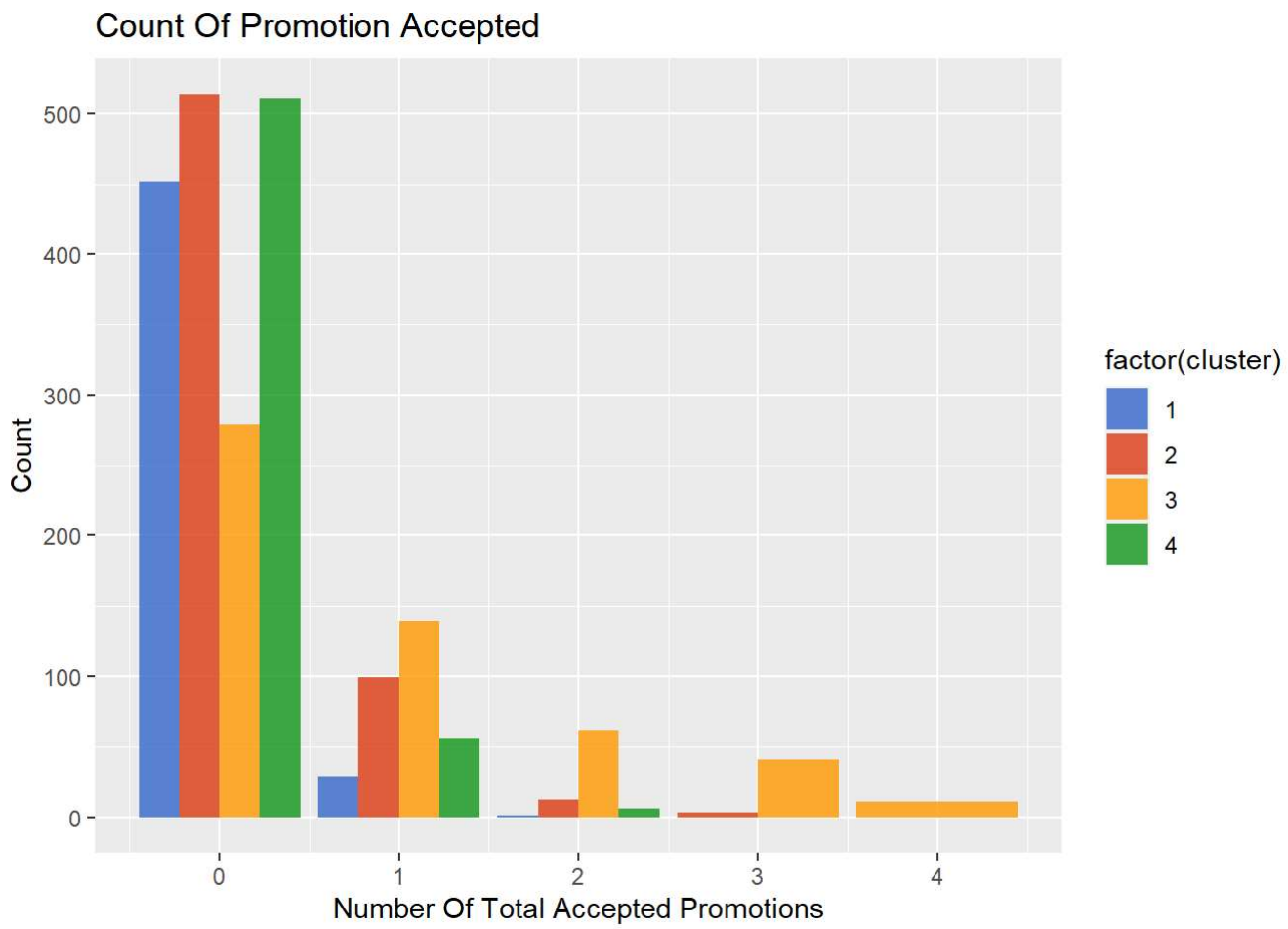
Green is high income while yellow is high income

```
ggplot(df, aes(x = cluster, y = Total_Purchaase)) +
  geom_point(size = 1,color = "#1f77b7", alpha = 0.5) +
  geom_boxplot(aes(fill = cluster), color = "black", outlier.shape = NA) +
  scale_fill_manual(values = c("#3366CC", "#DC3912", "#FF9900", "#109618")) +
  ggtitle("Cluster's Spending Distribution") +
  xlab("Clusters") +
  ylab("Spending")
```



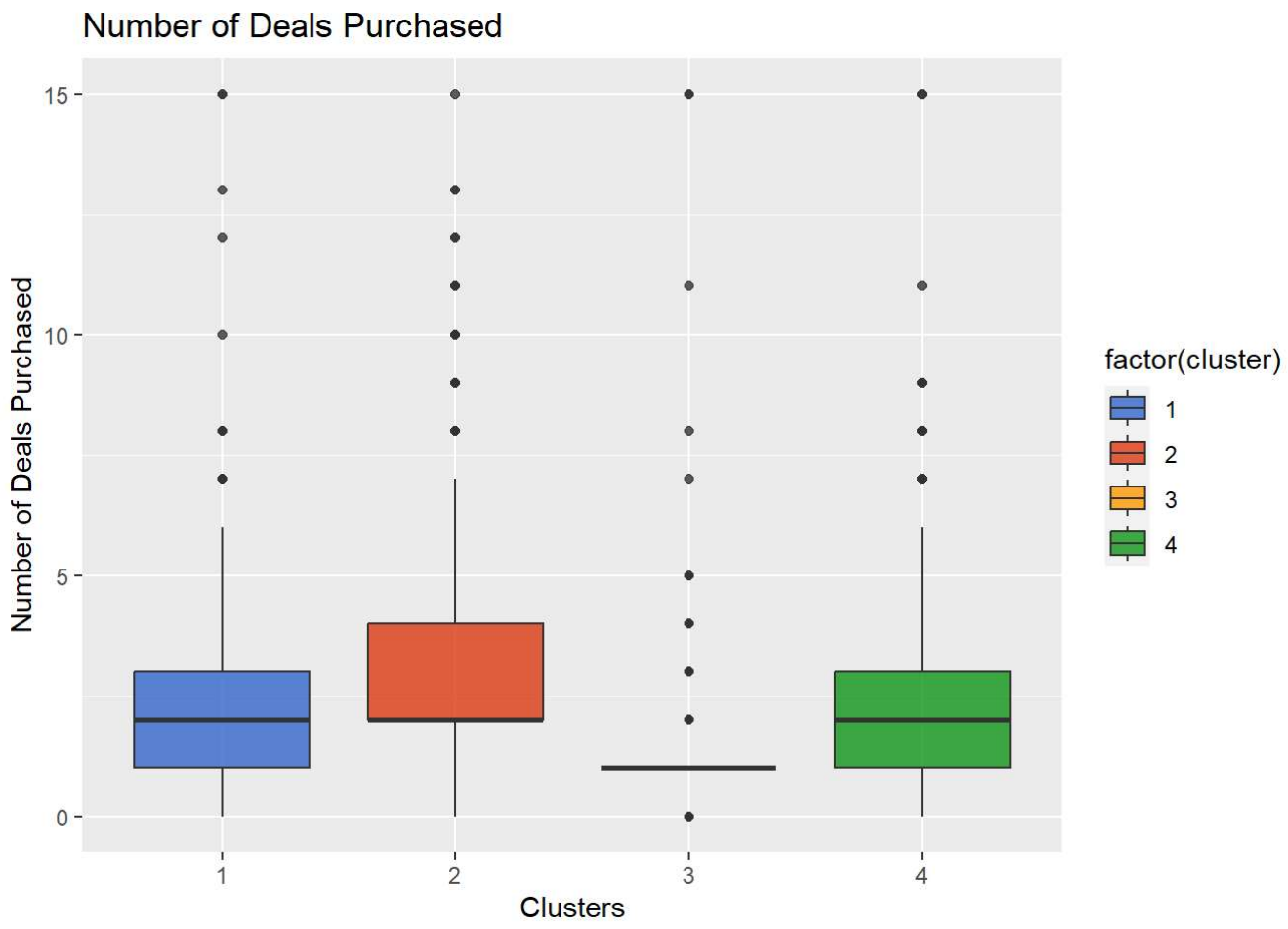
Green have high spending while yellow has low

```
df %>%
mutate(Total_Promos = AcceptedCmp1 + AcceptedCmp2 + AcceptedCmp3 + AcceptedCmp4 + AcceptedCmp5)
%>%
  ggplot(aes(x = Total_Promos, fill = factor(cluster))) +
  geom_bar(position ="dodge", alpha = 0.8) +
  scale_fill_manual(values = c("#3366CC", "#DC3912", "#FF9900", "#109618")) +
  ggtitle("Count Of Promotion Accepted") +
  xlab("Number Of Total Accepted Promotions") +
  ylab("Count")
```



The later campaign were most appealed to yellow cluster

```
ggplot(df, aes(x = factor(cluster), y = NumDealsPurchases, fill = factor(cluster))) +
  geom_boxplot(alpha = 0.8) +
  scale_fill_manual(values = c("#3366CC", "#DC3912", "#FF9900", "#109618")) +
  ggtitle("Number of Deals Purchased") +
  xlab("Clusters") +
  ylab("Number of Deals Purchased")
```

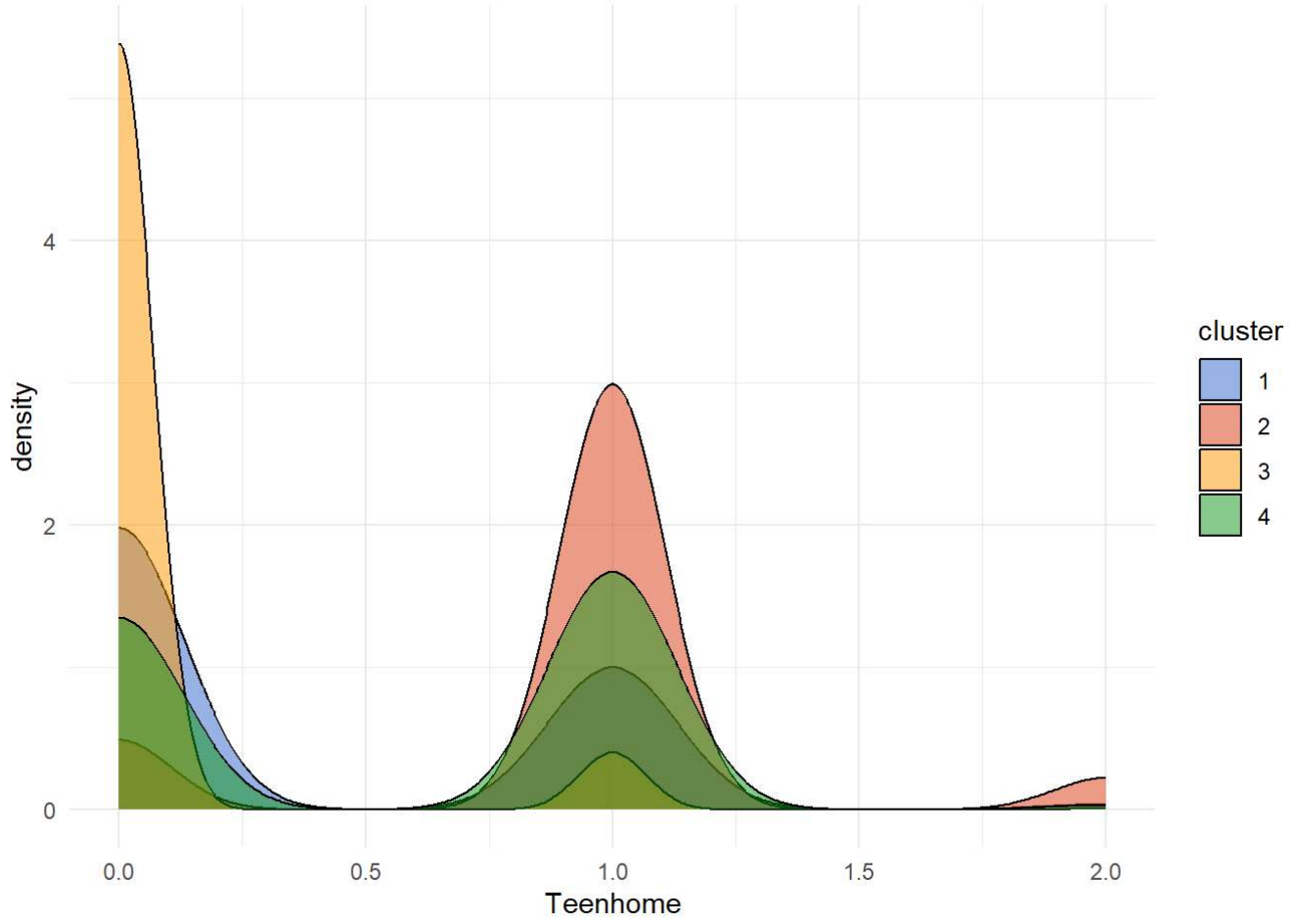


yellow did not get too many deals

```
df %>%
  group_by(cluster) %>%
  summarise(Wines = mean(MntWines), Fruits = mean(MntFruits), Meat = mean(MntMeatProducts), Fish
= mean(MntFishProducts), Sweet = mean(MntSweetProducts), gold = mean(MntGoldProds)) %>%
  ggplot(aes(x = cluster)) +
  geom_bar(aes(y = Wines, fill = "Wines"), stat = "identity", alpha = 0.7) +
  geom_bar(aes(y = Fruits, fill = "Fruits"), stat = "identity", alpha = 0.7) +
  geom_bar(aes(y = Meat, fill = "Meat"), stat = "identity", alpha = 0.7) +
  geom_bar(aes(y = Fish, fill = "Fish"), stat = "identity", alpha = 0.7) +
  geom_bar(aes(y = Sweet, fill = "Sweet"), stat = "identity", alpha = 0.7) +
  geom_bar(aes(y = gold, fill = "Gold"), stat = "identity", alpha = 0.7) +
  scale_fill_manual(values = c("Wines" = "#1F77B4", "Fruits" = "#FF7F0E", "Meat" = "#2CA02C", "F
ish" = "#D62728", "Sweet" = "#9467BD", "Gold" = "#8C564B")) +
  labs(title = "Average Spending on Product Categories by Cluster",
       x = "Cluster",
       y = "Average Spending",
       fill = "Type of Product")+
  theme_minimal() +
  theme(legend.position = "right")
```

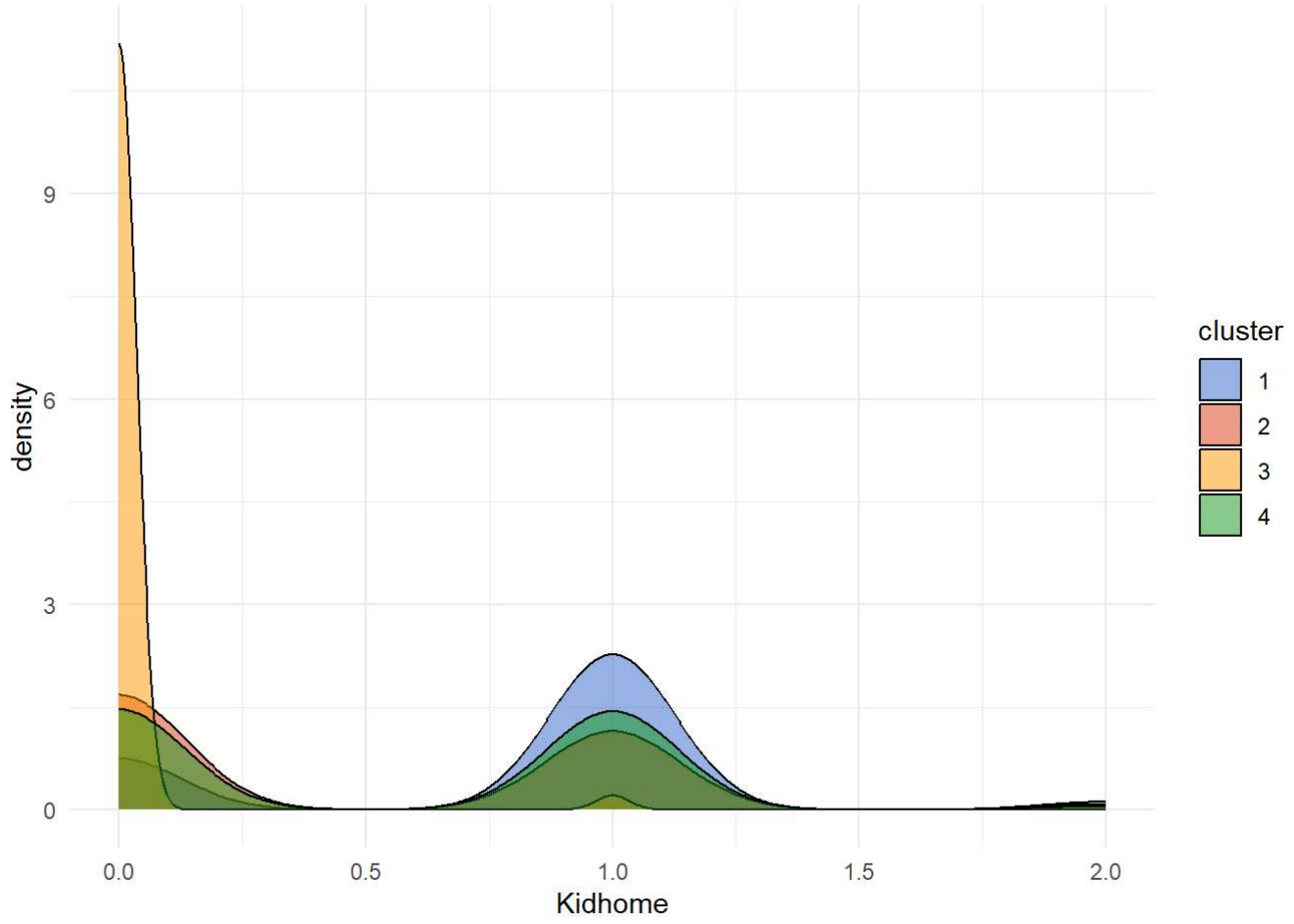
green consumes more meat

```
ggplot(df, aes(Teenhome, fill = cluster)) +
  geom_density(alpha = 0.5) +
  scale_fill_manual(values = c("#3366CC", "#DC3912", "#FF9900", "#109618")) +
  theme_minimal()
```



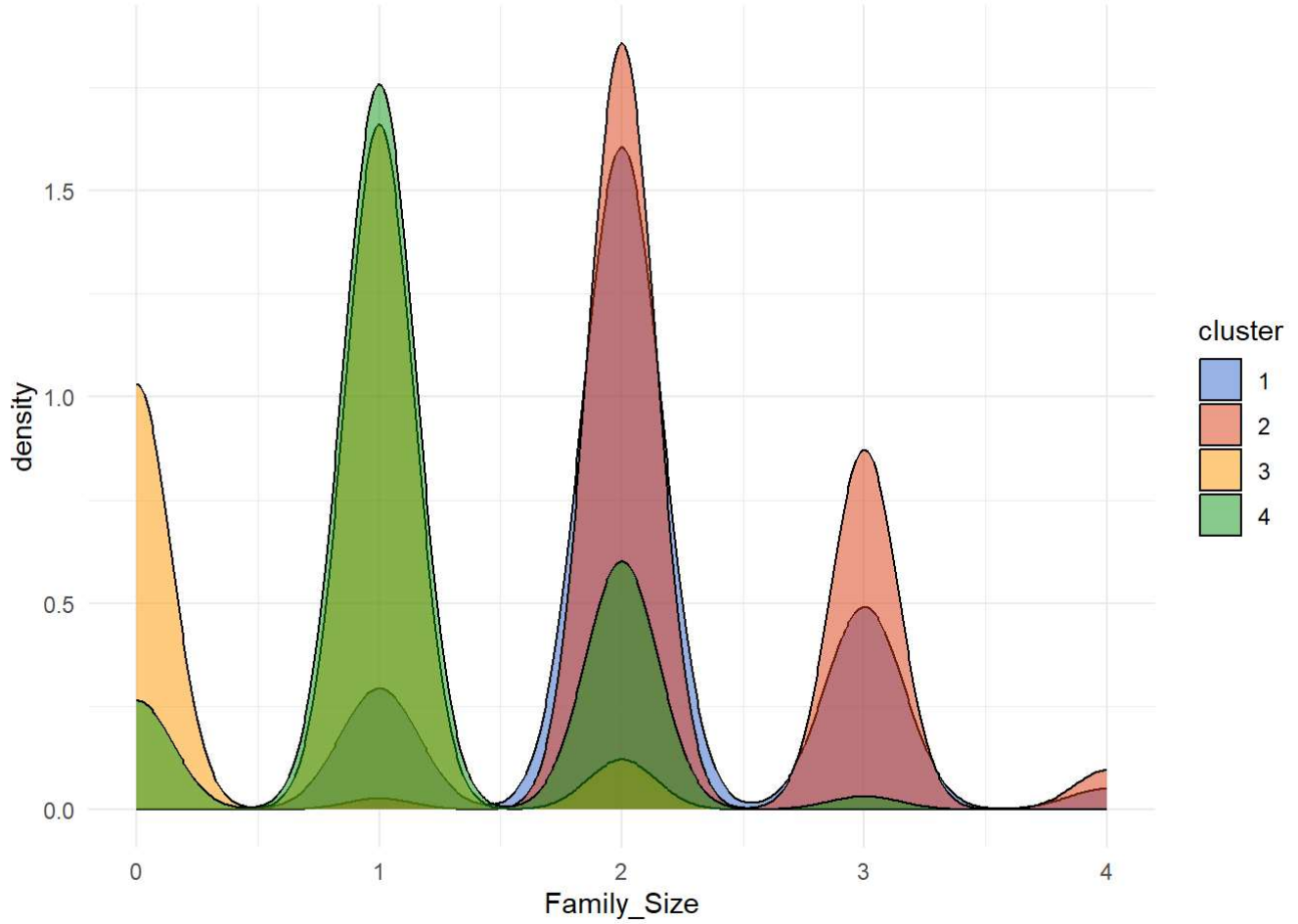
Red is a parent for sure while green isn't

```
ggplot(df, aes(Kidhome, fill = cluster)) +
  geom_density(alpha = 0.5) +
  scale_fill_manual(values = c("#3366CC", "#DC3912", "#FF9900", "#109618")) +
  theme_minimal()
```



Green doesn't have kid

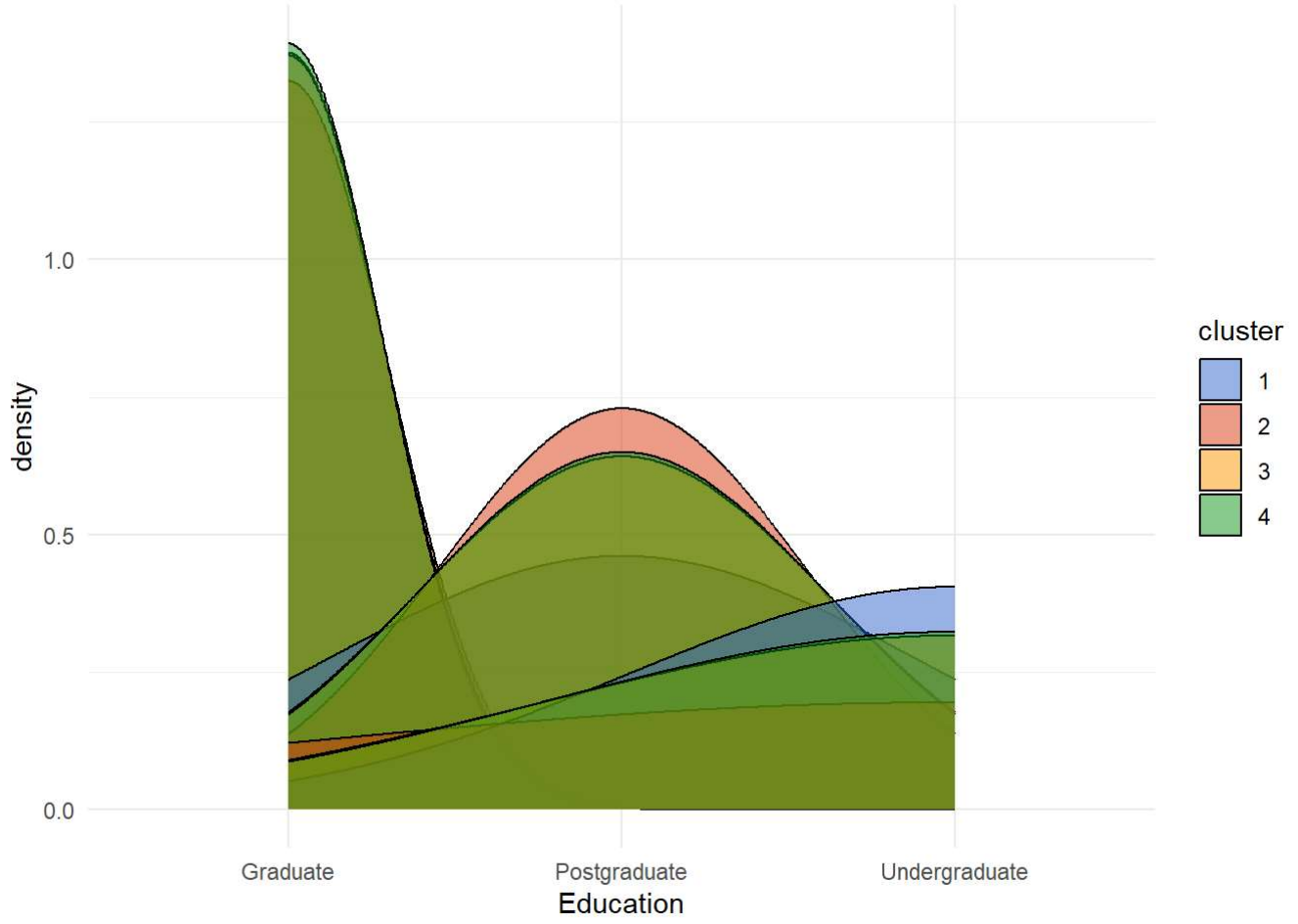
```
ggplot(df, aes(Family_Size, fill = cluster)) +
  geom_density(alpha = 0.5) +
  scale_fill_manual(values = c("#3366CC", "#DC3912", "#FF9900", "#109618")) +
  theme_minimal()
```



Red and Yellow are definitely have family while green and blue seem they don't

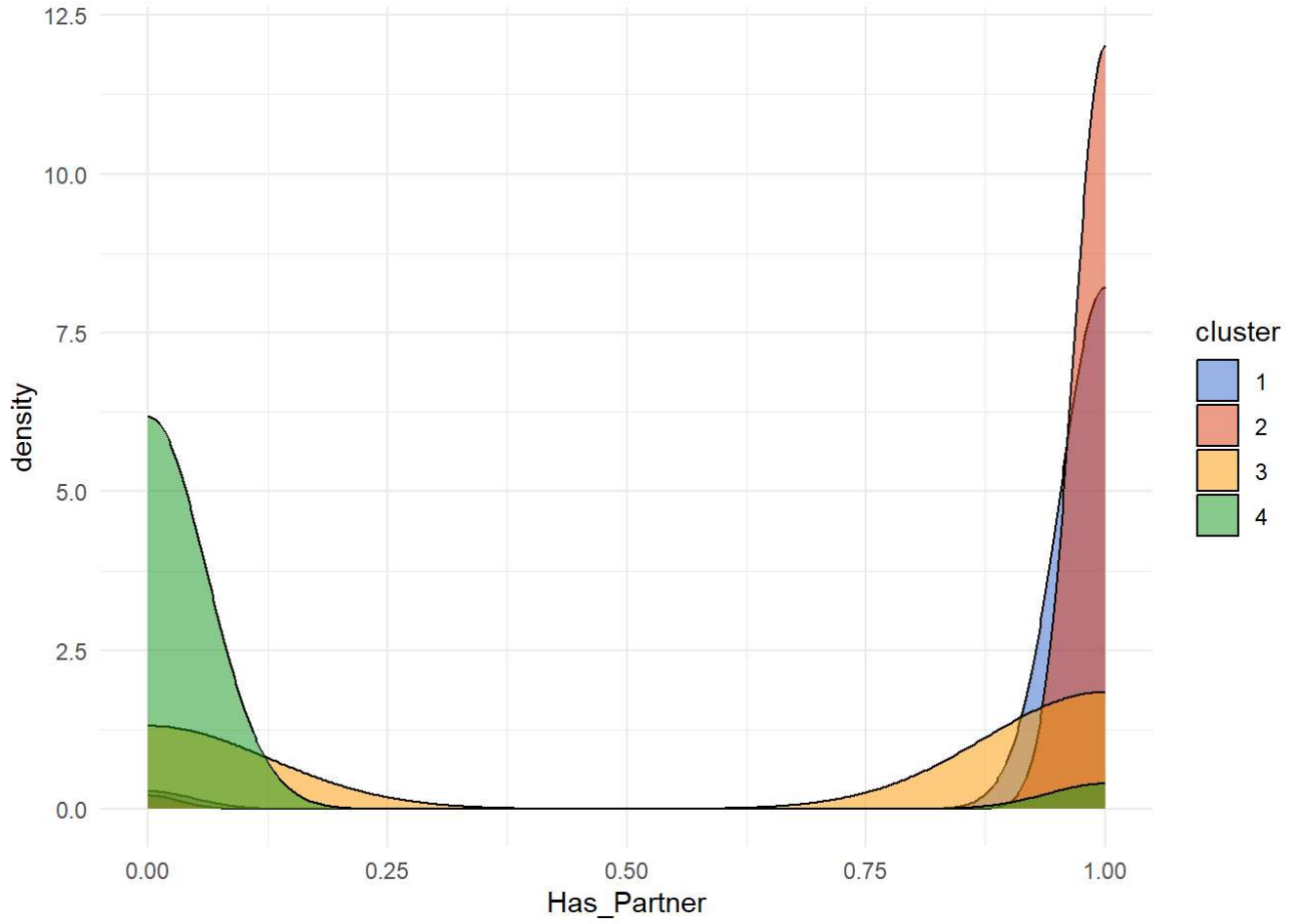
```
ggplot(df, aes(Age, fill = cluster)) +
  geom_density(alpha = 0.5) +
  scale_fill_manual(values = c("#3366CC", "#DC3912", "#FF9900", "#109618")) +
  theme_minimal()
```


```
ggplot(df, aes(Education, fill = cluster)) +
  geom_density(alpha = 0.5) +
  scale_fill_manual(values = c("#3366CC", "#DC3912", "#FF9900", "#109618")) +
  theme_minimal()
```



No Particular specific conclusion can be drawn

```
ggplot(df, aes(Has_Partner, fill = cluster)) +
  geom_density(alpha = 0.5) +
  scale_fill_manual(values = c("#3366CC", "#DC3912", "#FF9900", "#109618")) +
  theme_minimal()
```



Red and Yellow have a partner while blue mostly doesn't and nothing conclusive could be said about green.

```
write.csv(df,"Data/CustomerWithClusters.csv")
```

Conclusion

Red Cluster:

Income: Average

Teens at Home: 1-2 teens

Family Size: More than 2 members

Age: 40-60 years old

Parent: 1 partner

Blue Cluster:

Income: Low

Teens at Home: Somewhere in between (unclear range)

Family Size: 1-2 members

Age: 30-40 years old

Parent: 1 parent

Yellow Cluster:

Income: High

Teens at Home: No teens at home

Family Size: No family

Age: Not specified

Parent: Somewhere in between (unclear criteria)

###Green Cluster:

Income: Average

Teens at Home: Somewhere in between (unclear range)

Family Size: Single person household

Age: Not specified

Partner: Single (Could be Single Parents)