PROJECT #115

CUSTOMER PERSONALITY ANALYSIS

EXCELR
Raising Excellence

MENTOR : Rajshekhar

TEAM MEMBERS

1.Rohit Shirudkar
2.Akshay Subramanian
3.Alka Aswar
4.Junaid Hussain
5.Mohit Shrimali

BUSINESS PROBLEM

EXCELR
Raising Excellence

BUSINESS OBJECTIVE

Customer personality analysis helps a business to modify its product based on its target customers from different types of customer segments.
Finding the potential customers by analysing the behaviour of them is useful to understand the targeted customers.
For example, instead of spending money to market a new product to every customer in the company's database, a company can analyse which customer segment is most likely to buy the product and then market the product only on that particular segment.

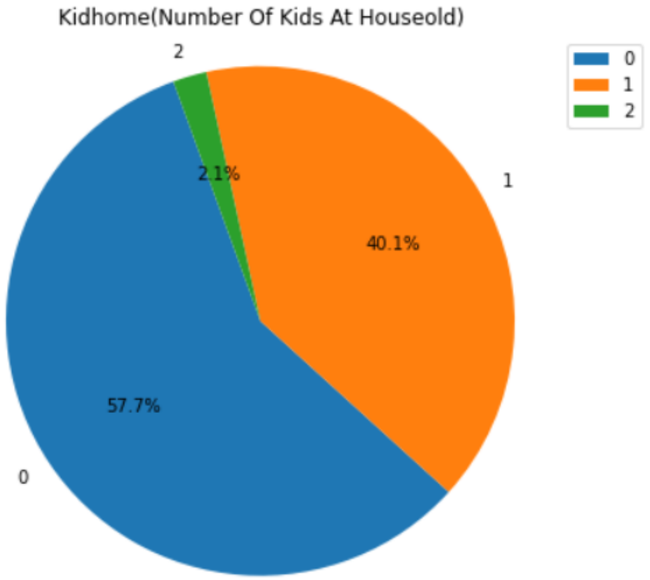
ARCHITECTURE LEVEL ANALYSIS

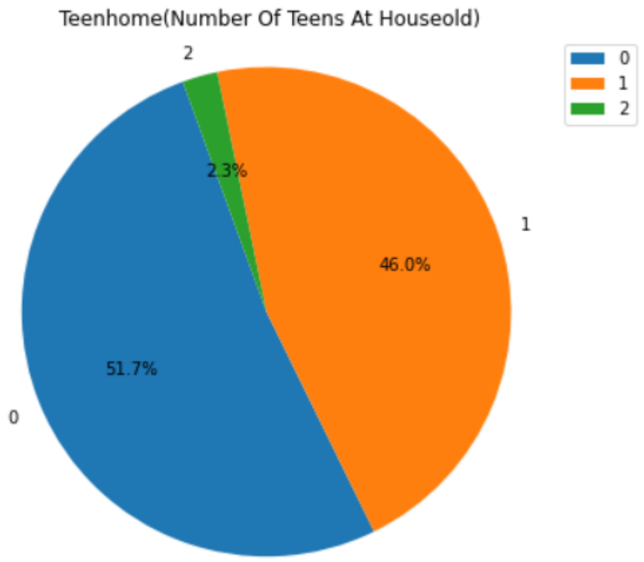
Customer personality analysis helps a business to modify its product based on its target customers from different types of customer segments.
Finding the potential customers by analysing the behaviour of them is useful to understand the targeted customers.
For example, instead of spending money to market a new product to every customer in the company's database, a company can analyse which customer segment is most likely to buy the product and then market the product only on that particular segment.

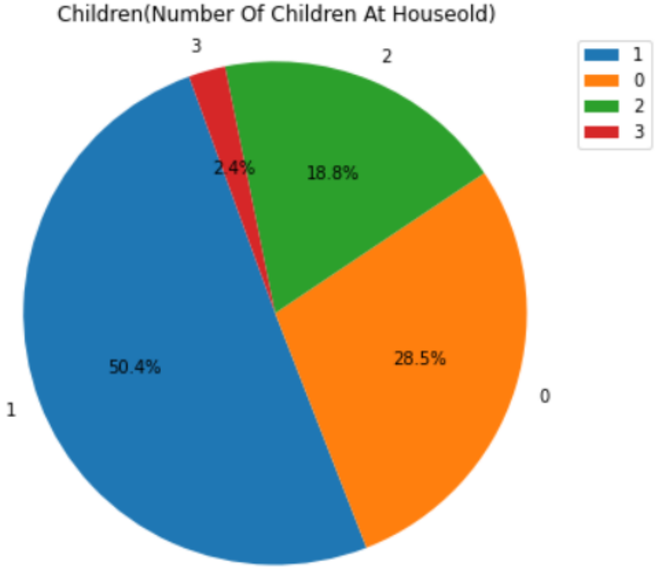
EXPLORATORY DATA ANALYSIS

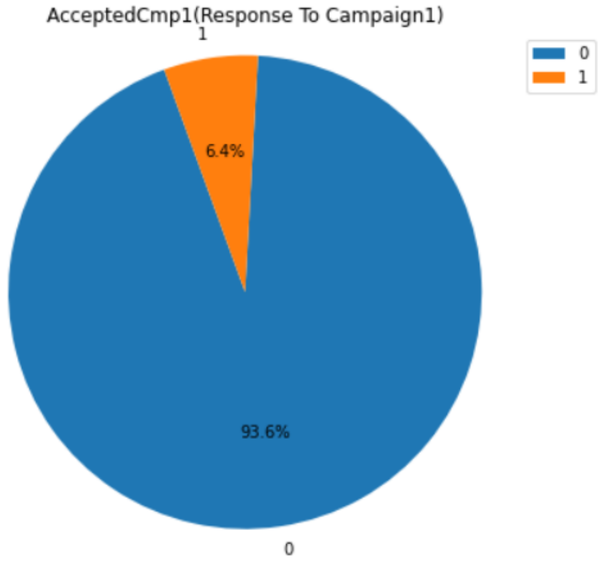
- 58% of customers have no kids in household
- 40% of customers have 1 kid and only 2% of customers have 2 kids in household
- 52% of customers have no Teens in household
- 46% of customers have 1 Teen and 2% cutomers have 2 Teens in household

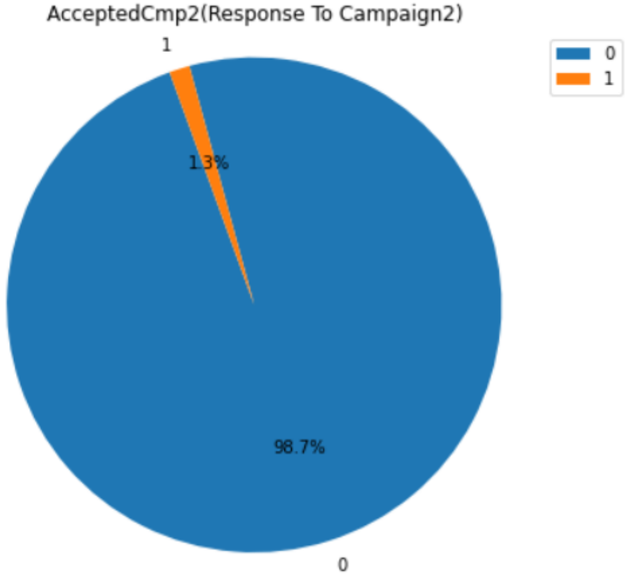
50% CUSTOMERS HAVE AT LEAST ONE KID OR TEEN IN HOUSEHOLD

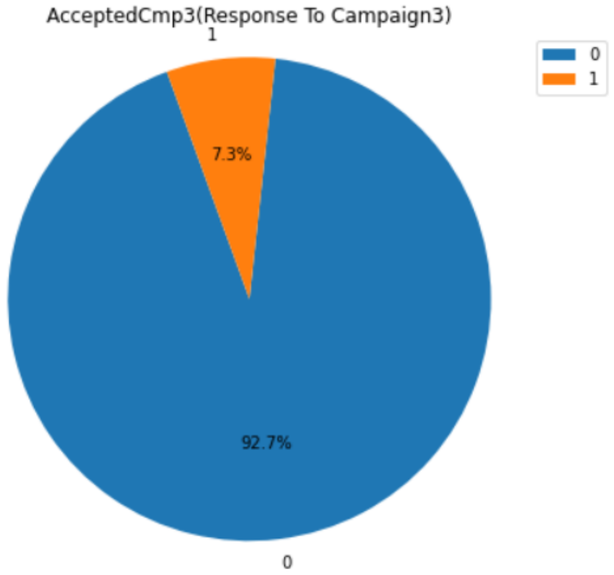


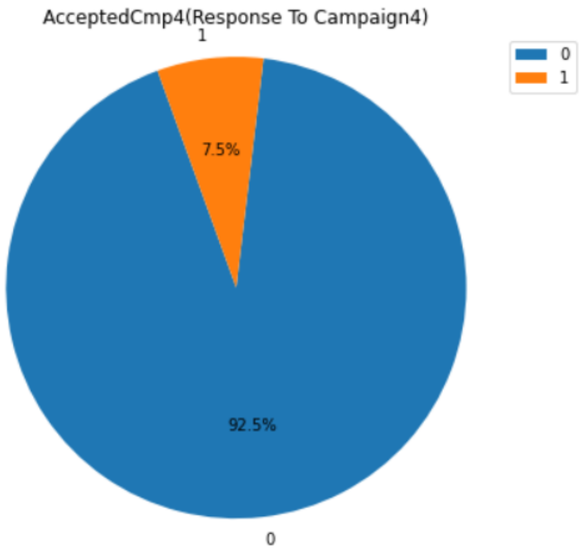


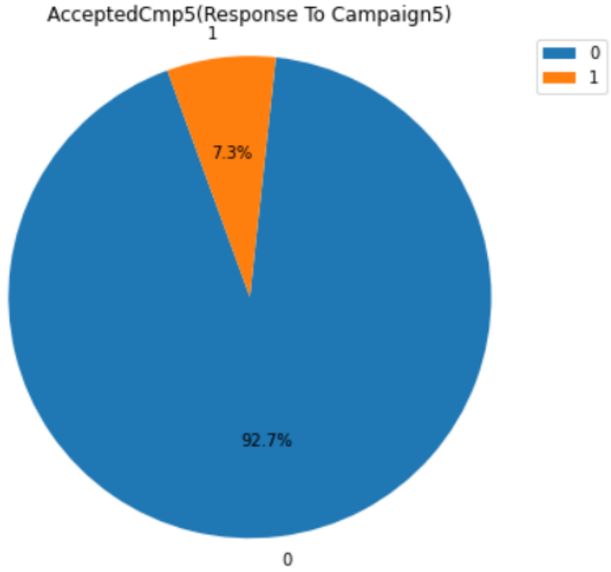




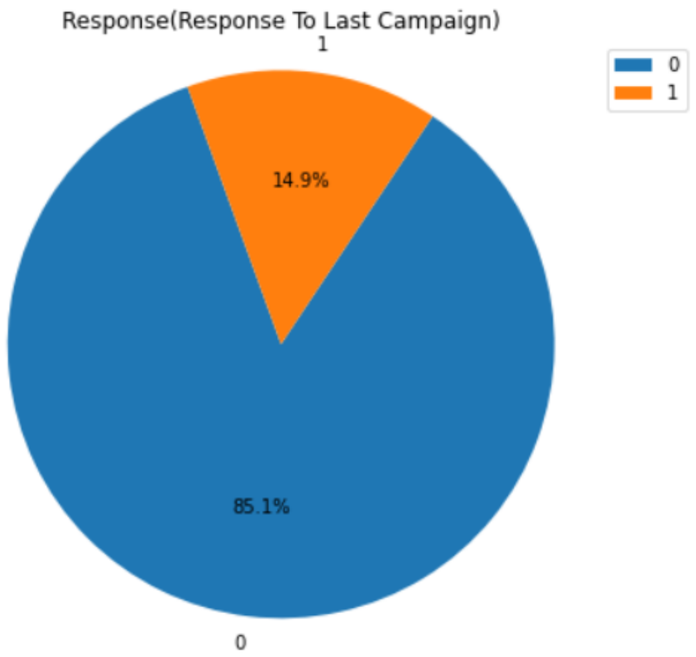


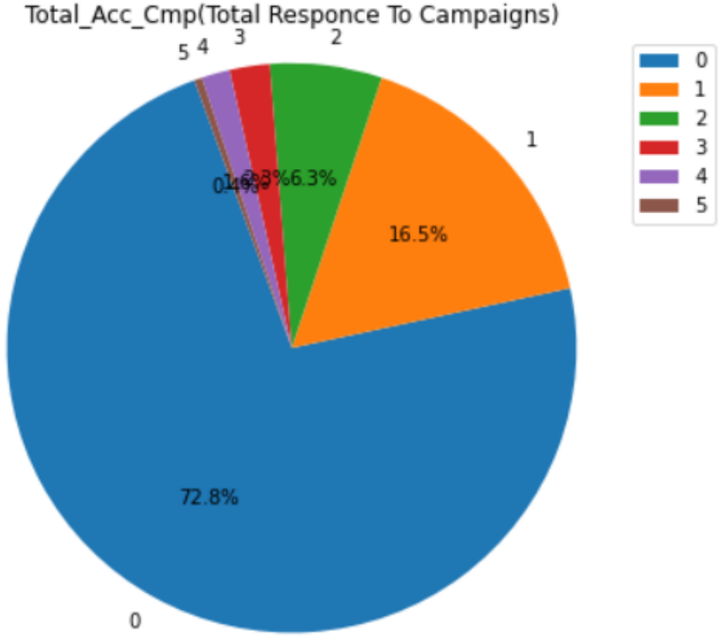






Accepted Campaigns

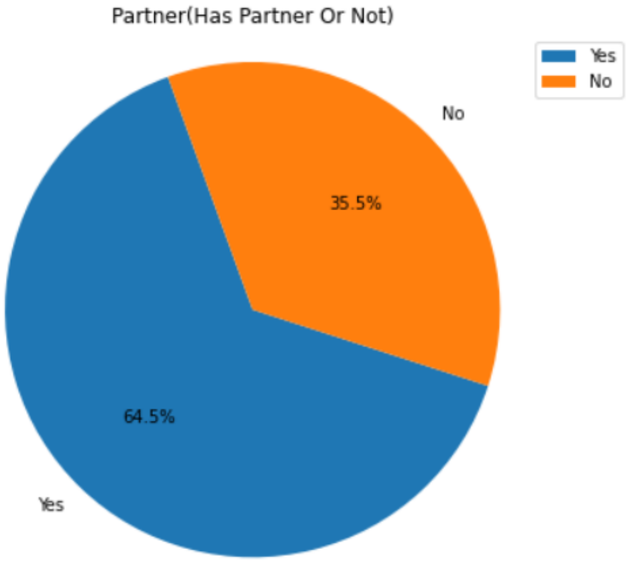




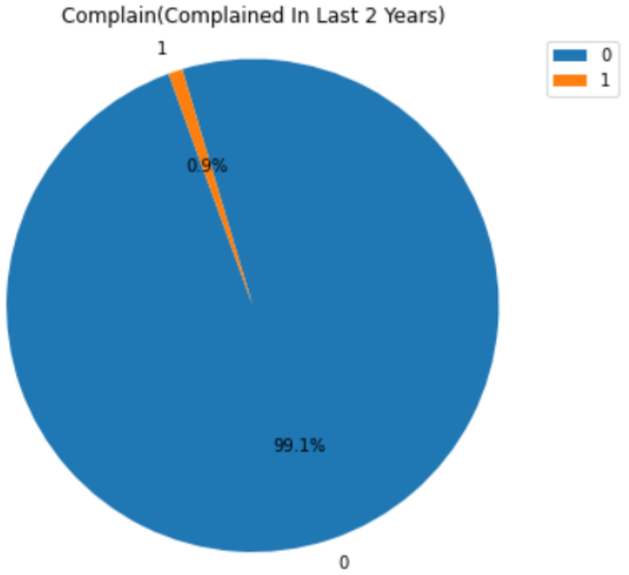
- Only 6% of customers accepted the offer in campaign 1
- Only 1.3% of customers accepted the offer in campaign 2
- Only 7.3% of customers accepted the offer in campaign 3
- Only 7.5% of customers accepted the offer in campaign 4
- Only 7.3% of customers accepted the offer in campaign 5

15% OF CUSTOMERS ACCEPTED THE OFFER IN LAST CAMPAIGN

73% OF CUSTOMERS DIDNT ACCEPTED ANY OFFER IN ANY CAMPAIGN

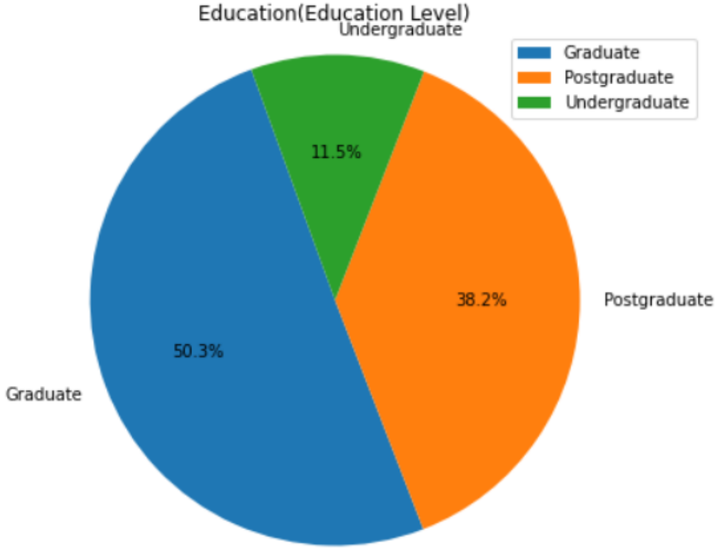


65.5% of customers have partner in life, which makes sense since majority of customers are well educated and might be ready to settle



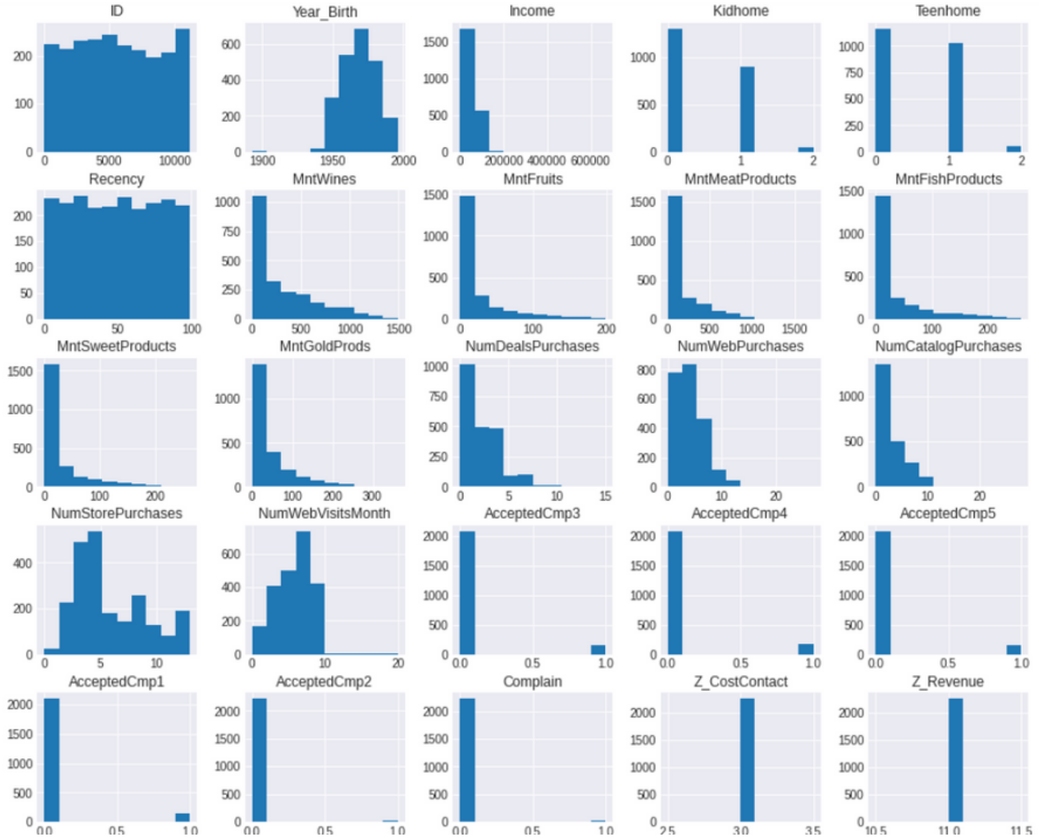
50% of customers are graduates and only 11% of customers are undergraduates, remaining are postgraduates.

So majority of customers are well educated.

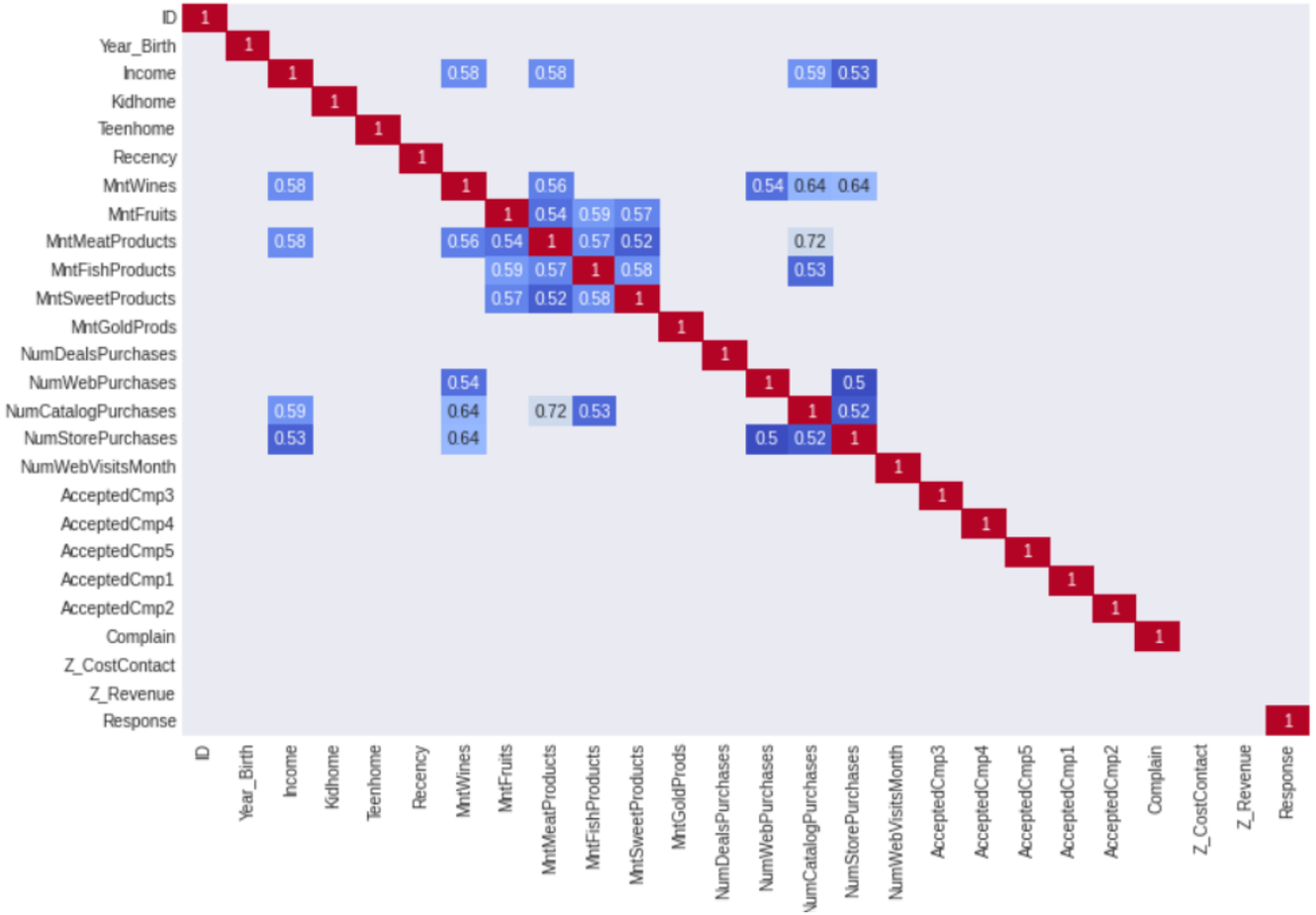


Only 1% of customers had any complaint in last 2 years.

Which is good thing or the customer service is really bad or not working properly or customers don't know how to complain

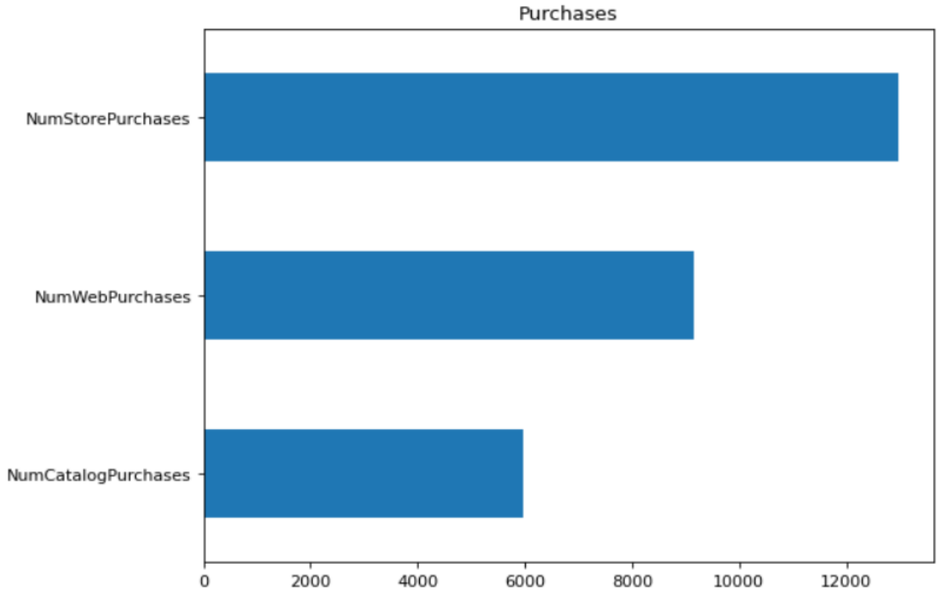


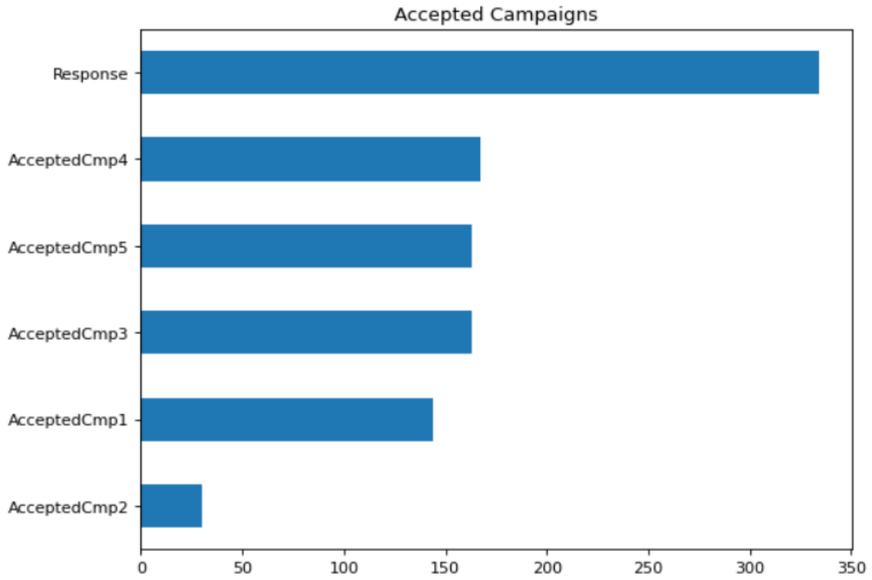
From the distribution of data we can see that Majority of data is positively skewed

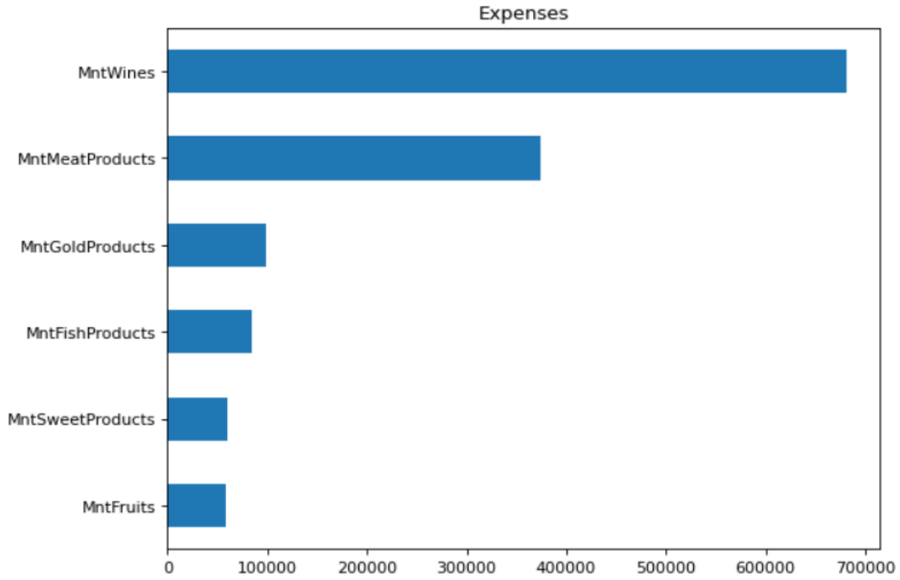


COLUMNS HAVING CO-RELATION GREATER THAN 0.5

Majority of customers like to by meat and fish products through catalog



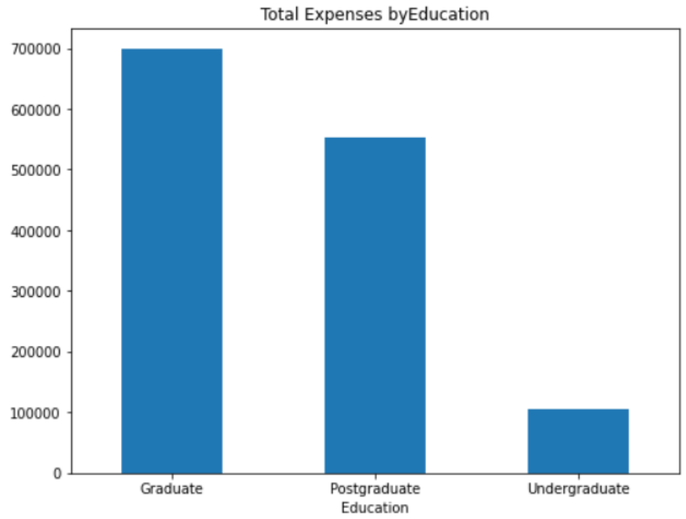


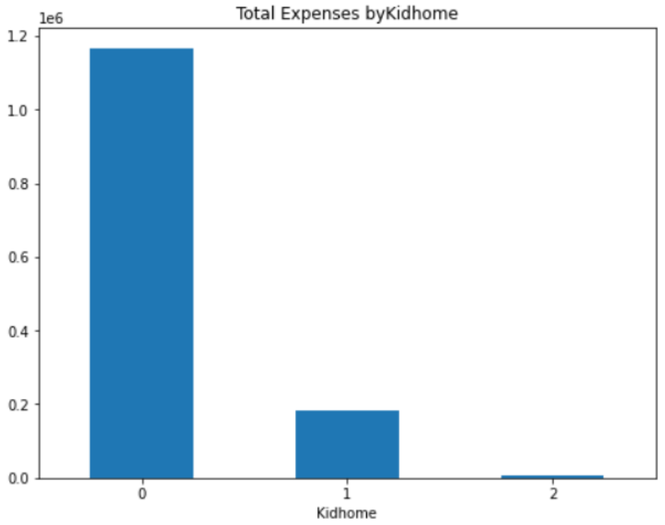


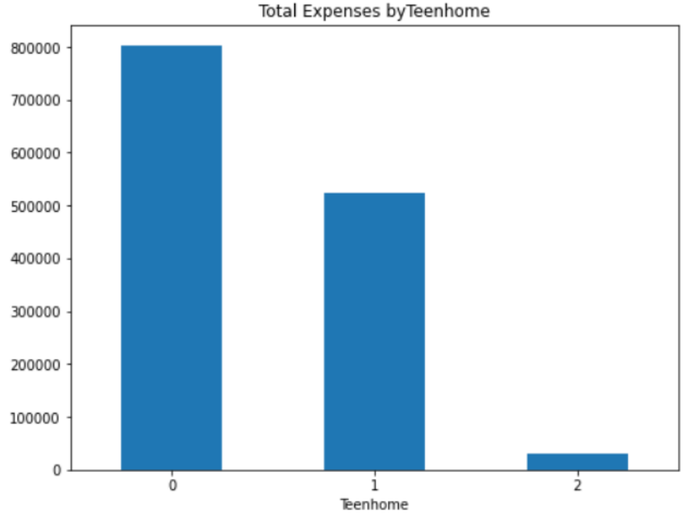
Most amount of money is spent on wine and meat by customers and minimum on fruits and sweets

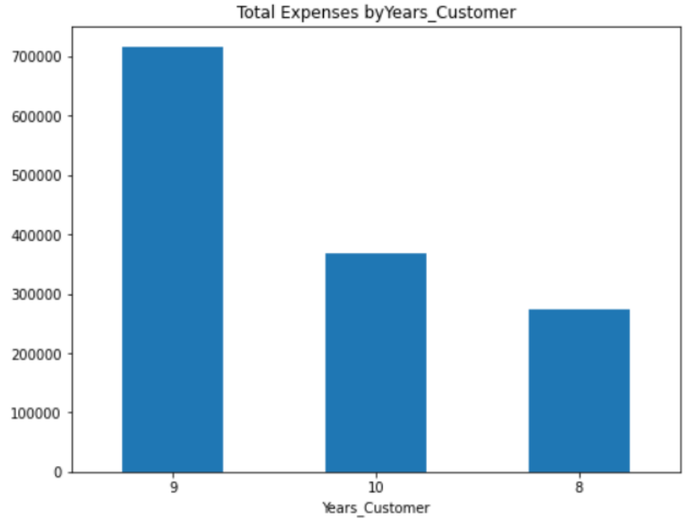
Majority of customers like to purchase through store

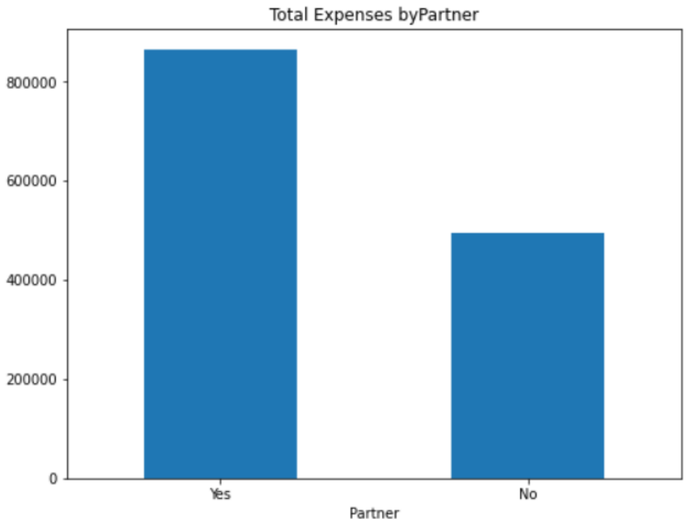
The chances of accepting an offer in campaign increases with the number of campaigns, maximum amount of customers accepted the last campaign



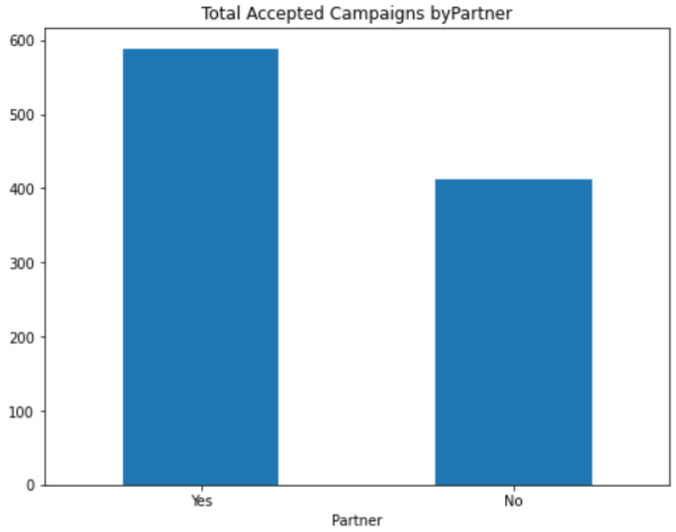


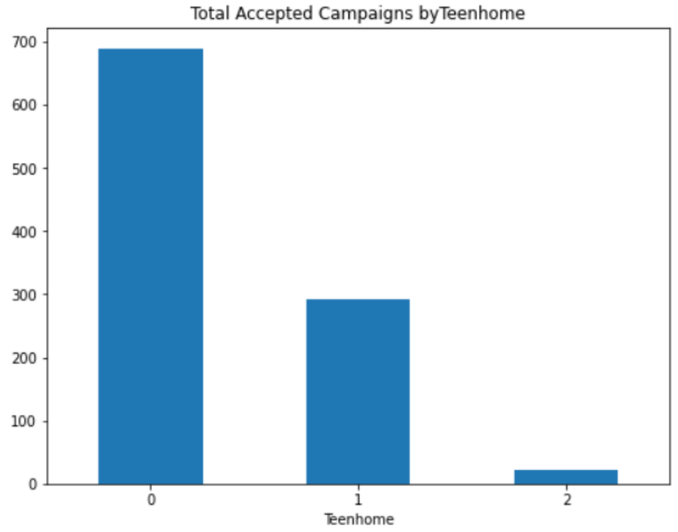


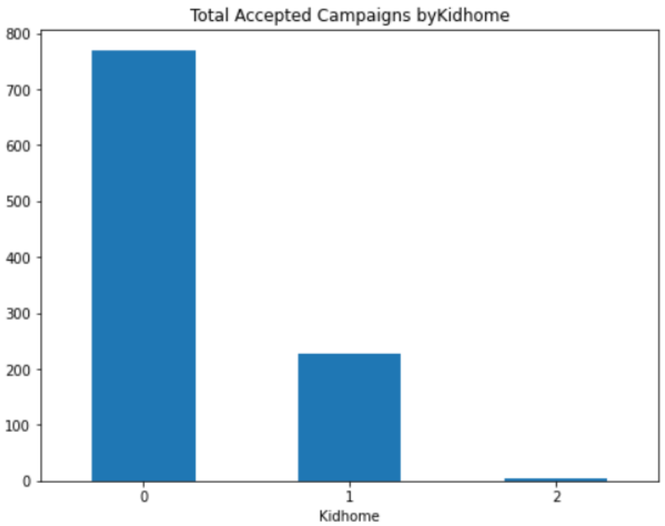


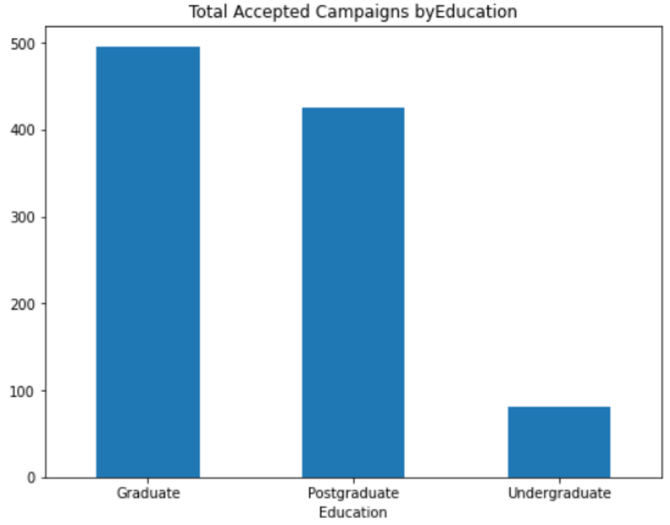


- Undergraduates have spent less money as compared to others, which is understandable
- Majority of customers with no kid and teen at home have spent more.
- Customers who have partner seems like to spent more money.

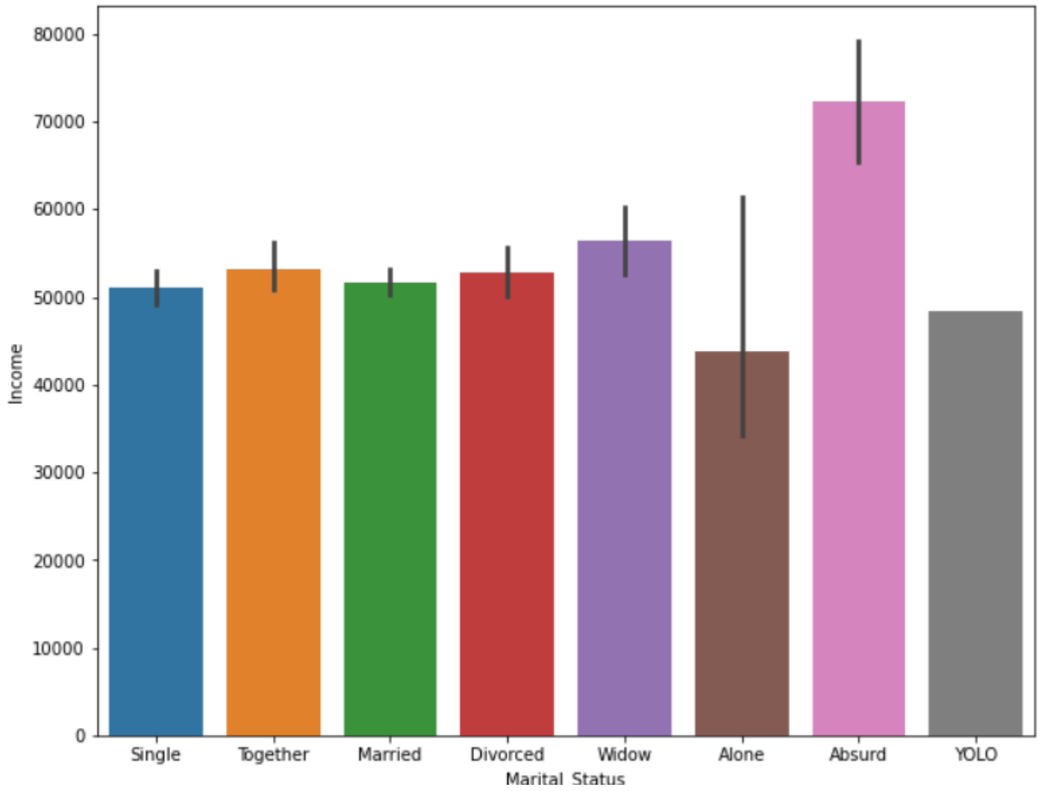


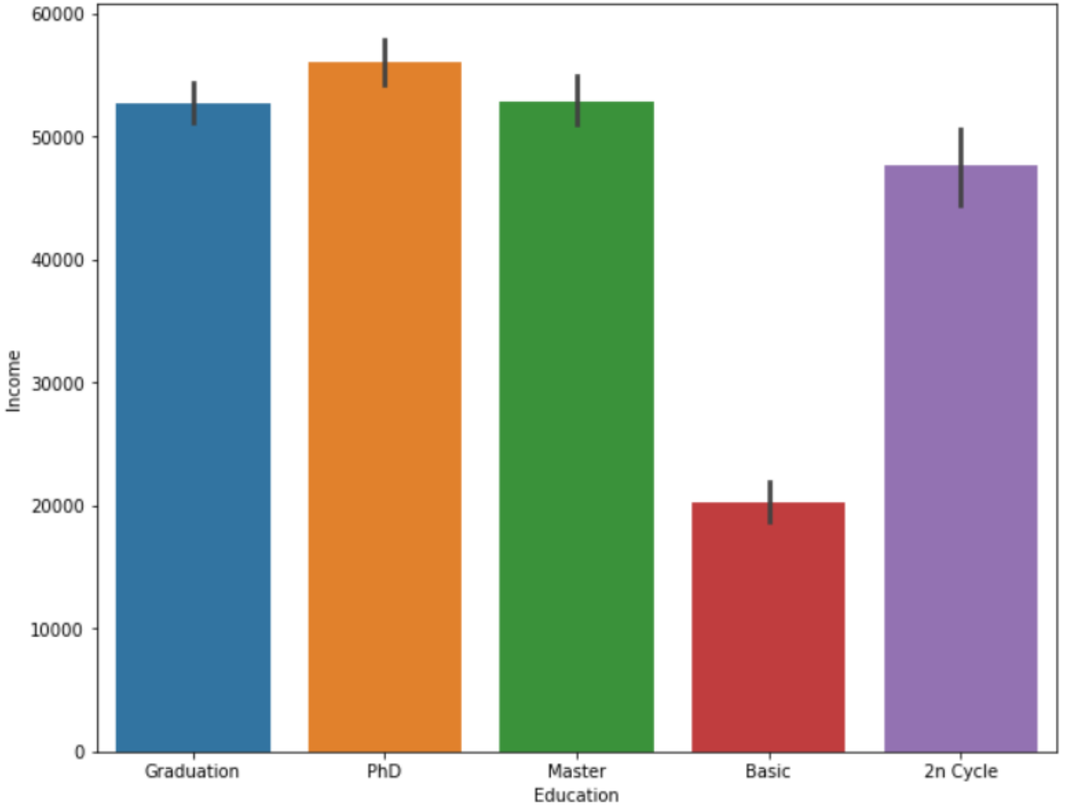






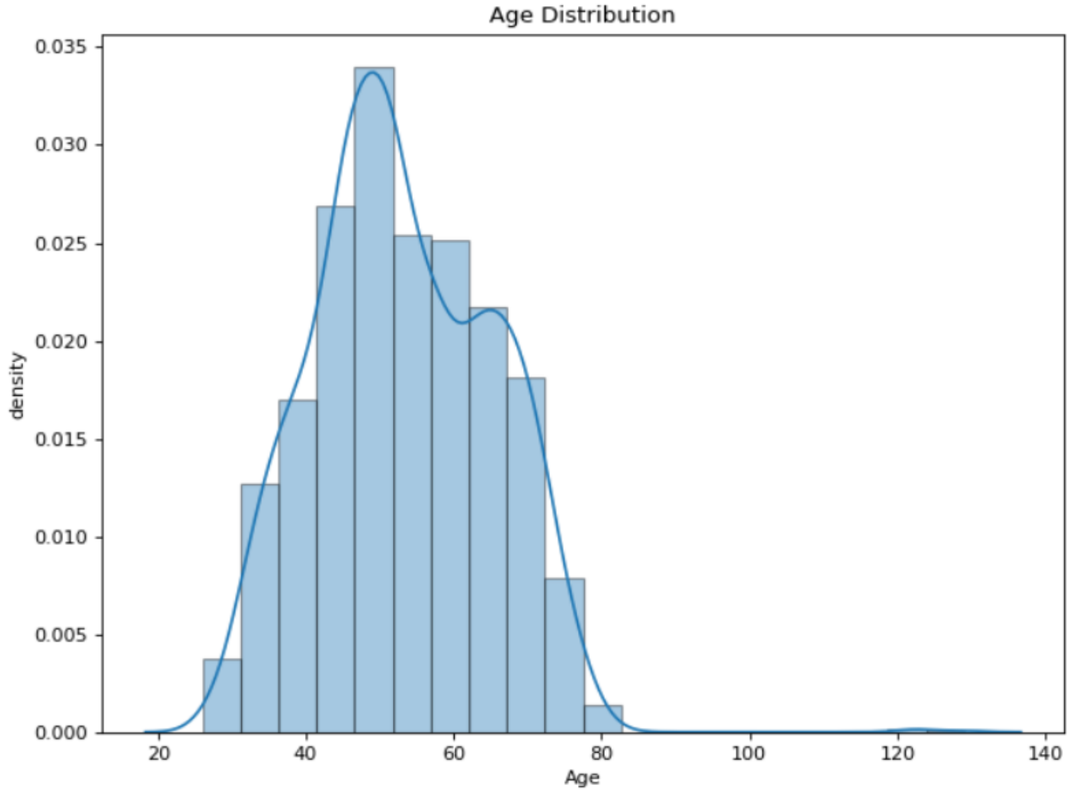
- Campaign acceptance rate is high with graduates and postgraduates
- majority of campaigns accepted by customers having partner
- Having kids at home does effect the campaign acceptance rate, maybe customer gets busy with kids and teen at home

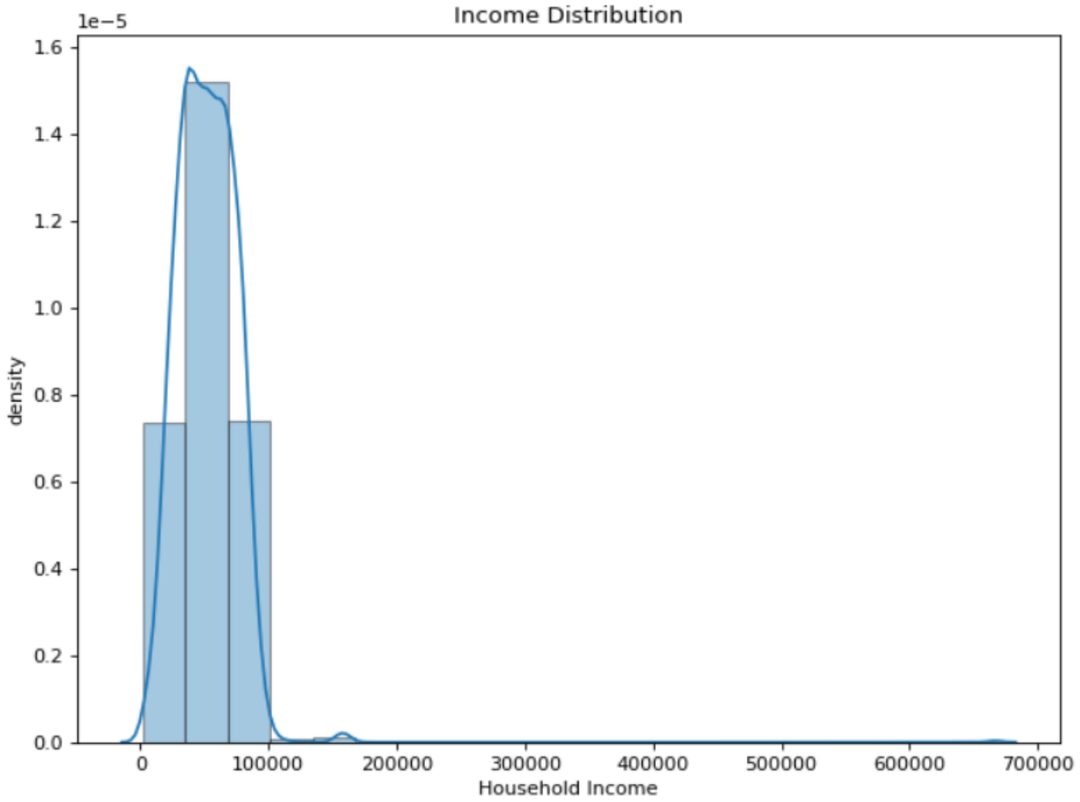




- It seems like income is almost similar for each category in maratial_status, except Alon, Absurd and YOLO
- YOLO seems to have highest income, however we have very little data for those three categories, so we can't be sure!
- So income has nothing to do with maratial_status.

- Income is low for customers with basic education, which is understandable, but customers with 2n cycle education have really good income
- Customers with PHD have highest income.



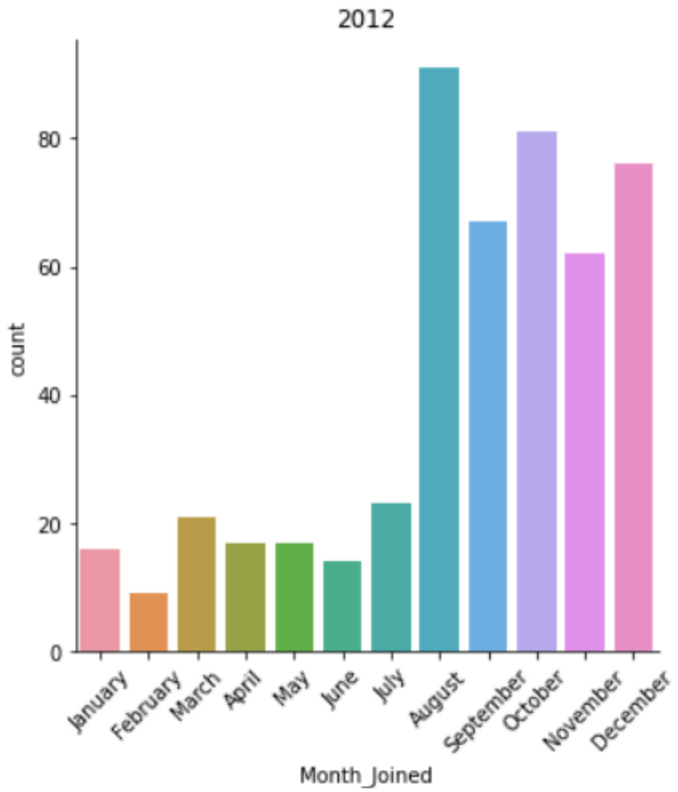


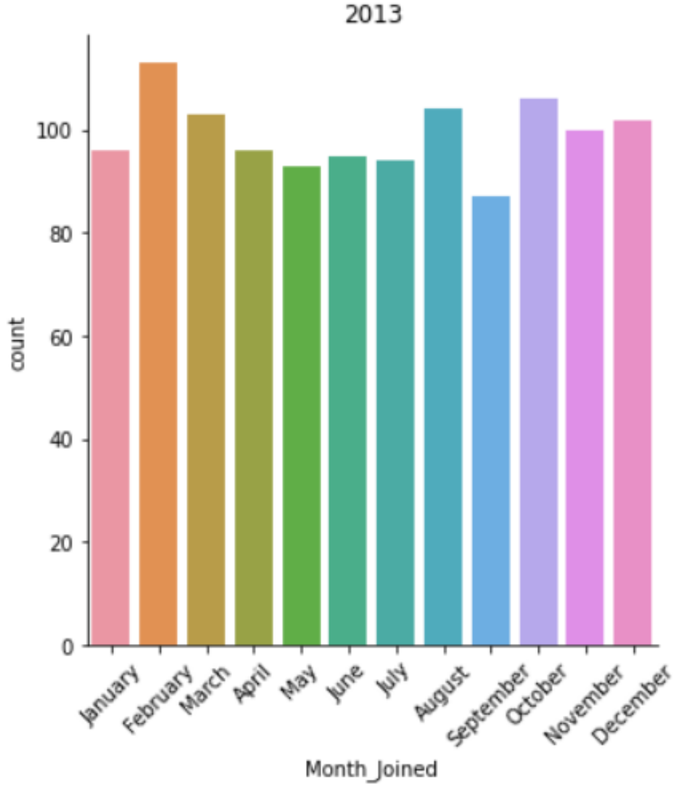
The majority of customers have age within 45 to 65
And there are few outliers on the right side

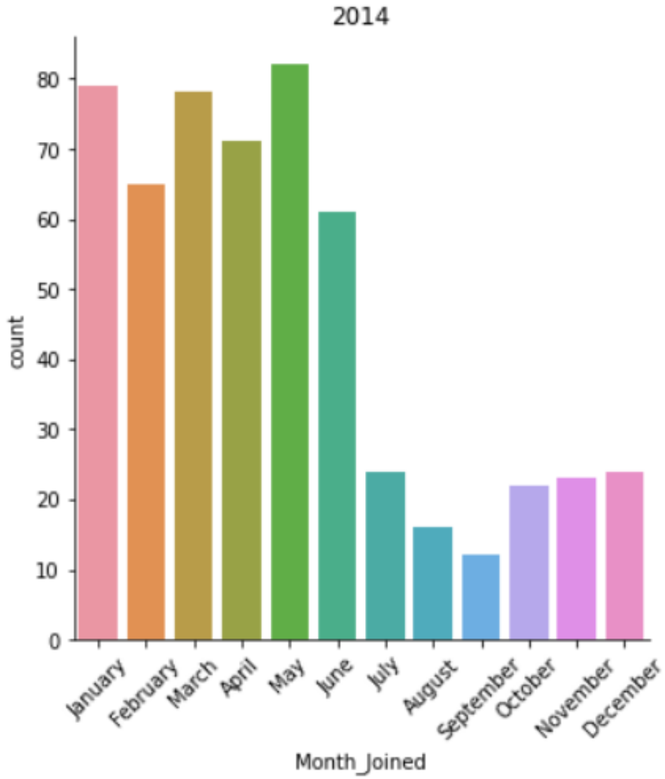
We can see from the distribution majority of customers have income within 30000–800000

We can see few outliers tho

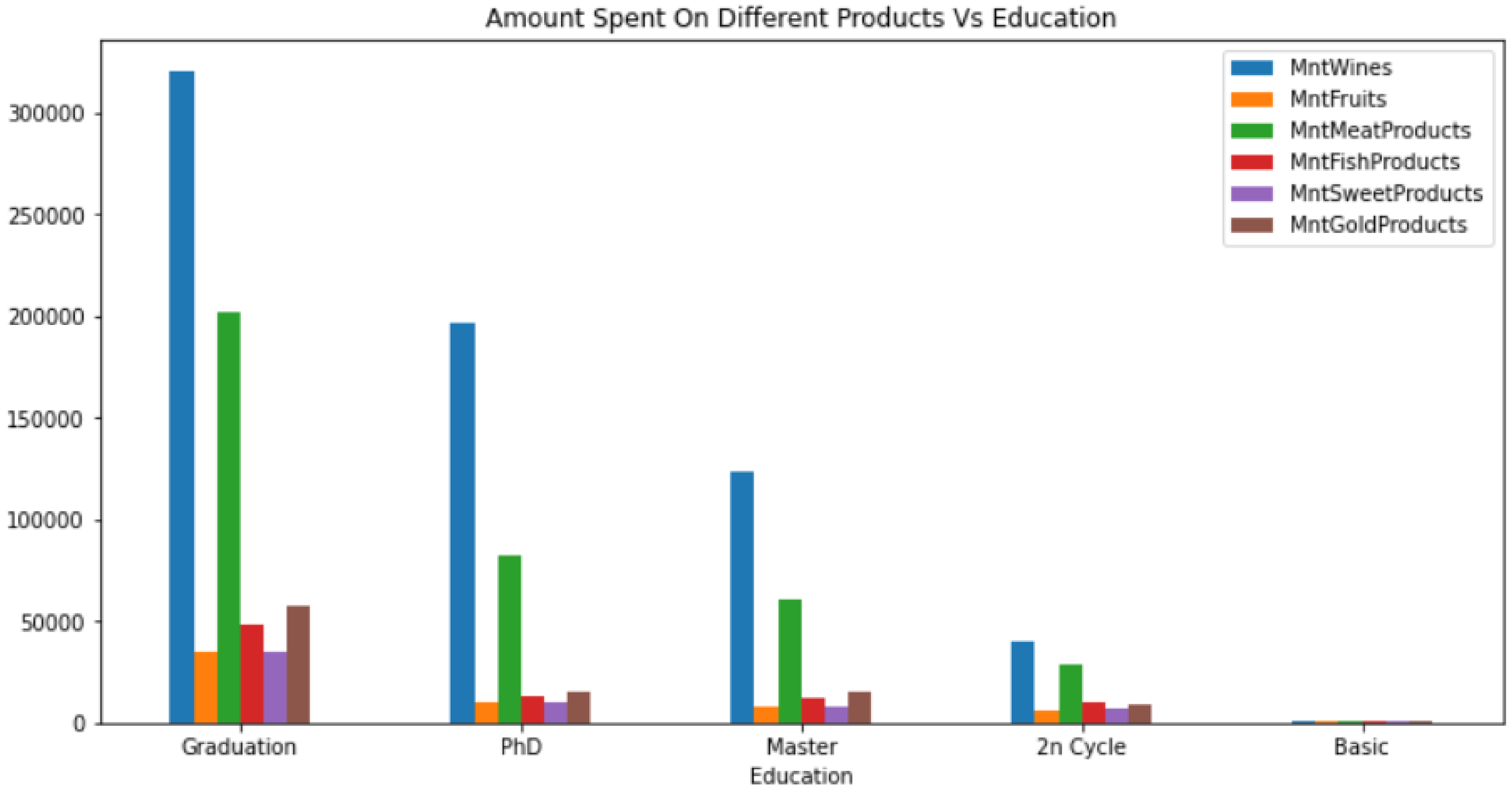
CUSTOMER ACQUISITION TREND



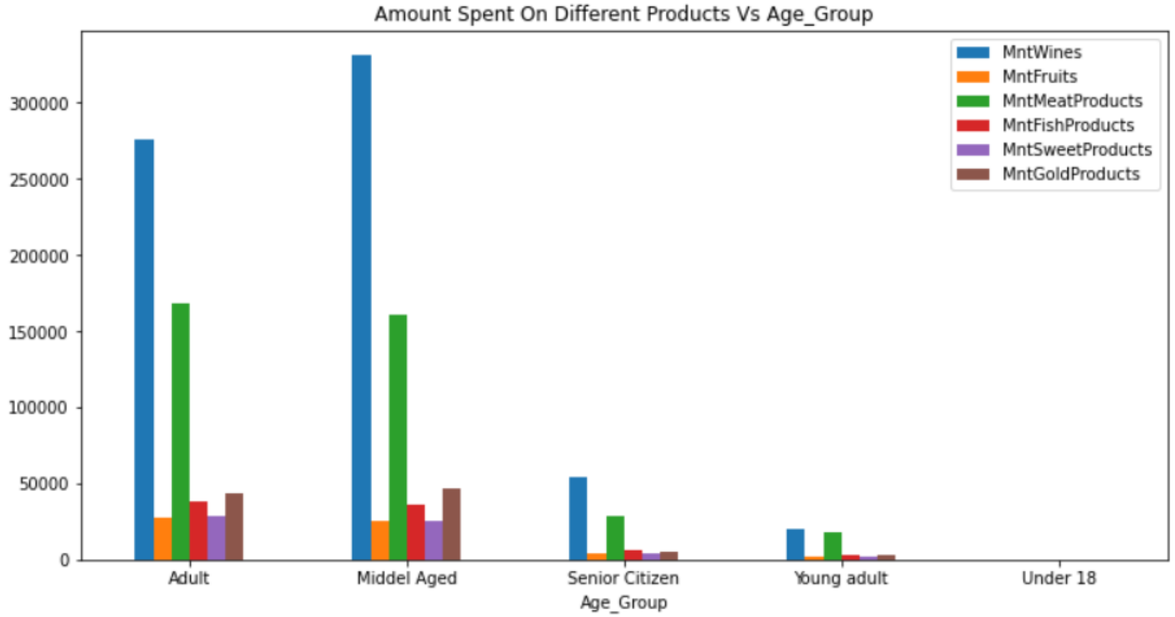


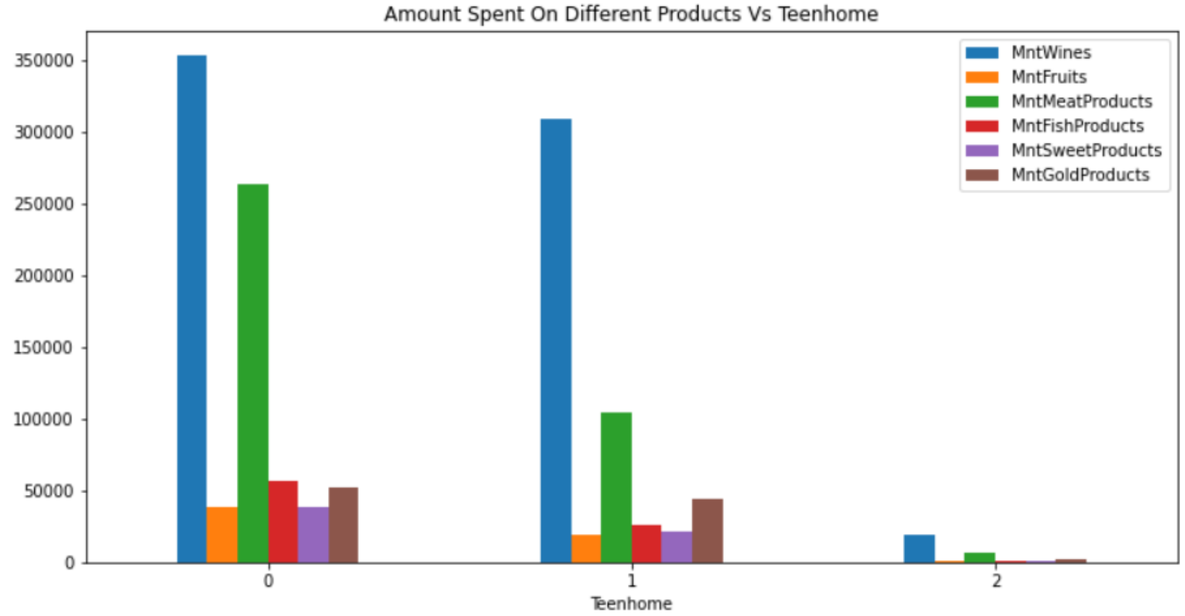


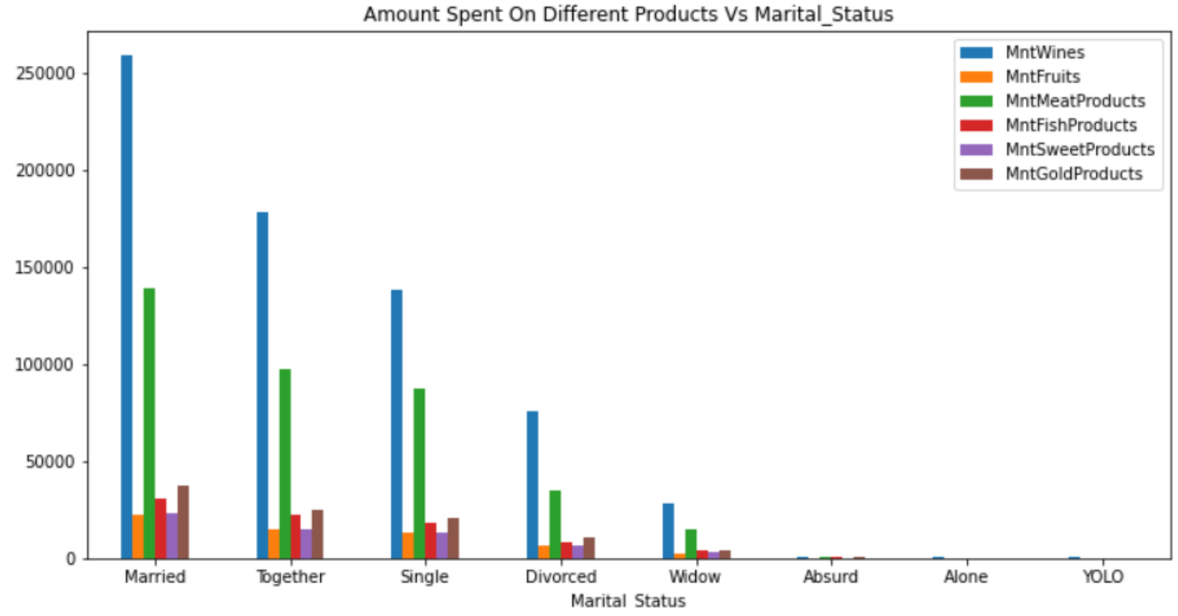
AMOUNT SPENT ON DIFFERENT PRODUCTS BY CUSTOMERS

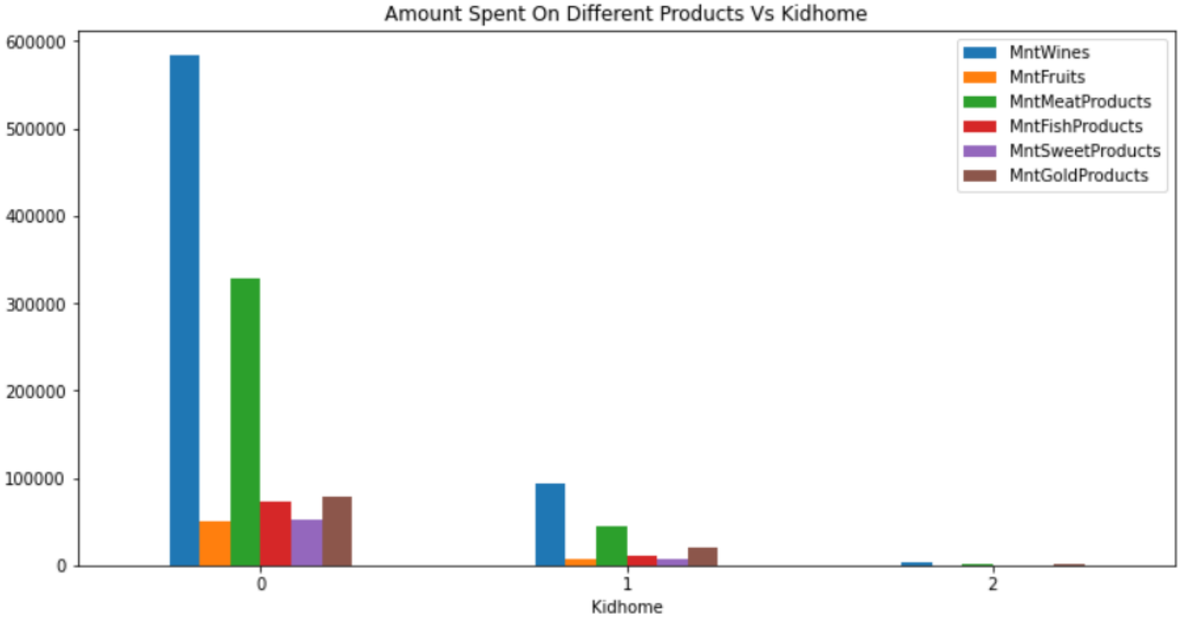


AMOUNT SPENT ON DIFFERENT PRODUCTS BY CUSTOMERS

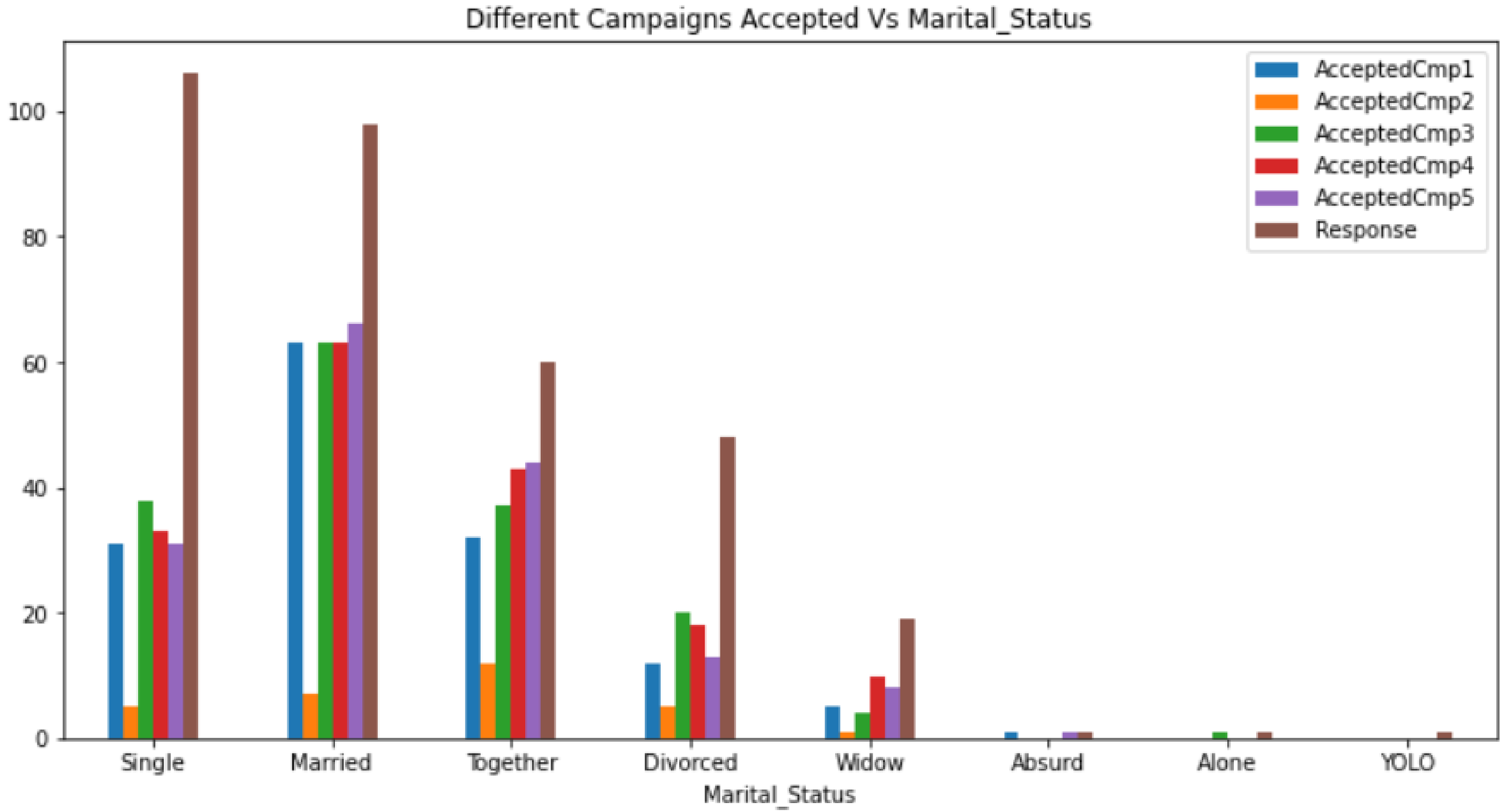




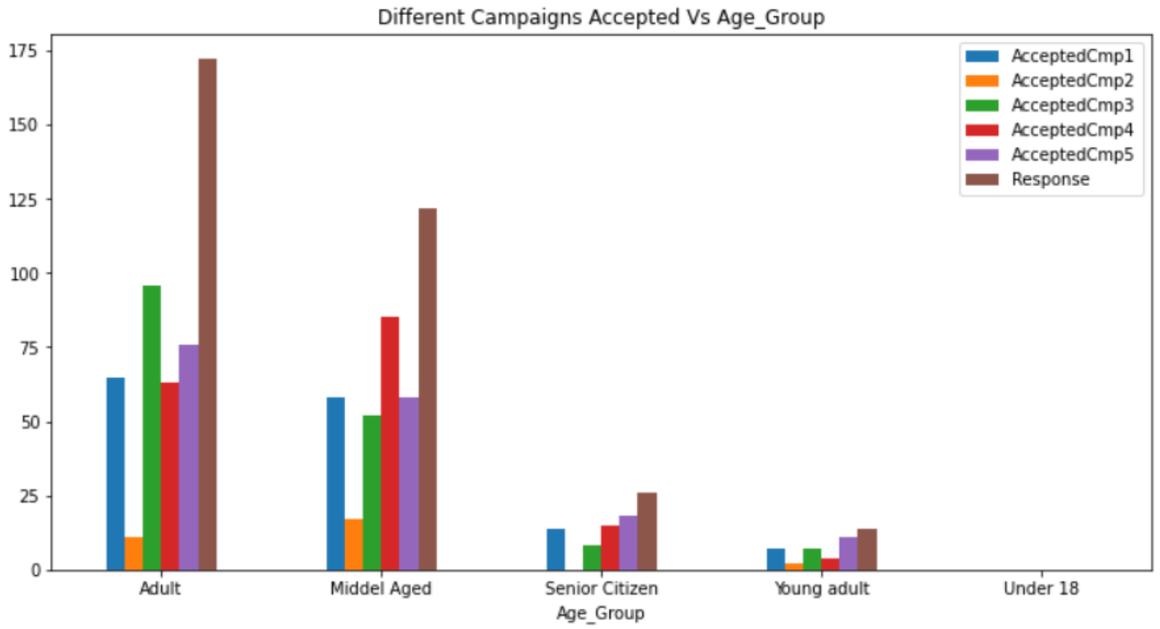


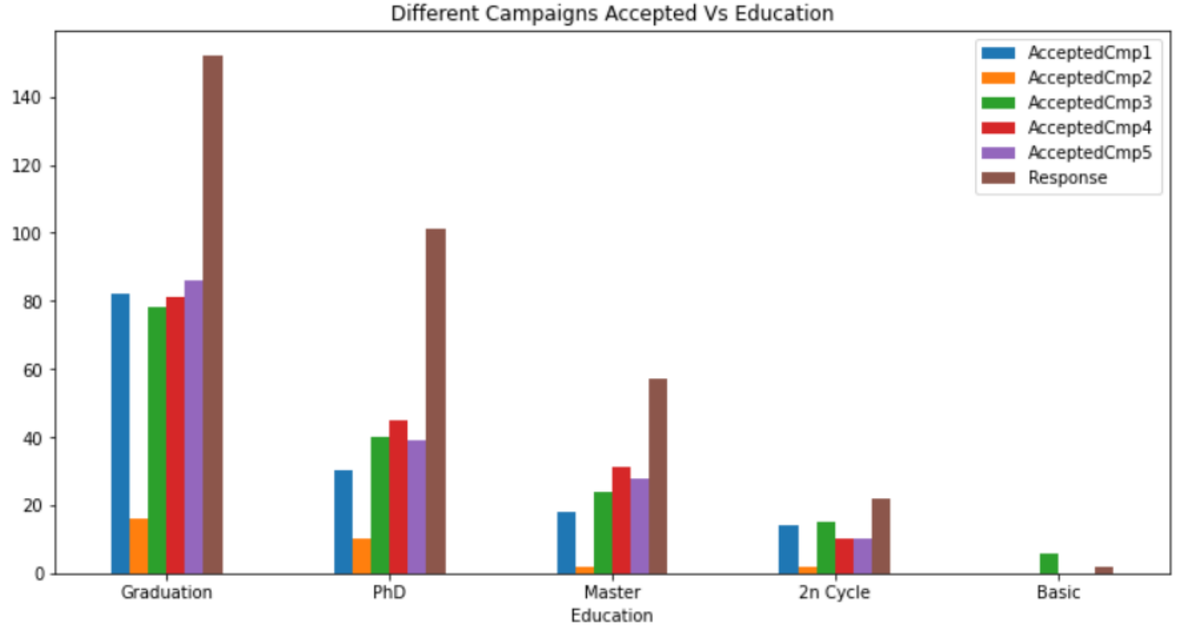


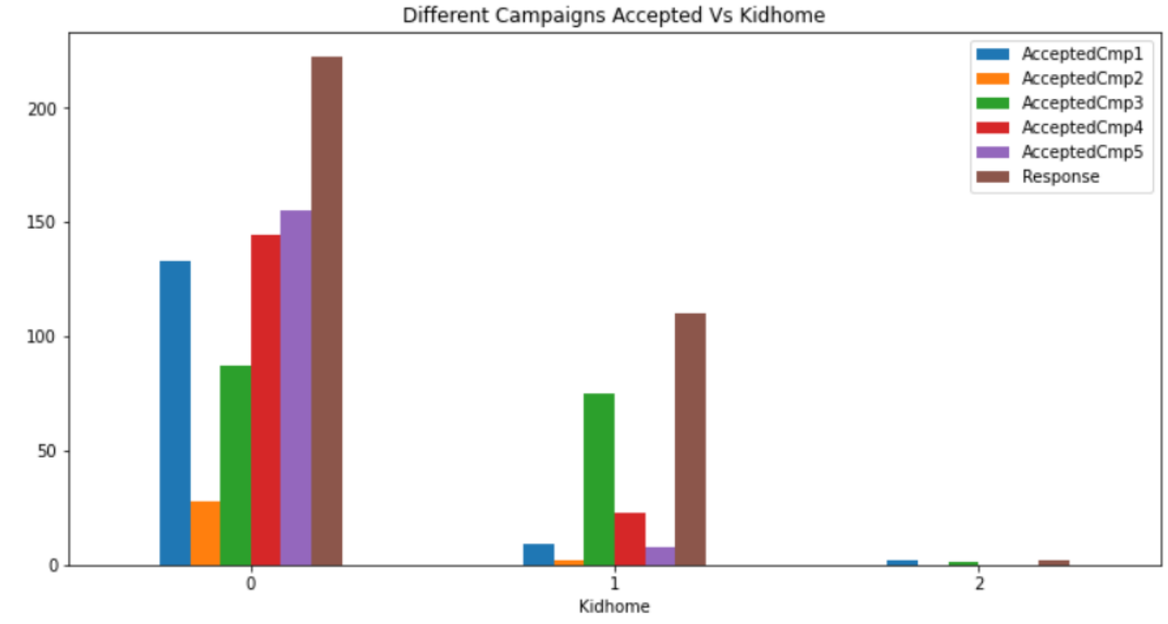
DIFFERENT CAMPAIGNS ACCEPTED BY CUSTOMERS

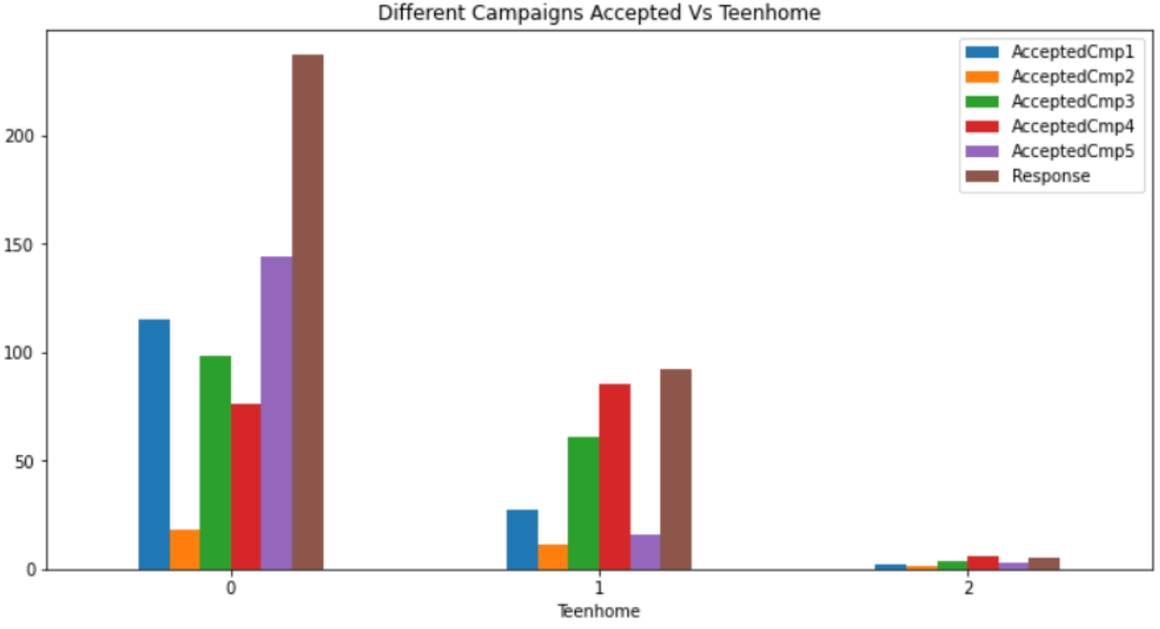


DIFFERENT CAMPAIGNS ACCEPTED BY CUSTOMERS

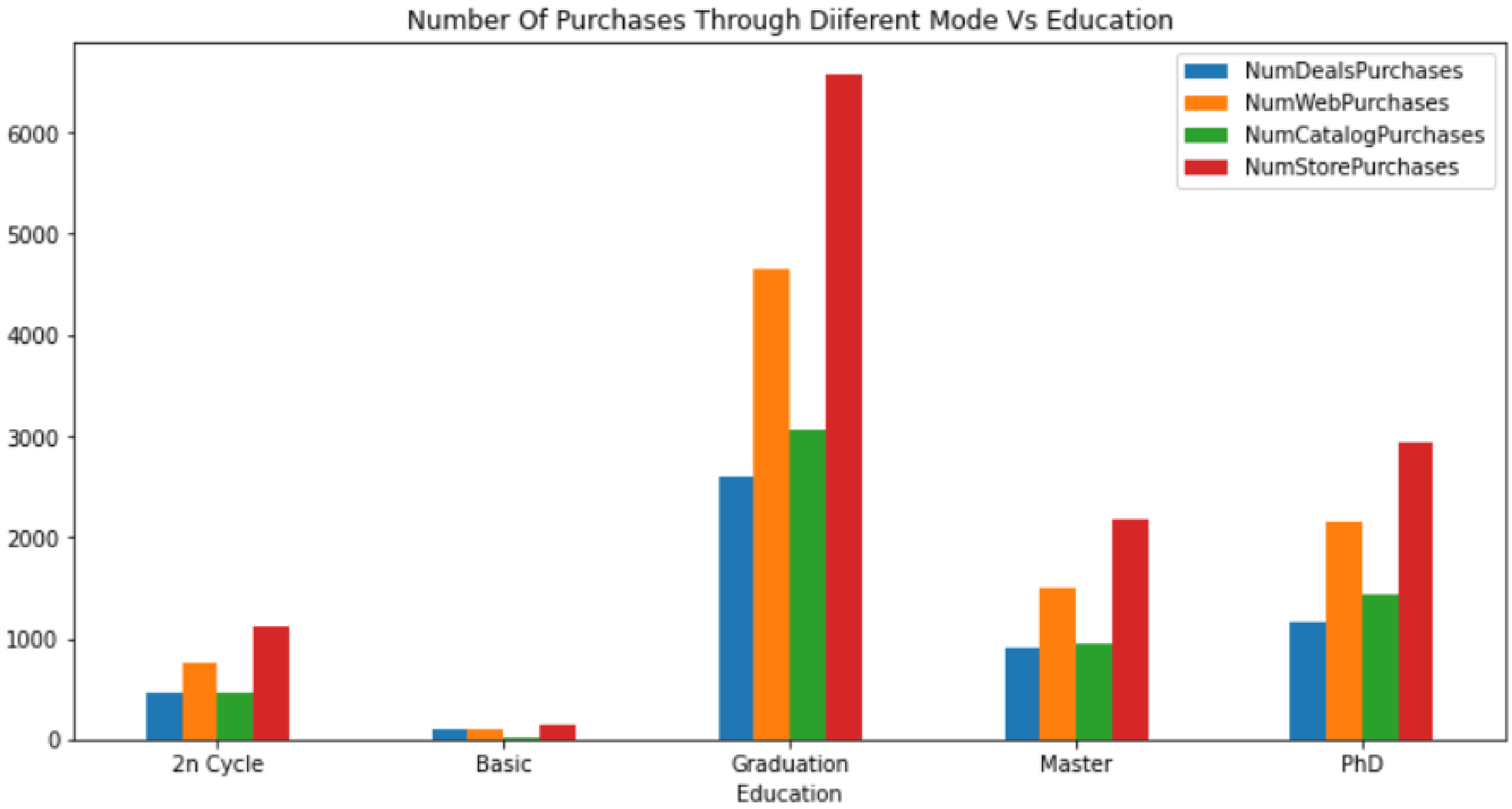




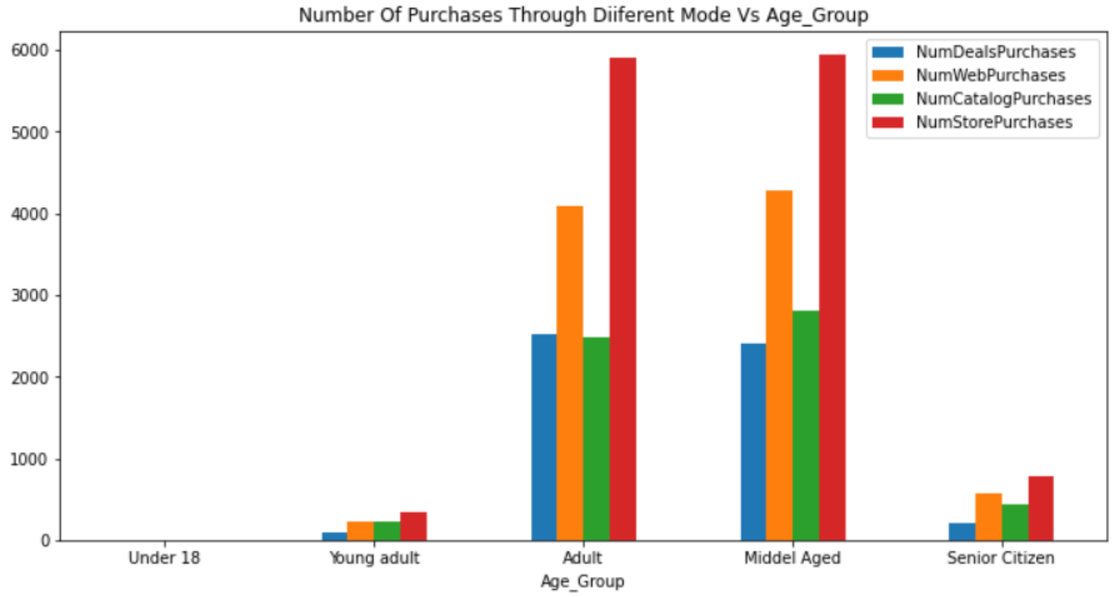


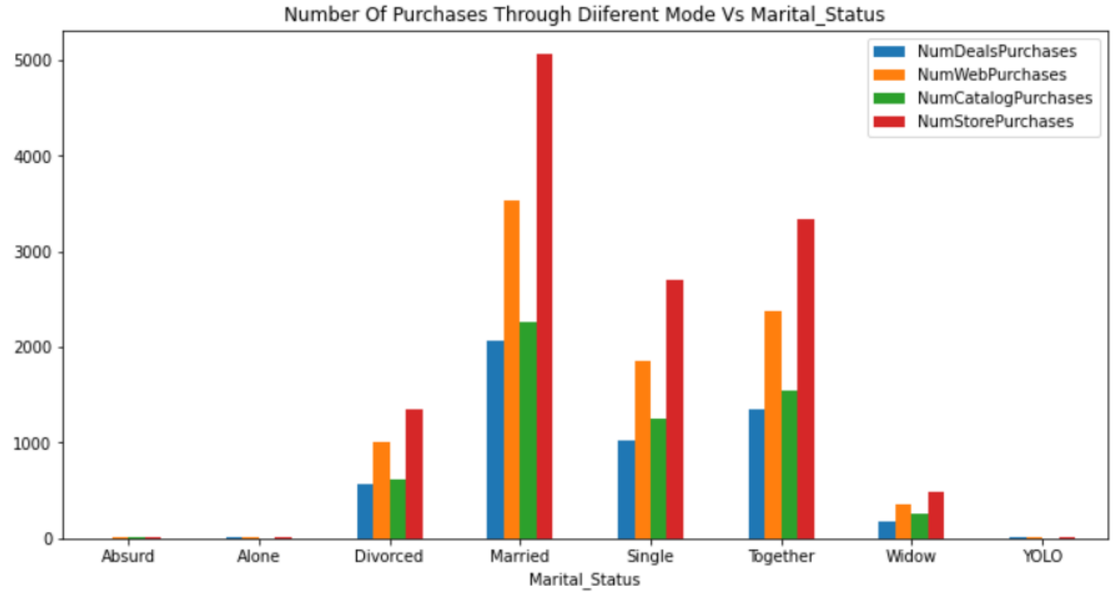


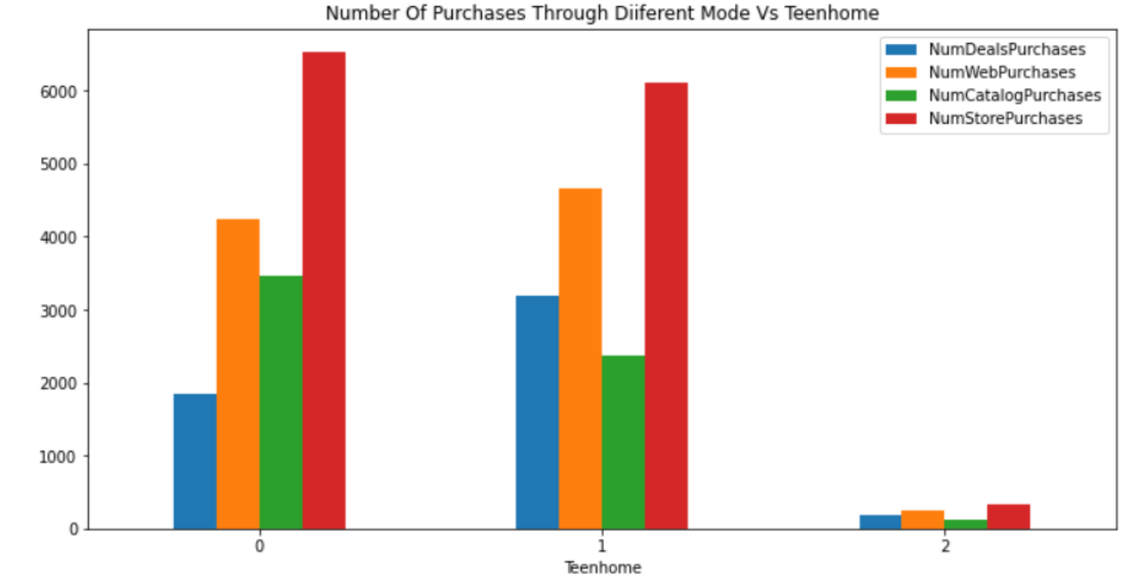
NUMBER OF PURCHASES THROUGH DIFFERENT MODES

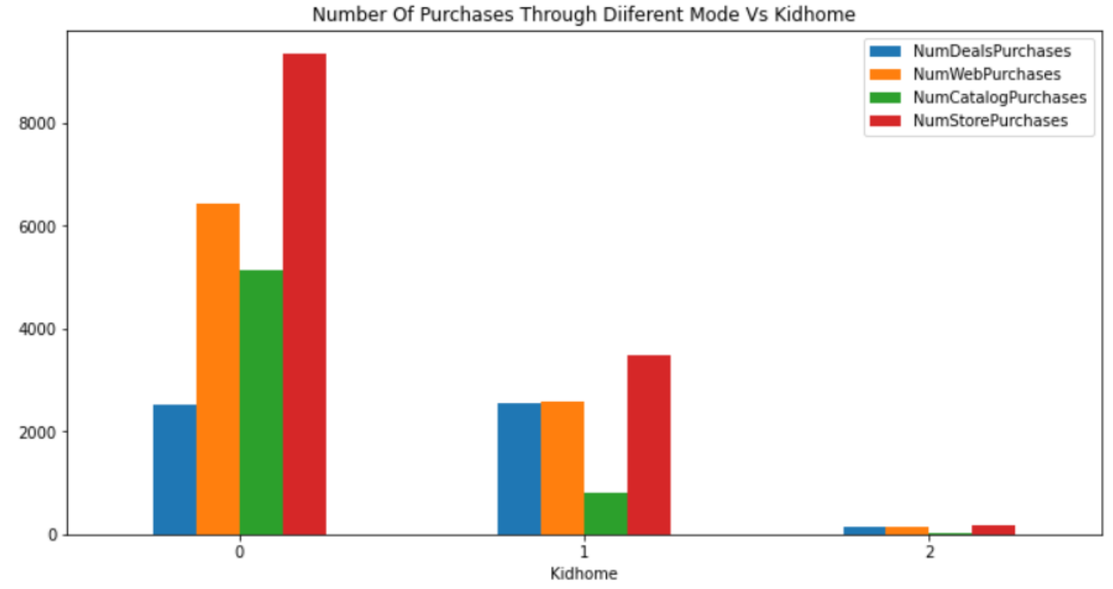


NUMBER OF PURCHASES THROUGH DIFFERENT MODES









CONCLUSION

TARGET AUDIENCE

- Age = 45 to 65
- Education = Graduates & Post Graduates
- Kids = 0
- Teens = 0 or 1
- Marital Status = Married or together
- Income = 30000–800000

MOST REVENUE GENERATING PRODUCTS

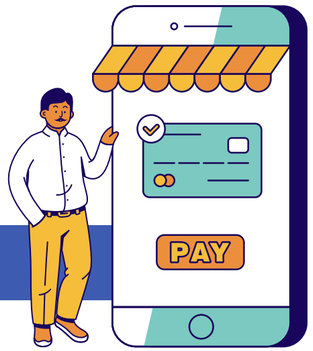
MEAT

WINE

MOST PREFERED PLACES FOR PURCHASING

STORE

WEB

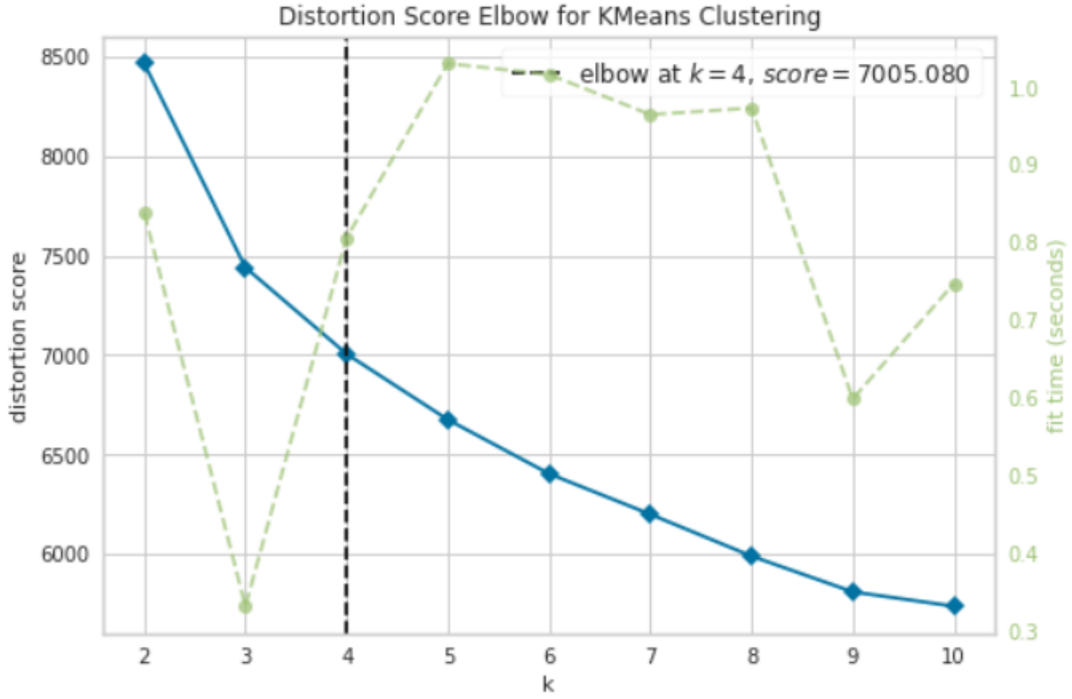


STEPS FOLLOWED AFTER EDA

1. Drop redundant columns
2. Treat outliers with capping
3. Created transformer pipeline

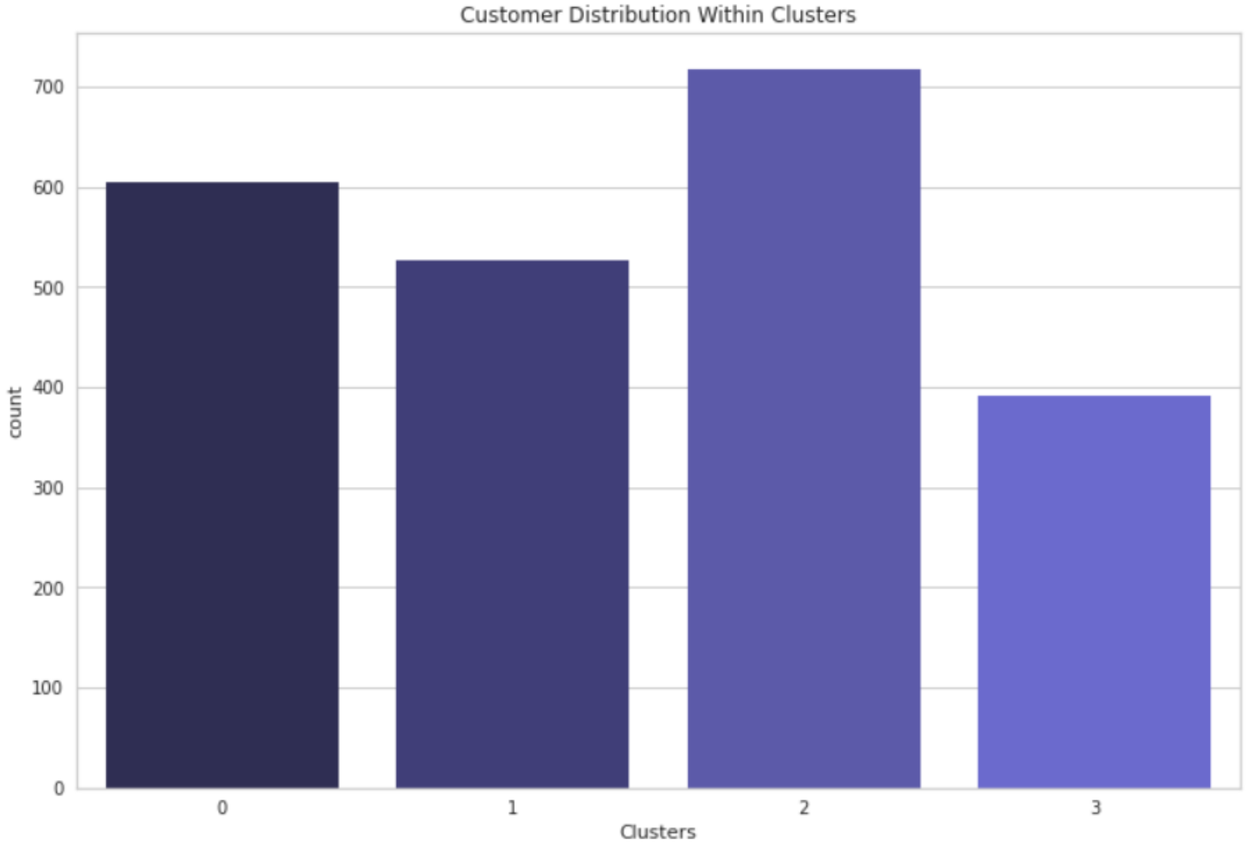
```
ColumnTransformer(transformers=[('num',
                                 Pipeline(steps=[('standardscaler',
                                                  StandardScaler())]),
                                 ['Income', 'Age']),
                                ('ordinal',
                                 Pipeline(steps=[('ordinalencoder',
                                                  OrdinalEncoder(categories=[['Undergraduate',
                                                                              'Graduate',
                                                                              'Postgraduate']]))]),
                                 ['Education_Level']),
                                ('nominal',
                                 Pipeline(steps=[('onehotencoder',
                                                  OneHotEncoder())]),
                                 ['Partner'])])
```

K-MEANS CLUSTERING

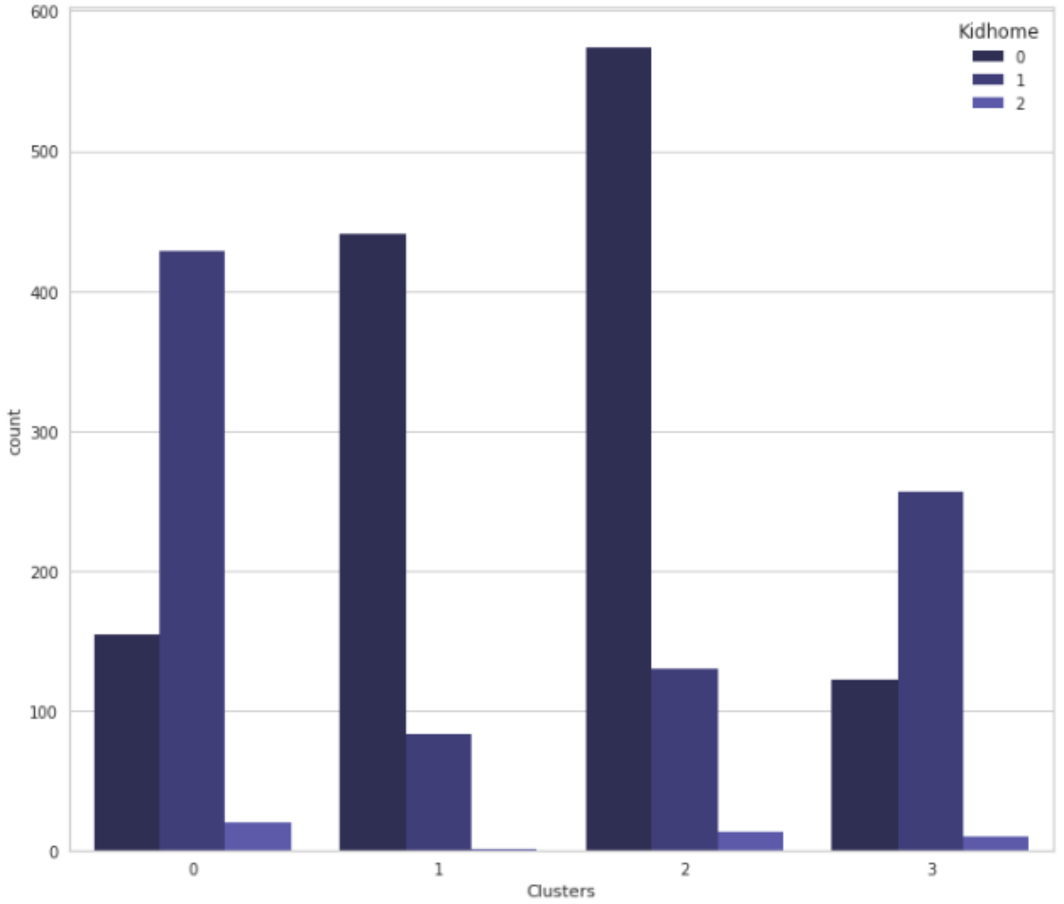


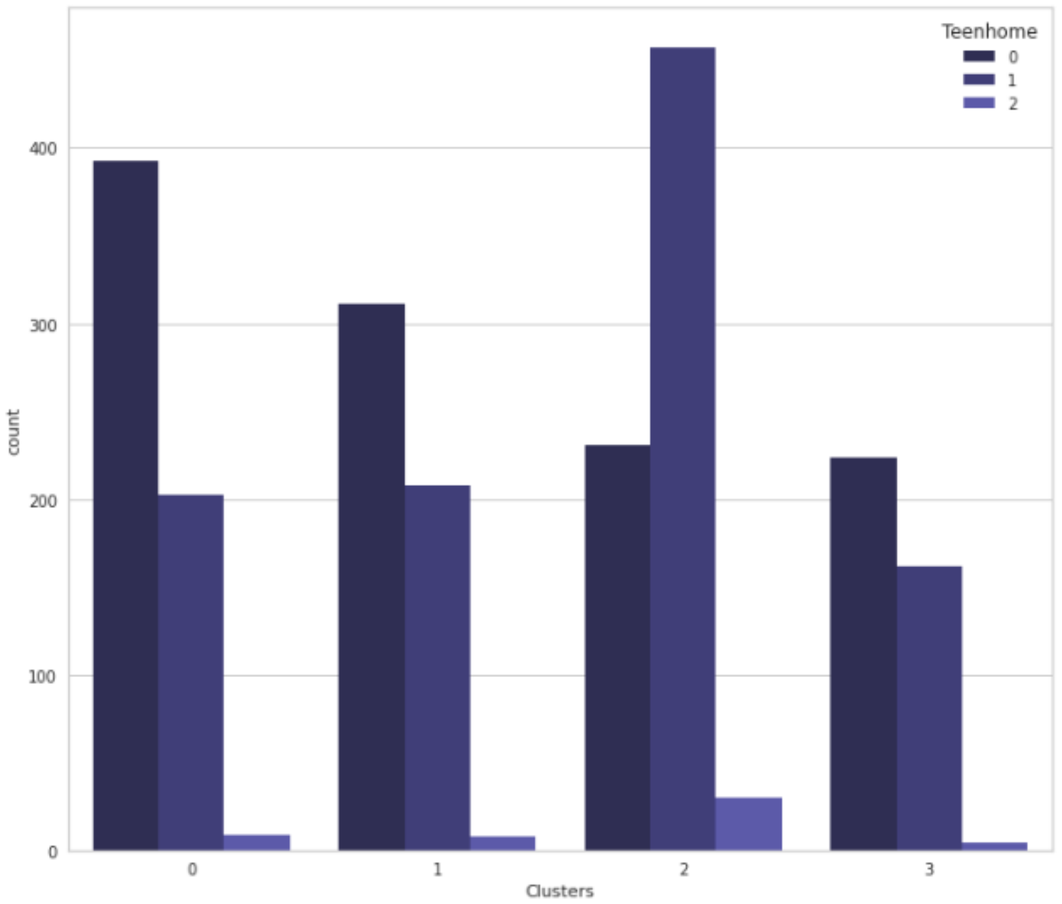
According to Kelbow method we should choose k=4

CLUSTER DISTRIBUTION



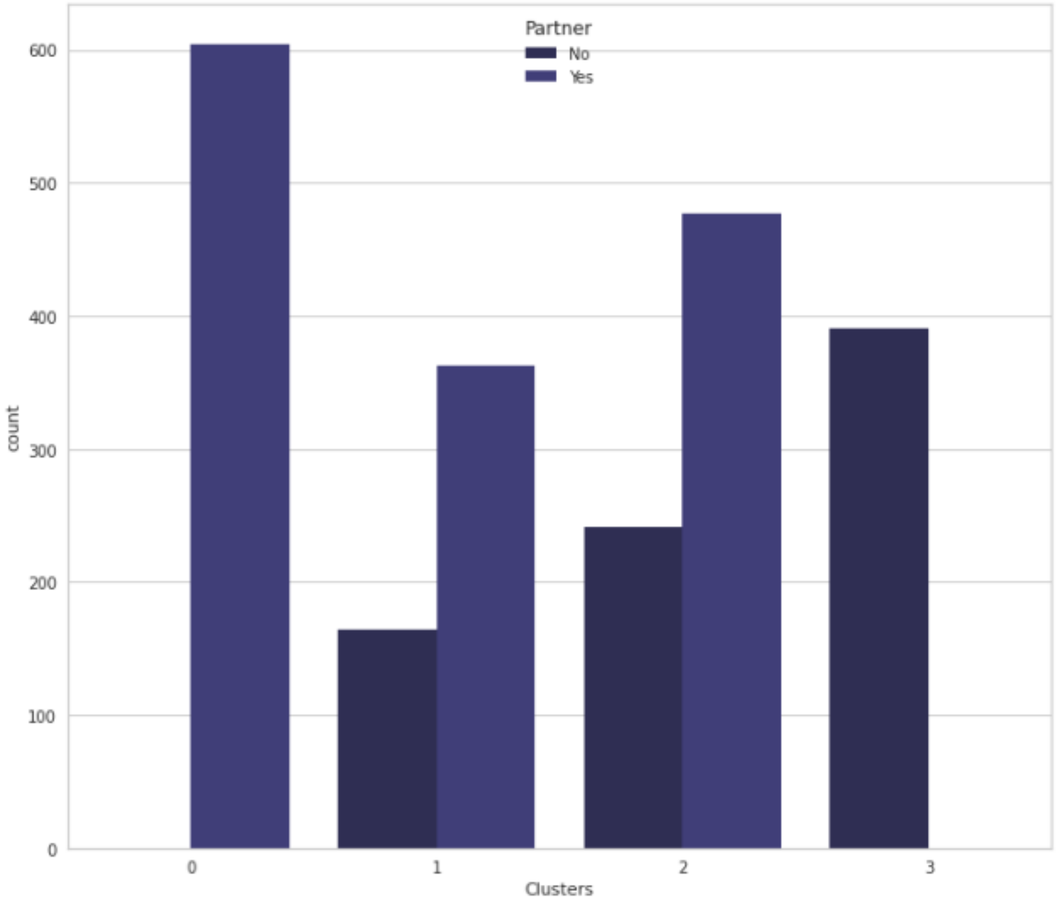
- Cluster 2 has the highest number of customers
- Cluster 3 has the least number of customers



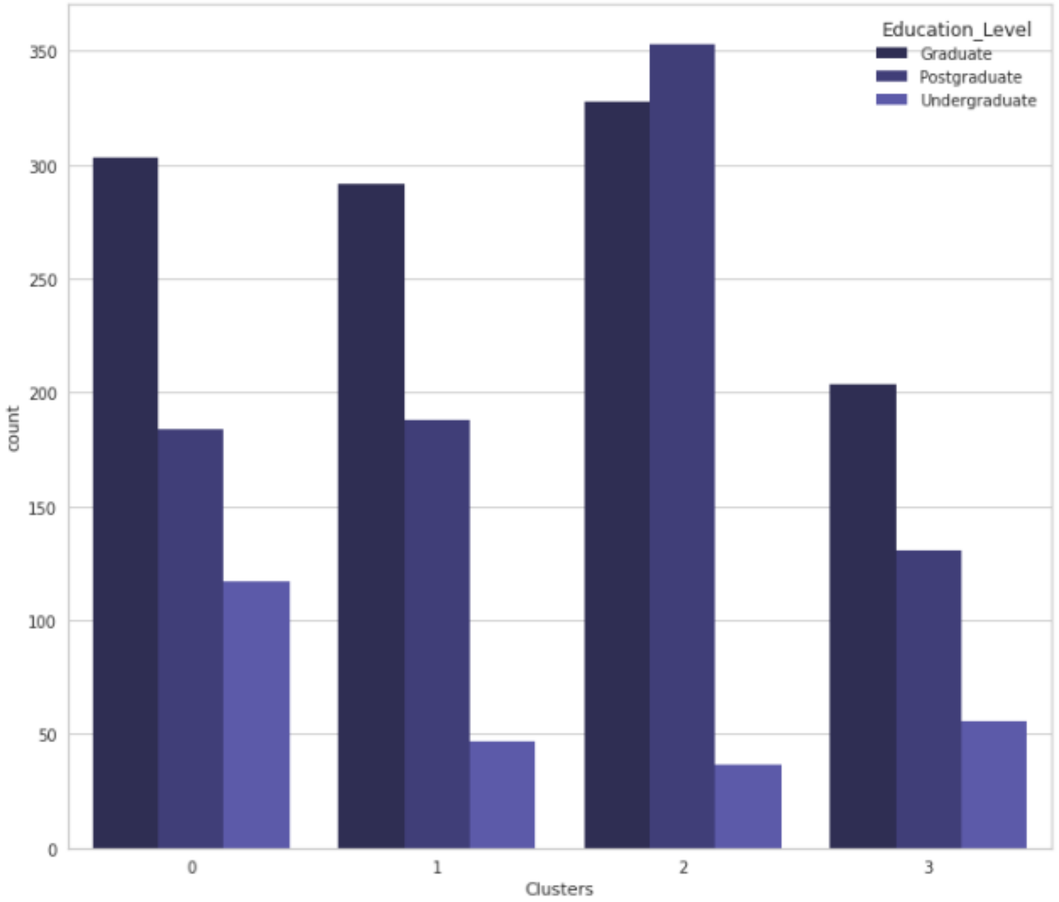


- Cluster 0 mostly has customers with 1 kid in the household
- Cluster 1 has customers with no kids in the household
- Cluster 2 also has a large number of customers with no kids in the household
- Cluster 3 has customers with 0 and 1 kid in household

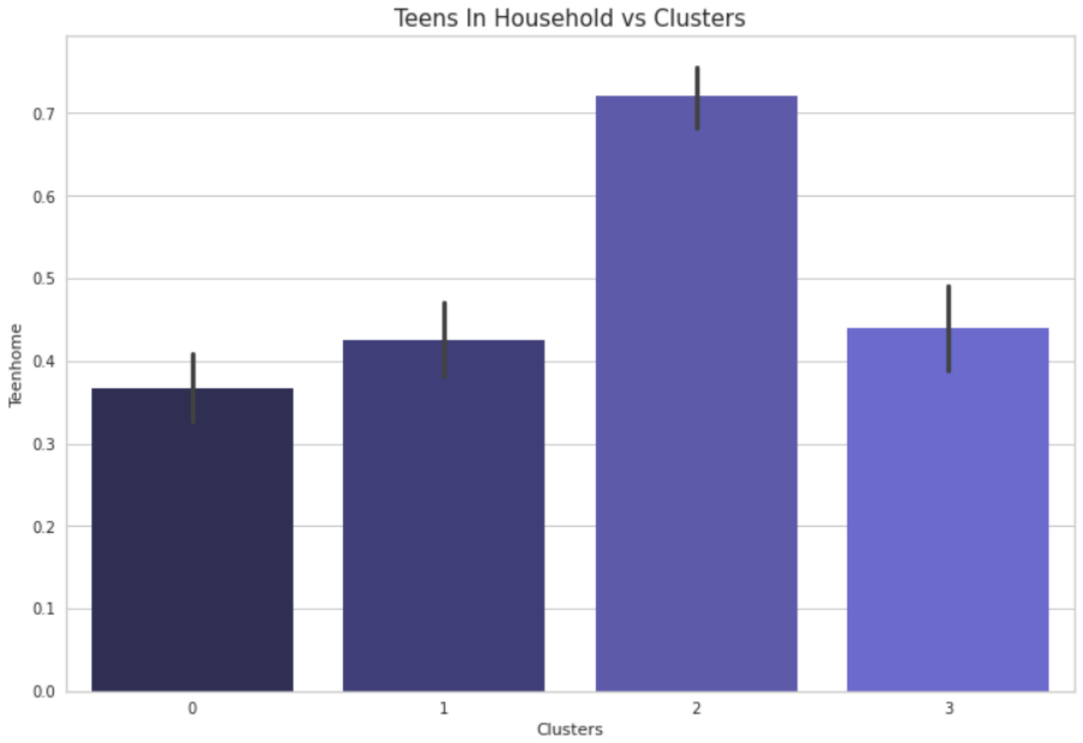
- Cluster 0 consists of customers with no teen in the household & few of them have 1 Teen in household
- Same goes for the cluster 1 & 3
- Cluster 2 has customers with 1 Teen in household

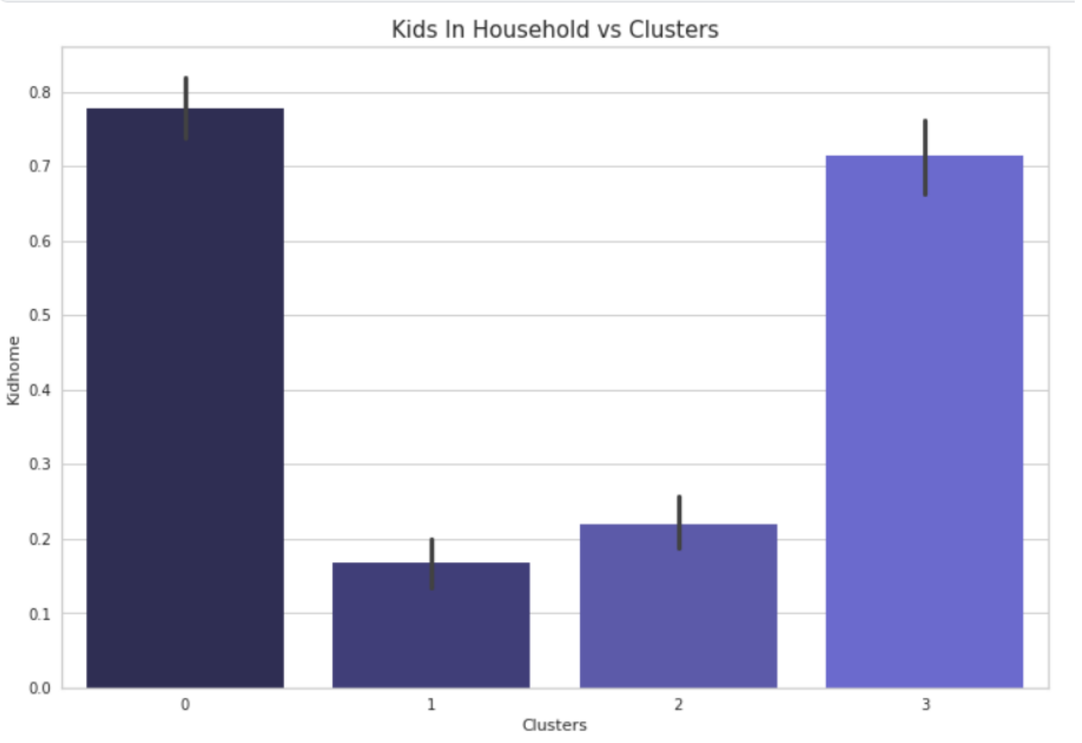


- All the customers in cluster 0 have a partner
- All The customers in cluster 3 have no partner
- Cluster 1 & 2 has customers with and without a partner, but most of them have a partner



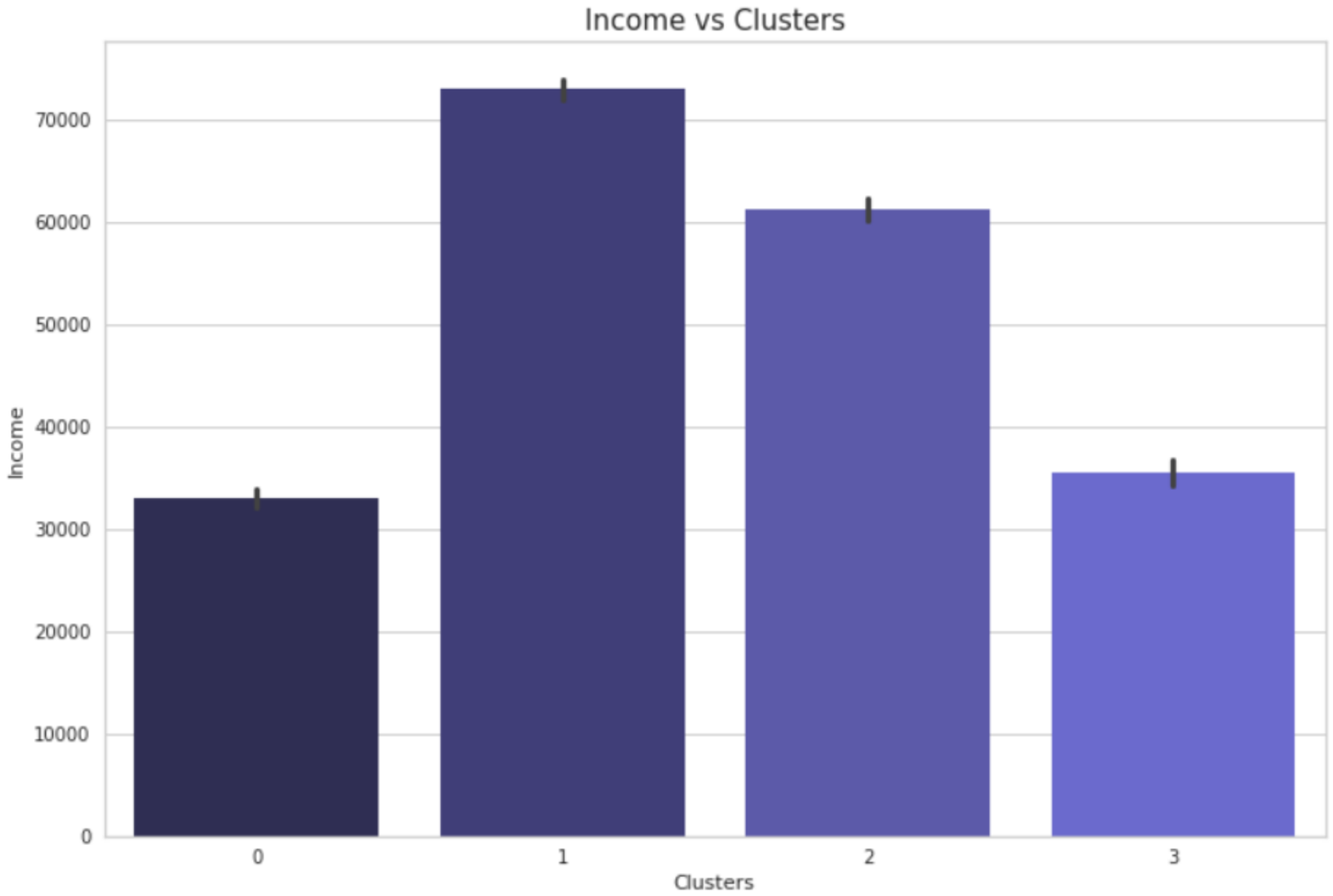
- All clusters have customers with a graduate, postgraduate and undergraduate background
- All clusters have less number of customers with an undergraduate background
- Cluster 2 has the highest number of postgraduates and graduates



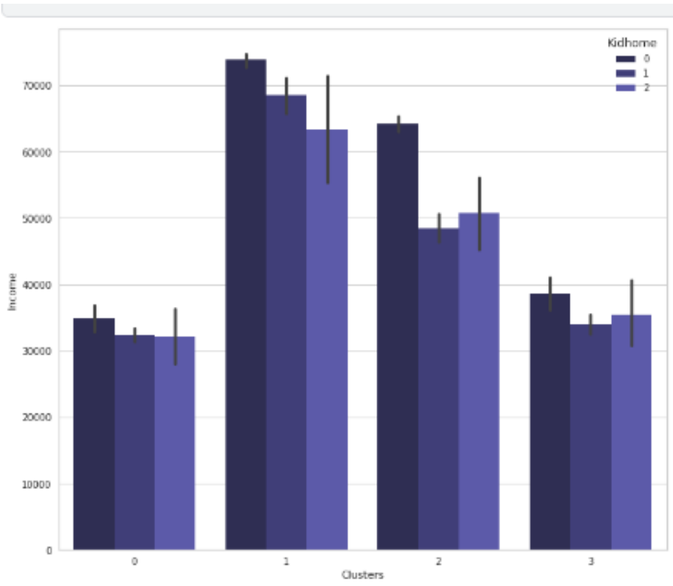


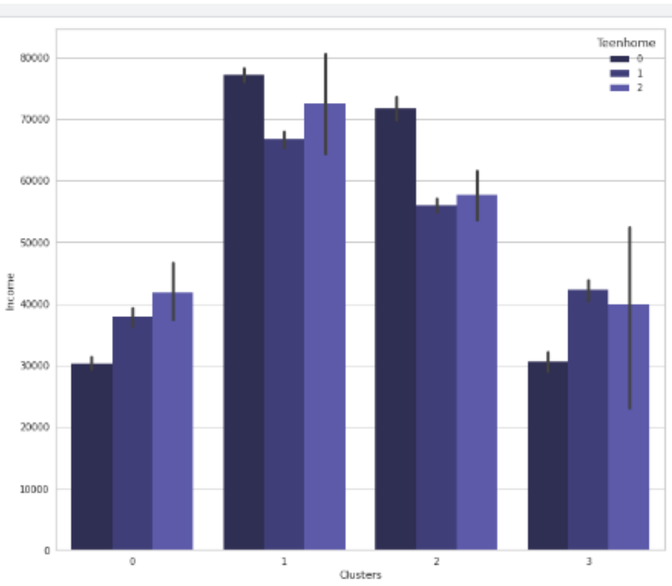
- Cluster 0 and 3 has the maximum number of customers with kids in the household
- Cluster 1 and 2 has the least number of customers with kids in the household

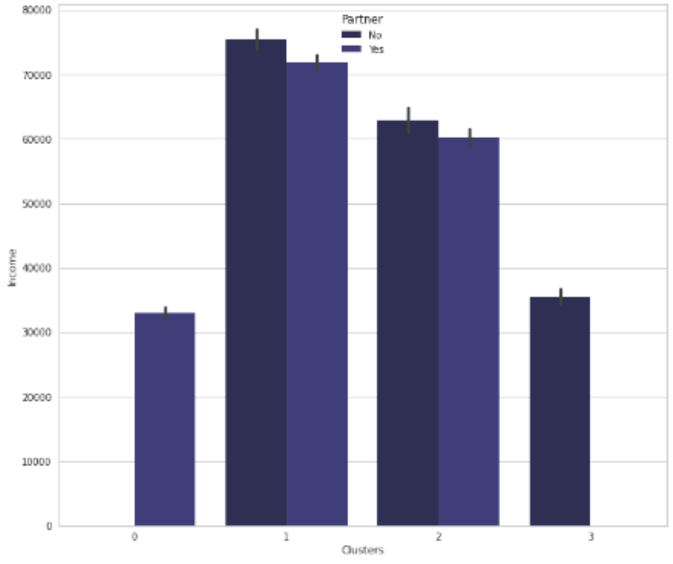
- Cluster 2 has maximum number of customers having Teens in household
- Remaining Clusters also have customers with Teens in household but they are less as compared to cluster 2

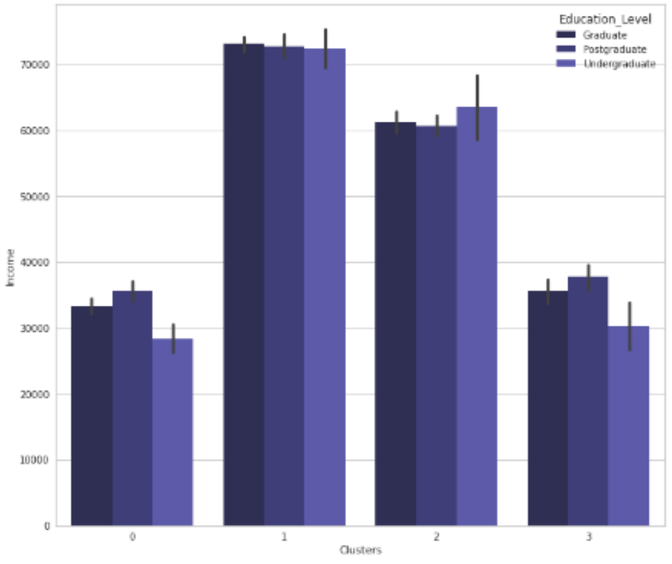


- Cluster 1 has high Income followed by cluster 2. Which is little strange cause cluster 2 has highest number of customers and most number of post graduates & graduates as compared to cluster 1
- Cluster 0 and 3 has least income









- 1. The very interesting thing I can see from the first 2 plots is, as We saw before the number of customers with 2 kids or teens is very much less still their income is similar to the customers with no kids or teens or who have 1 kid & teen. So I conclude that the customers with more than 1 kid or teen in the household have high household income 1

- 2. From the last 2 plots, I think education_level and partner have nothing to do with income. Undergraduates are earning equal to or more than graduates and postgraduates within each cluster.

Cluster 1

- fewer customers but with the highest income
- No kids, few have 1 Teen
- Graduates & postgraduates
- Most of them have partner

Cluster 2

- Max number of customers & high income
- No kids, few have 1 or 2 Teen
- High number of postgraduates & graduates
- Most of them have partner

Cluster 0

- Least income
- 1 kid & few have 1 Teen also
- Graduates & postgraduates but also has most undergraduates than any cluster
- All have partner

Cluster 3

- fewer customers & less income
- 1 kid & few have 1 Teen also
- Graduates & postgraduates
- All have no partner