Analysis of Black Friday Sales

14.07.2019

Amit Verma, Gunnika Singh Sobti, Karan Patel, Ravinder Singh, Sparsh Sharma

Goal

Our aim was to find the relationship between number of purchases on Black Friday sales and variables such as occupation, marital status, age groups, gender, city category and number of years lived in a city. Another objective was to establish a market basket of items from product category A, B and C.

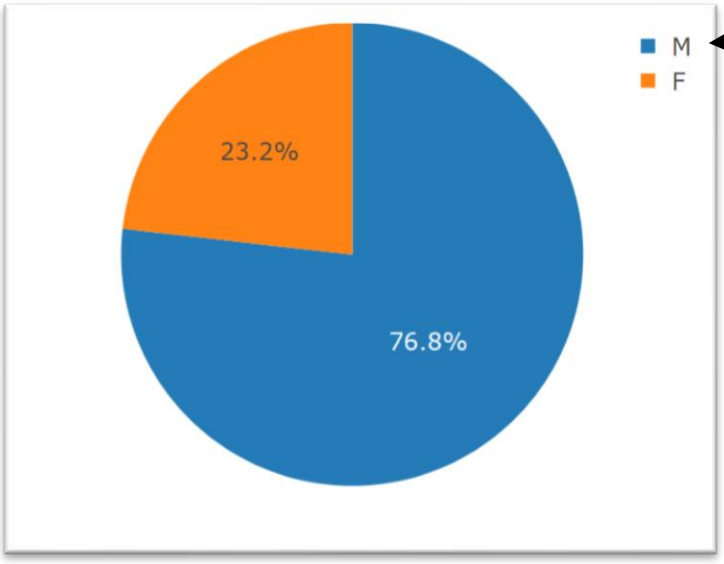
Key Findings

- Number of Male Shoppers > Female Shoppers.
- People in age range 26-35 purchase most.
- People generally spent over $7500 in sales.
- Item 8 of product category A, item 5 of product category A, and combination of items 5-8 from Product category A and B respectively had the highest frequency of purchase among Males and Females.
- Approximately 0.9% of the variation in number of items purchased could be explained by Gender + Age + City Category where the p-value[1] = 2.2e-16.

Purchase by Consumers

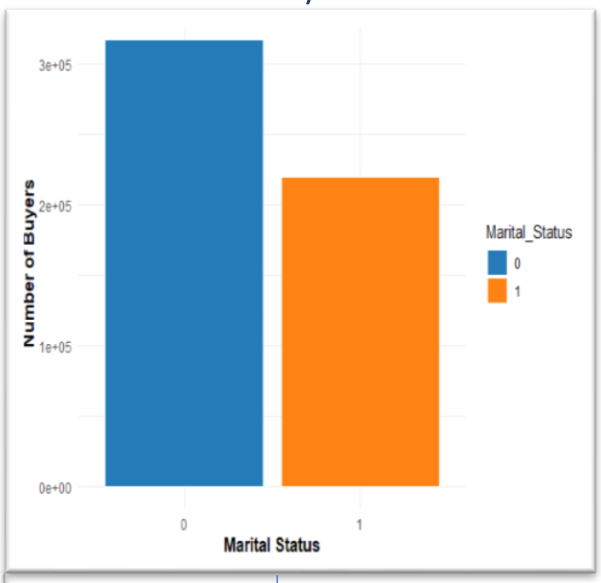
[1] The p-value for each term tests the null hypothesis that the coefficient is equal to zero (no effect). A low p-value (< 0.05) indicates that you can reject the null hypothesis. In other words, a predictor that has a low p-value is likely to be a meaningful addition to your model because changes in the predictor's value are related to changes in the response variable.

Purchase Distribution of Consumers Based on Gender



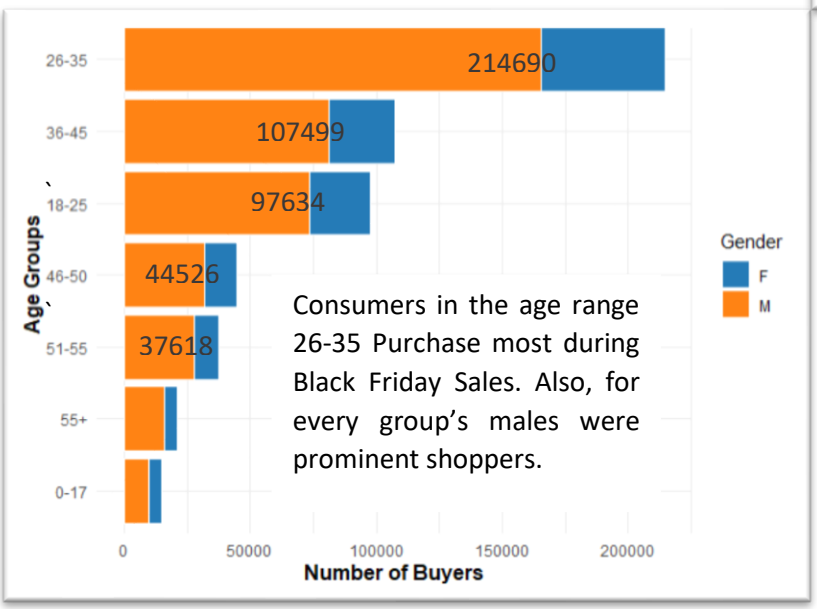
It is clear from the pie chart that Male shoppers at the black Friday spent considerably more money than Females. (or the payment is done by Males.)

Purchase Distribution by Marital



More Unmarried people buy products on Black Friday as compared to the married ones, but do they spend more?

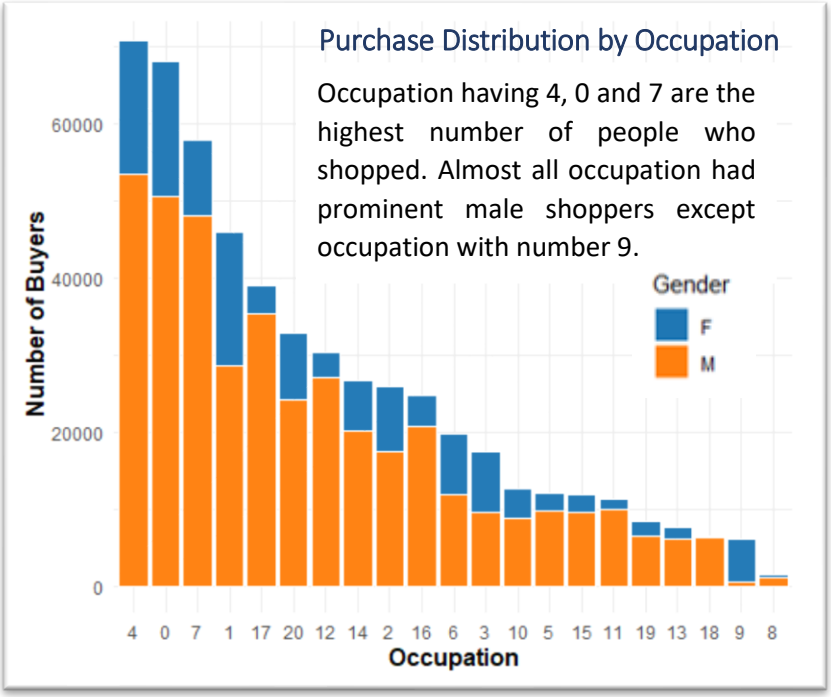
Items purchased by consumers based on Age Groups



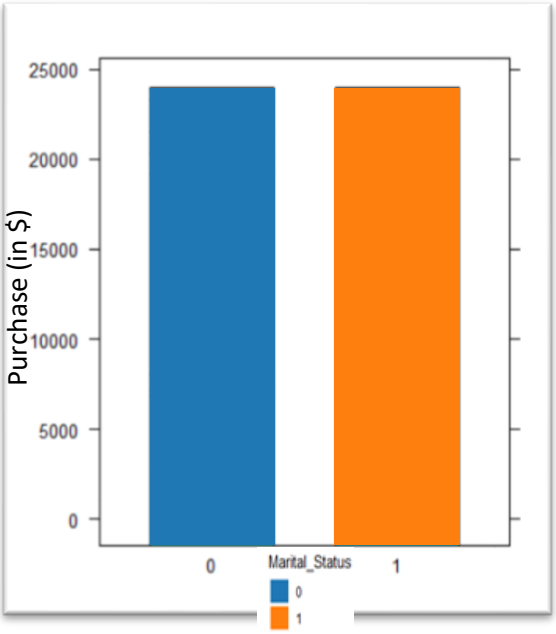
Consumers in the age range 26-35 Purchase most during Black Friday Sales. Also, for every group's males were prominent shoppers.

Although, we saw more single people buy products on Black Friday but looking at the purchase distribution they spent almost equal amount of money on products.

Purchase Distribution by Occupation

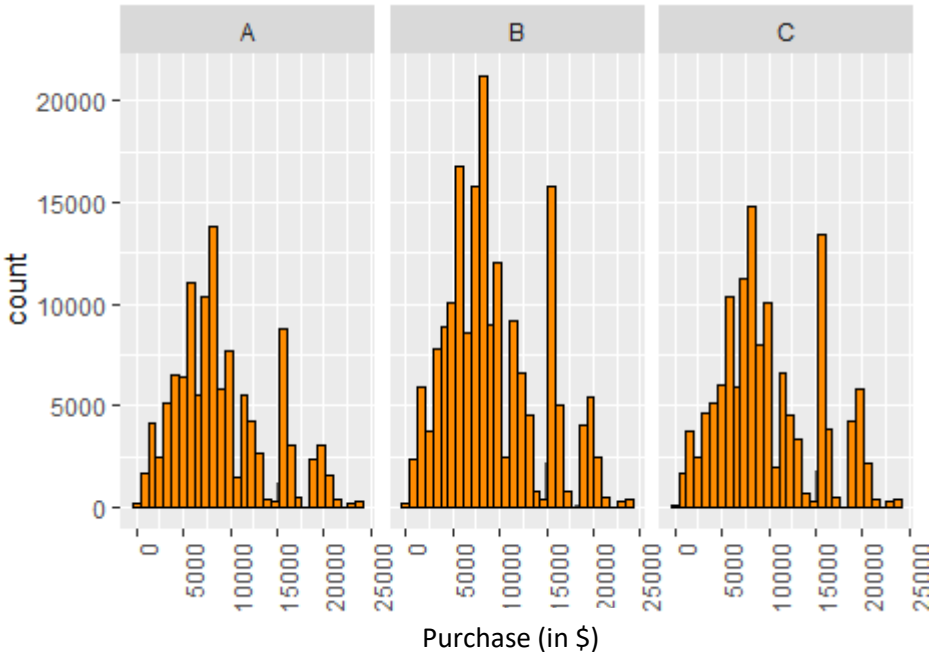


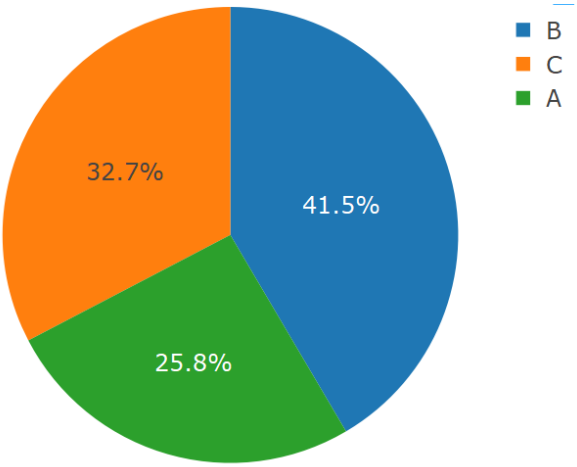
Occupation having 4, 0 and 7 are the highest number of people who shopped. Almost all occupation had prominent male shoppers except occupation with number 9.

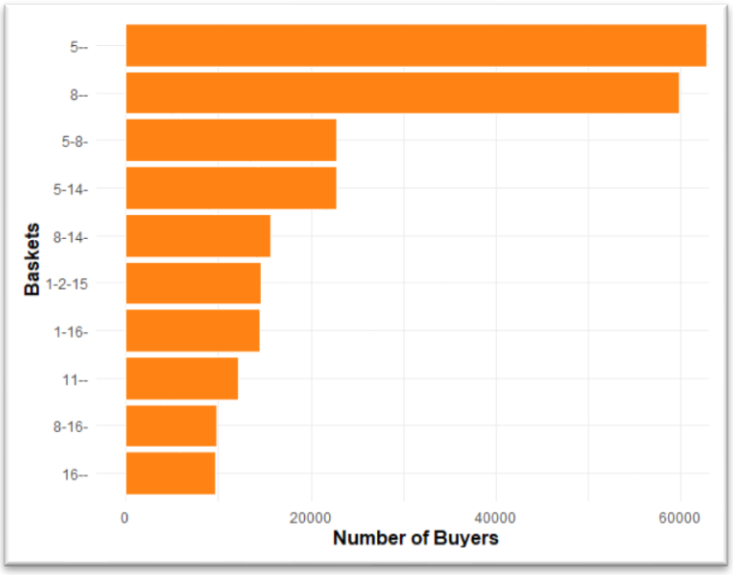


Purchase Distribution by City Category

The graph and pie chart both show that maximum number of items were purchased by consumers residing in City B i.e. approximately 42% of purchases were made by residents of city B.





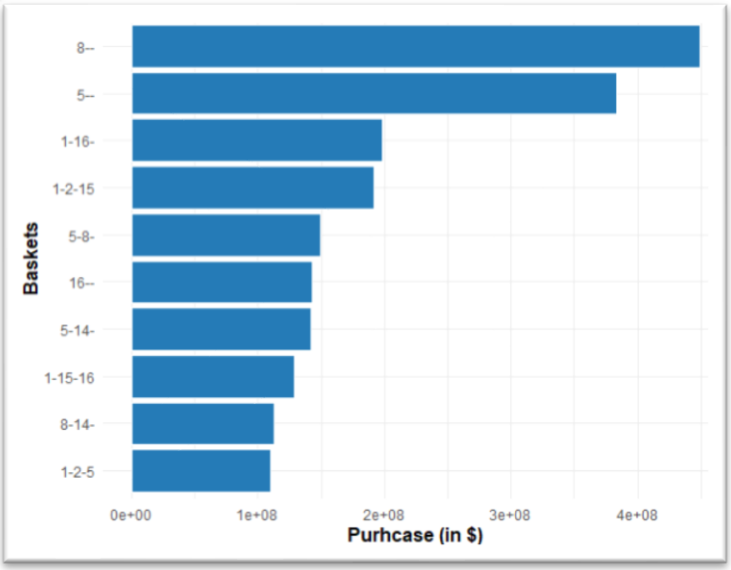


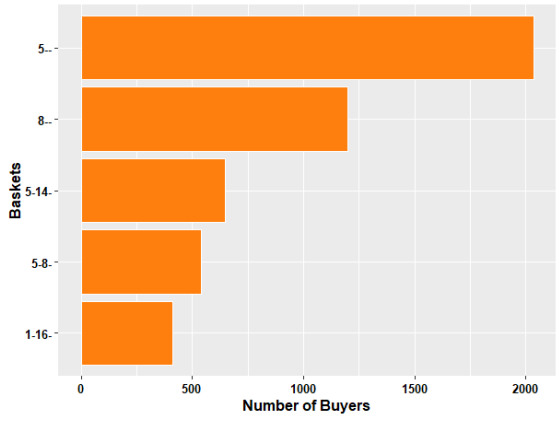
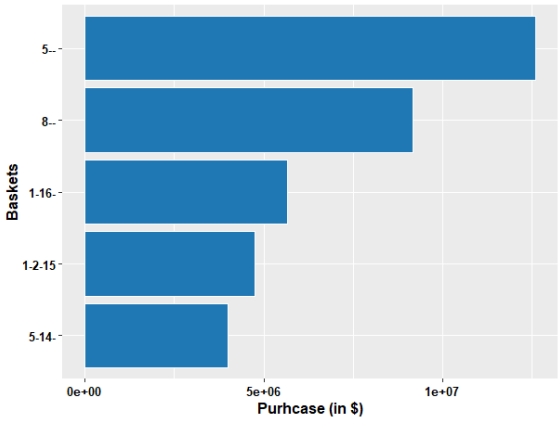
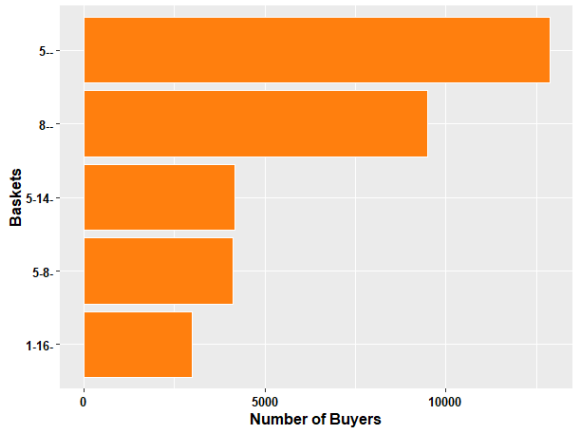
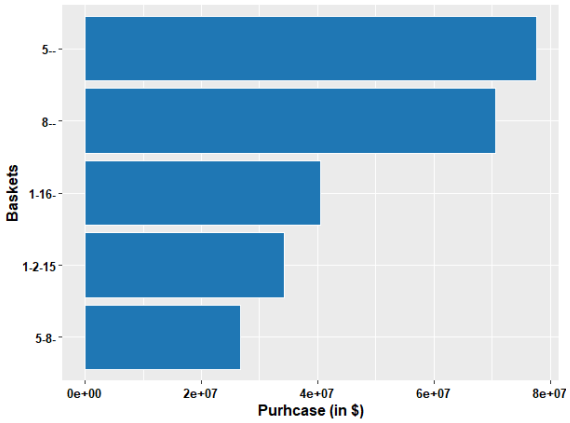
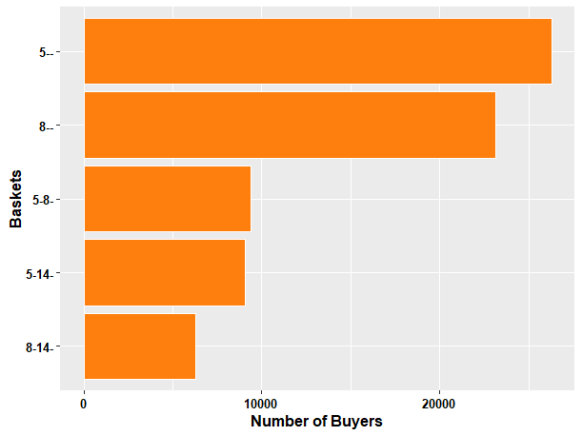
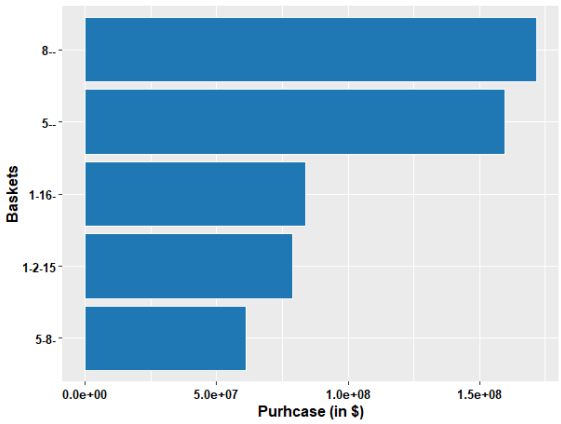
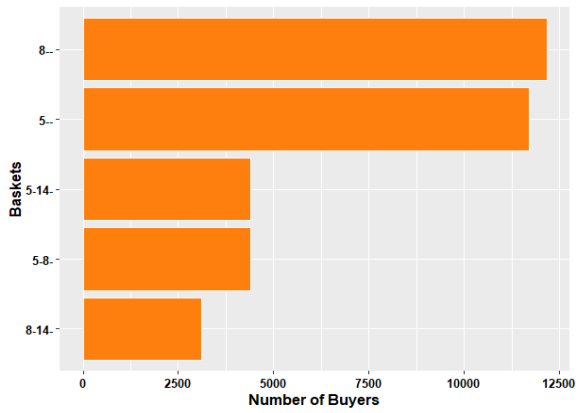
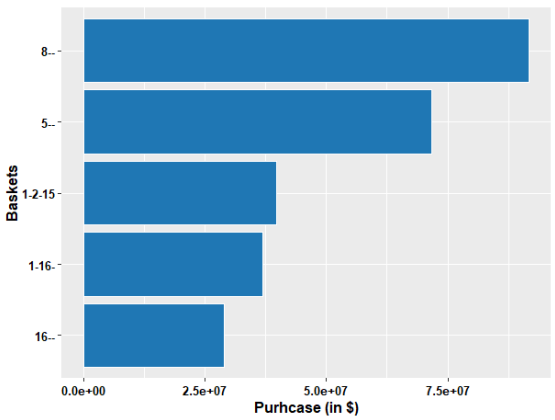
From Basket analysis of buyers on Black Friday sales, single category baskets (5 – –) and (8 – –) were the most preferred followed by double category baskets (5 – 8 –), (5 – 14 –) and (8 – 14 –) and a three category basket which was most preferred was (1 – 2 –15).

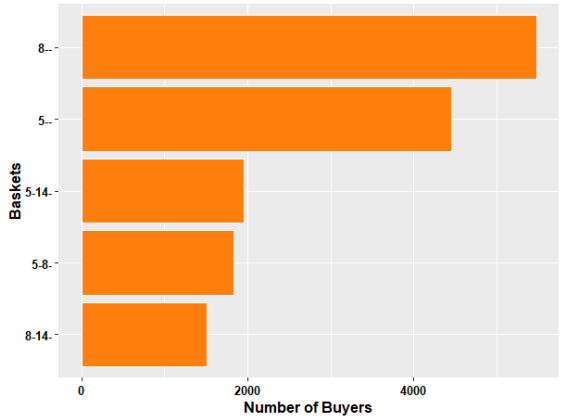
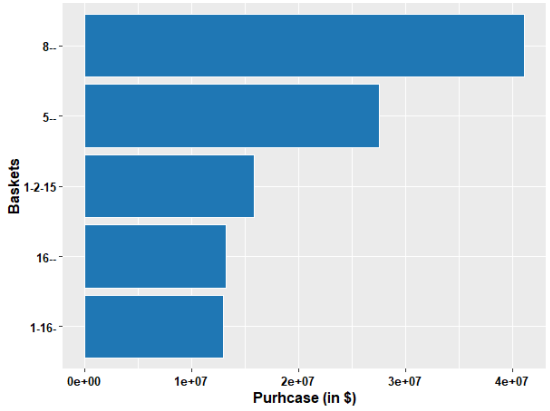
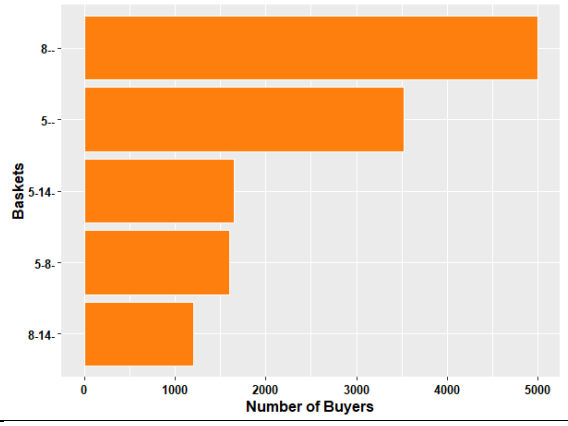
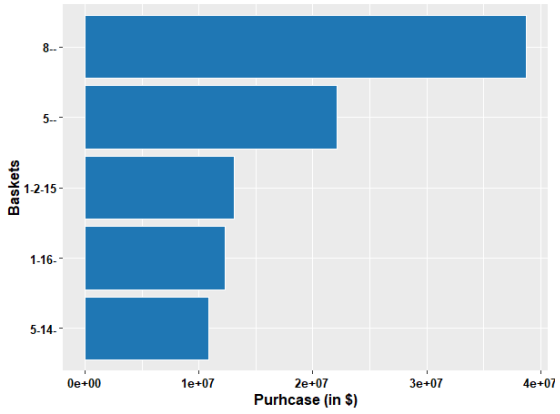
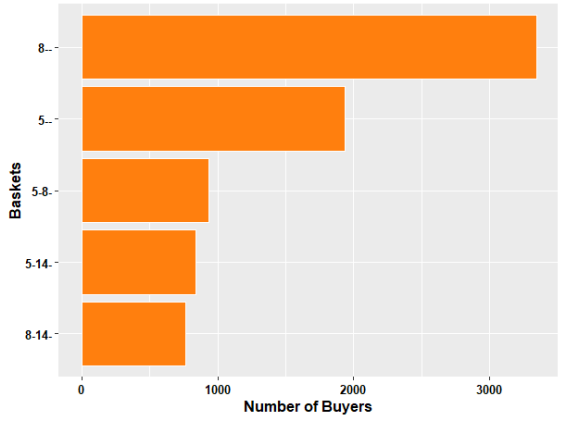
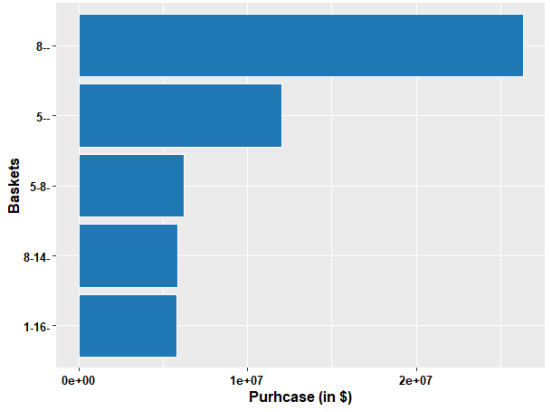
Baskets	Percentage of Buyers
5 – –	11.7%
8 – –	11.1%
5 – 8 –	4.24%
5 – 14 –	4.23%
8 – 14 –	2.92%
1 – 2 – 15	2.73%

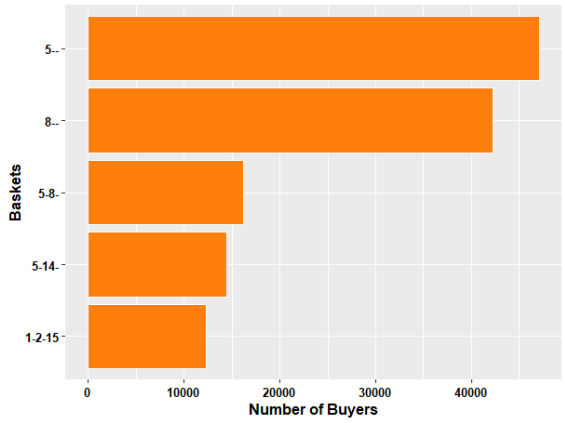
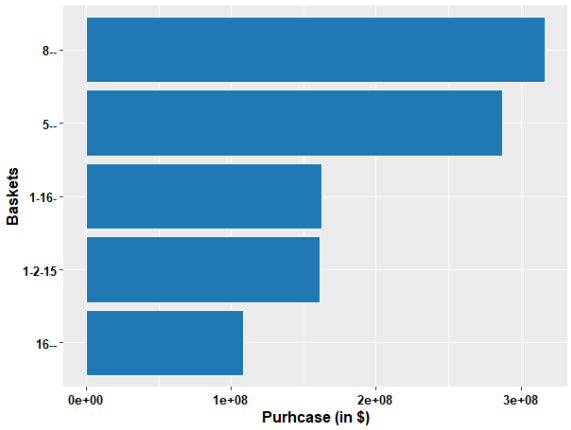
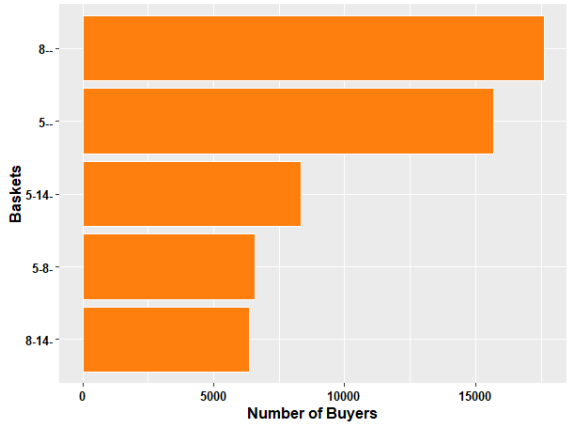
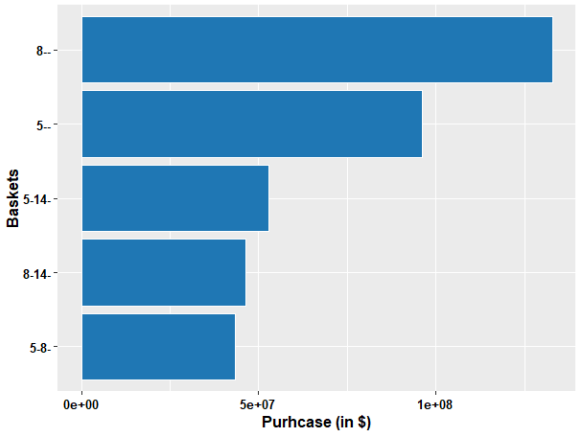
Even though baskets (5 – –) was the most preferred basket but it was basket (8 – –) that contributed in terms of sales. Also, basket (1 – 16 –) was not much prominent in terms of preference, whereas, sale figures show a different story and it contributed being

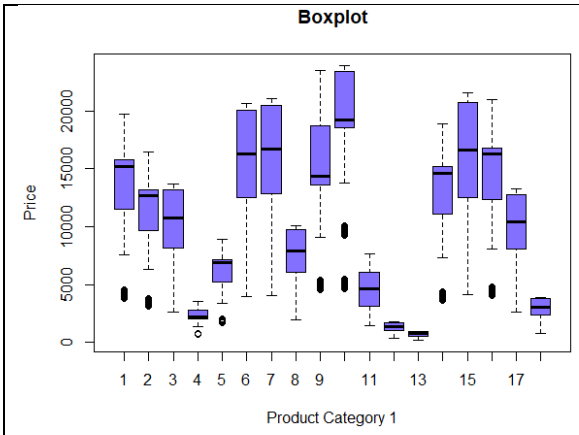
Baskets	Purchase Distribution
8 – –	8.96%
5 – –	7.64%
1 – 16 –	3.95%
1 – 2 – 15	3.83%
5 – 8 –	2.97%
16 – –	2.85%

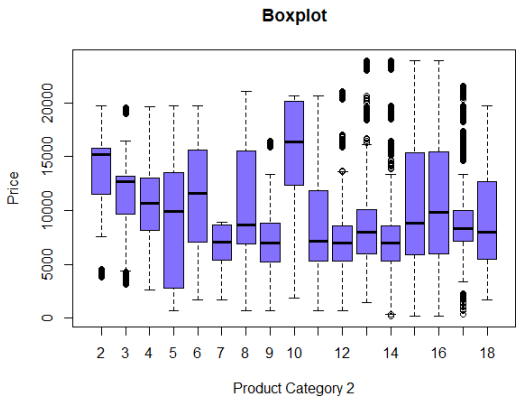


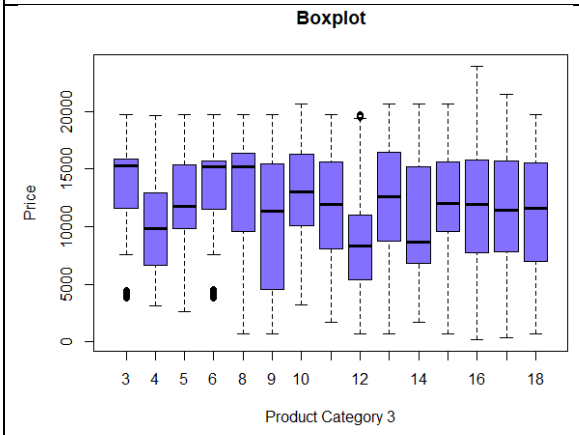
Basket analysis on the basis of Age Group		
Ages	Number of Buyers by Age Group	Purchase Distribution by Age Group
0-17		
18-25		
26-35		
36-45		

46-50		
51-55		
55+		
Insights	• Overall, all age groups preffered Product Category 1: item 8 and item 5 the most. • All age groups bought more of Product Category 1:item 5 and Product Category 2: item 15, together.	
Suggestion	• The shop keepers should place Product Category 1: item 5 and item 6 and start targeting age group of 26-35 for more profit. • To increase sales the shop keepers must keep the item together as per product baskets preferred by individuals age group. • The number of Product Category 1:item 1 and Product Category 2:item 16 were sold less even though they prominent contributers in sales price. Hence, shopkeepers should focus on selling Product Category 1:item 1 and Product Category 2:item 16 even more. Moreover, they should target the age group 0-17 and 18-25 as they tend to buy these items more.	

<table>
<tr><th colspan="3">Basket analysis on the basis of Gender</th></tr>
<tr><th>Gender</th><th>Number of Buyers by Gender</th><th>Purchase Distribution by Gender</th></tr>
<tr><td>Male</td><td> 
</td><td> 
</td></tr>
<tr><td>Female</td><td> 
</td><td> 
</td></tr>
<tr><td>Insights</td><td colspan="2">• Overall, males and females both prefer Product Category 1: item 8 and item 5 the most.
• Males prefer to buy Product Category 1: item 1 and Product Category 2: item 16, together, whereas, Females prefer to buy Prodcut Category 1:item 5 and Product Category 2:item 14, together.</td></tr>
<tr><td>Suggestion</td><td colspan="2">• The shop keepers should place Product Category 1: item 1 and Product Category 2: item 16, together and Prodcut Category 1:item 5 and Product Category 2:item 14, together on a shelf for increase in sales.</td></tr>
</table>







Insights:

- Product Category 1:item 10 is the most expensive item amongst all the product categories.
- Items of Product Category 1 are having variations in their prices from very high to very low, whereas Product Category 3 seems to be evenly priced.

Suggestion:

- Shopkeepers should focus more on selling Product Category 1:item 10,6,7,15; Product Category 2:item 10 and most of the items from Product Category 3 to gain huge profit.

Regression Model

- Model: Purchase = Gender + Age + City Category + Occupation
- This model was achieved using the All Subset method for variables.
- Variation explained by model (Adjusted R- Squared): 0.93%.
- Interpretation: 0.93% of the variation in Purchase is due to variables such as Gender, Age, City Category, Occupation.
- The Mean Square Prediction Error (MSPE[2]) is high for our model, hence we are not confident in our model.
- The predictors of purchase i.e. Gender, age, City Category and Occupation do not hold equal weightage. Gender explains maximum variation out of the given predictors i.e. 0.3% variation in purchased is explained by gender alone.

Limitations

- Product Category which is an important variable in estimating Purchase count of items could not be taken into the model as it had a lot of levels and the outcome would have not given proper insights as to how it was affecting the Purchasing behaviour of consumers.
- There was no column for discounts which plays a major role in explaining increased Purchasing by consumers.

[2] In statistics the mean squared prediction error of a smoothing or curve fitting procedure is the expected value of the squared difference between the fitted values implied by the predictive function {\displaystyle {\widehat {g}}} and the values of the (unobservable) function g.