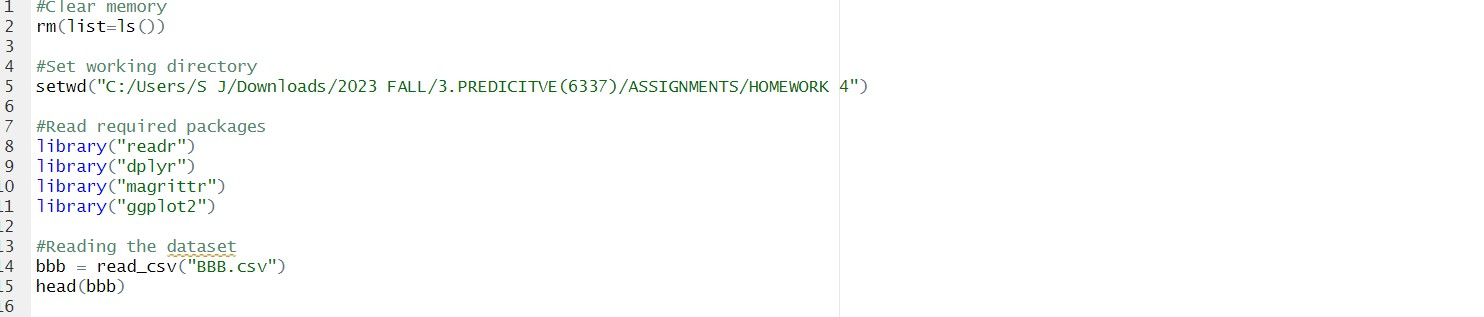
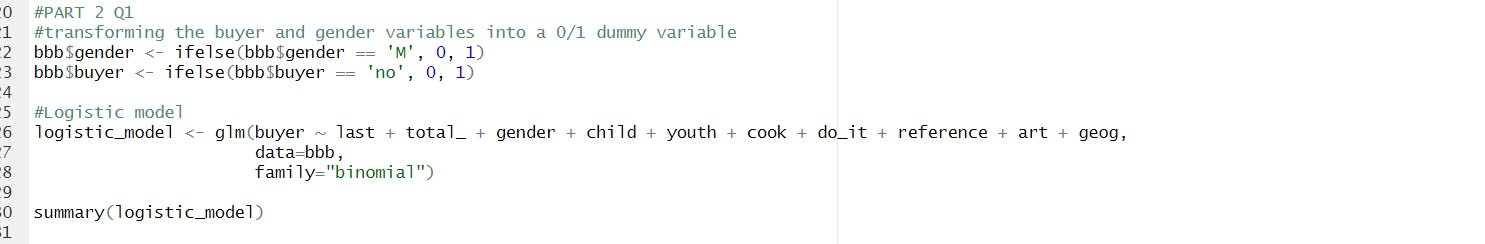
**Logistic Regression**

**1)**

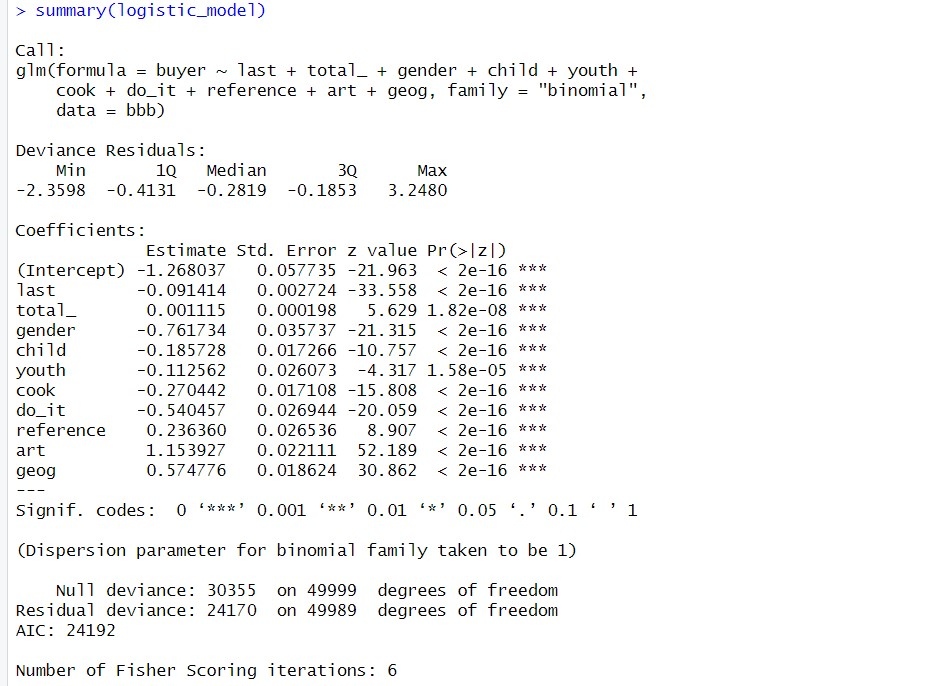
Importing libraries and reading the dataset



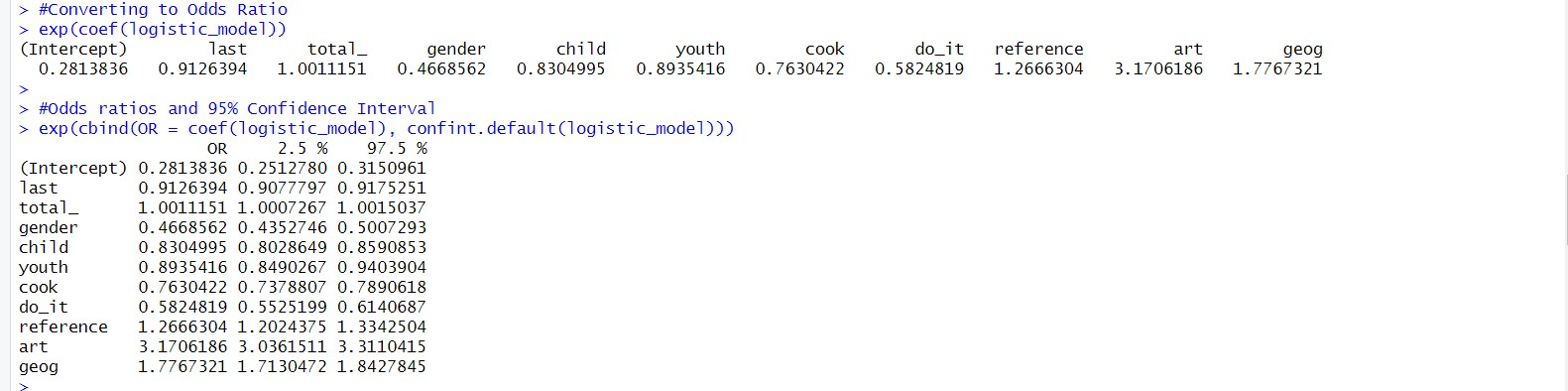
Transforming buyer and gender variables to dummy variables and running the logistic regression model



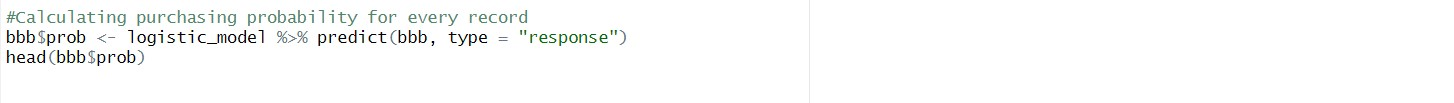
Output for logistic regression (summary table)

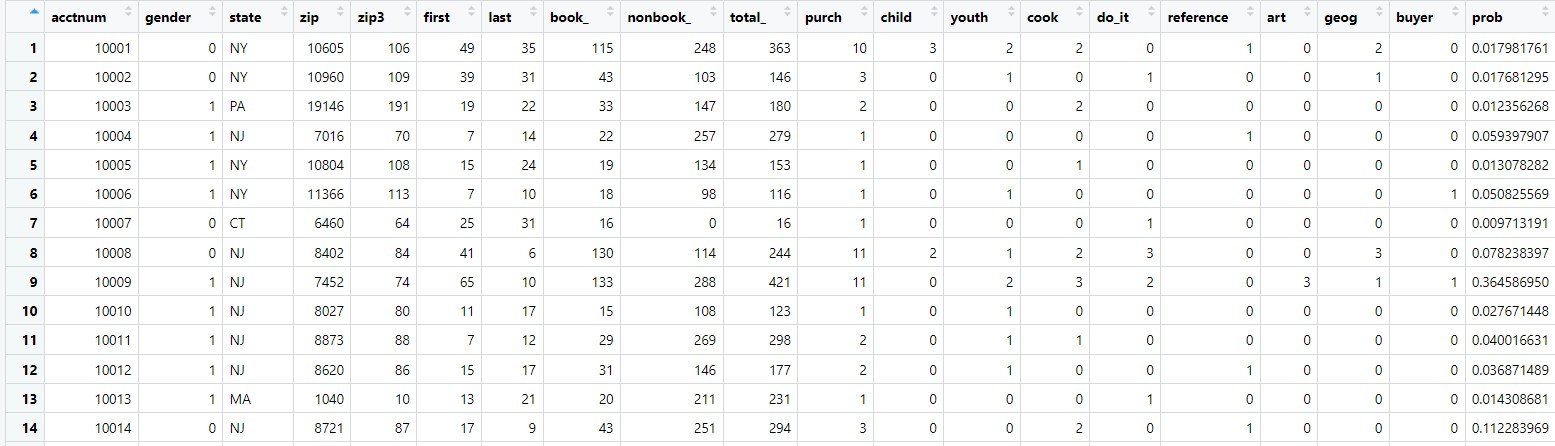


Output for exp



New variable (prob) that contains the predicted probability of purchase for each consumer.





**2)**

From the summary of logistic regression model, we can see that the p value for all the variables are very small therefore all the variables are statistically significant and will contribute in the final decision of buying or not. For the odds ratio as all the variables are statistically significant so variables with odds ratio greater than 1 can be identified as variables having the highest positive impact on buying, while variables with odds ratio less than 1 can be identified as variables having negative impact on buying.

Variables last, art, geog and reference have positive impact on buying probability and are more economically important.

Variables gender, cook, and do\_it have negative impact on buying probability.

Interpretation of each variable’s odd ratio is as follows:

last: As the odds-ratio is less than 1, we can say that for 1 unit increase in last variable there is a decrease in the odds of being a buyer of the book. One unit increase in month since most recent purchase decreases the odds of customer of buying “The Art History of Florence” by a 0.91.

Total\_ : Odds ratio for this is approximately equal to 1. There is no significant relationship between the total dollars spent and probability of buying.

Gender: Odds ratio is 0.46. It is less than 1. We can say that females are less likely to be a buyer (we have taken M=0 and F=1) and increasing female population will reduce the odds of book sale. One unit increase in month since most recent purchase decreases the odds of customer of buying “The Art History of Florence” by a 0.46.

child: Odds ratio is 0.83 (less than 1). So, increase in the number of children books purchase would reduce the odds of buying the book. One unit increase in number of children book purchased would reduce the odds of customer of buying “The Art History of Florence” by a 0.83.

youth: Odds ratio is 0.89 (less than 1). So, increase in the number of youth books purchase would reduce the odds of buying the book. One unit increase in number of youth books purchased would reduce the odds of customer of buying “The Art History of Florence” by a 0.89.

cook: Odds ratio is 0.76 (less than 1). So, increase in the number of cookbooks purchase would reduce the odds of buying the book. One unit increase in number of cook books purchased would reduce the odds of customer of buying “The Art History of Florence” by a 0.76.

do\_it: Odds ratio is 0.58 (less than 1). So, increase in the number of do-it-yourself books purchase would reduce the odds of buying the book. One unit increase in number of do-it-yourself books purchased would reduce the odds of customer of buying “The Art History of Florence” by a 0.58.

reference: Odds ratio is 1.26. We can say that for buyers of reference books have a significant positive relationship with buying the book. One unit increase in number of reference books purchased would increase the odds of customer of buying “The Art History of Florence” by a 1.26.

art: Odds ratio is 3.17. We can say that for buyers of art books have a significant positive relationship with buying the book. One unit increase in number of art books purchased would improve the odds of customer of buying “The Art History of Florence” by a 3.17.

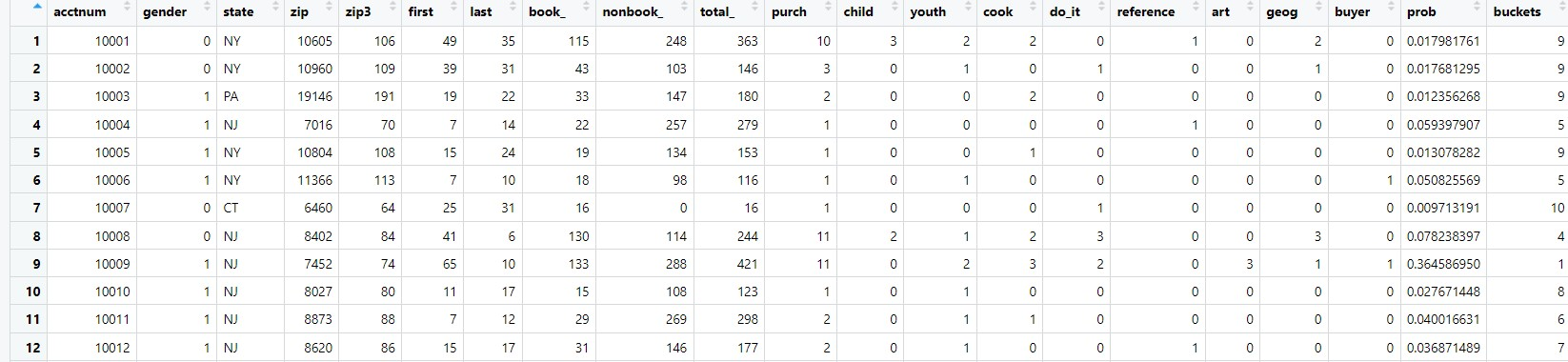
geog: Odds ratio is 1.77. We can say that for buyers of geography books have a significant positive relationship with buying the book. One unit increase in number of geography books purchased would reduce the odds of customer of buying “The Art History of Florence” by a 1.77

**Part II: Decile Analysis of Logistic Regression Results**

**1)**

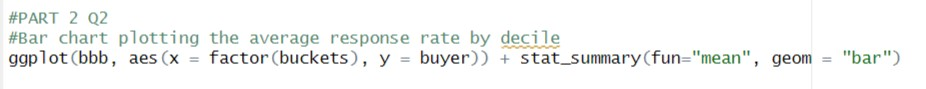
Assigning each customer to a decile based on his or her predicted probability of purchase with ‘bucket 1’ being the highest average purchase probability.

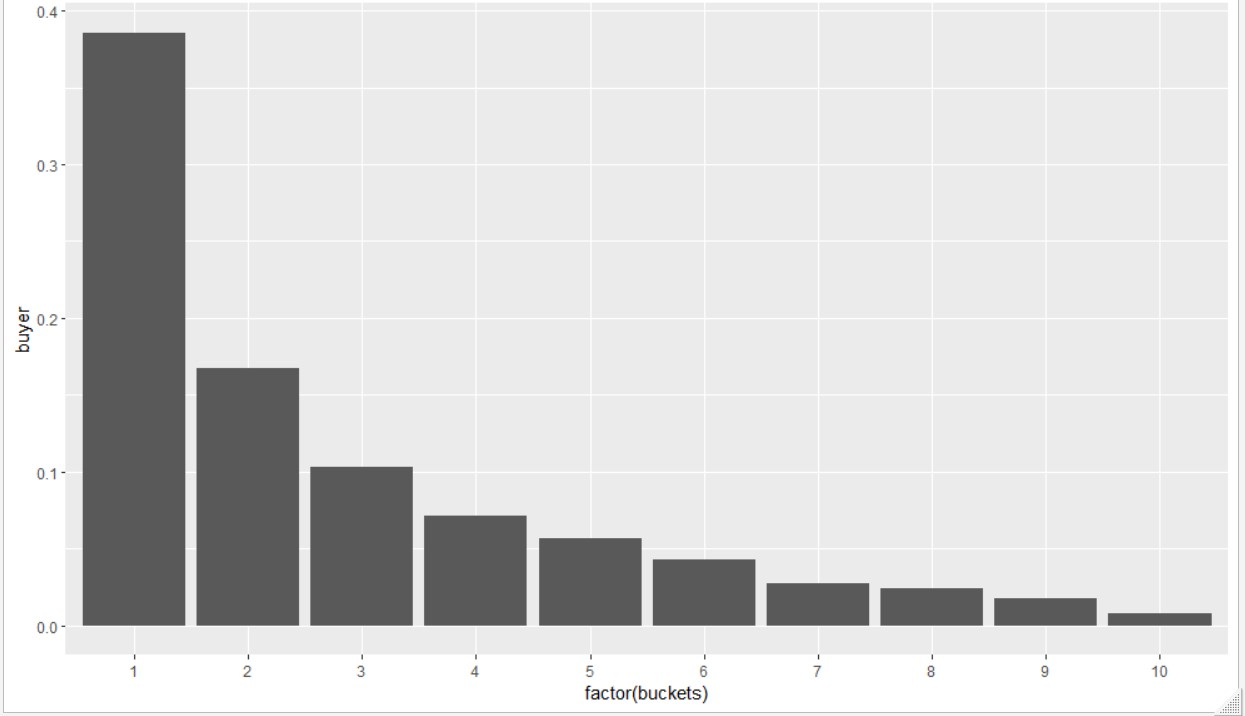




**2)**

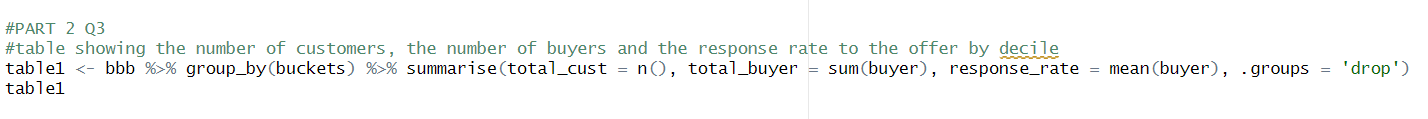
Bar chart plotting the average response rate by decile.

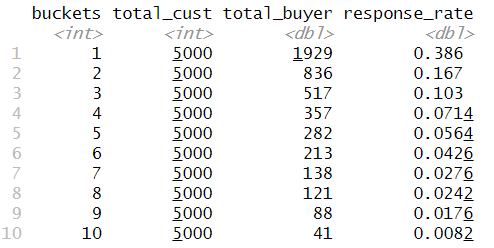




**3)**

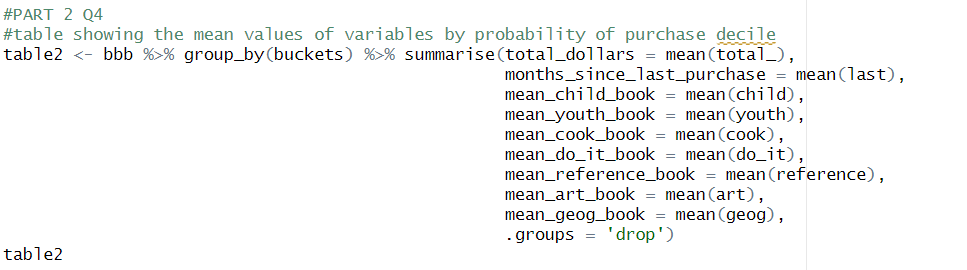
Table showing the number of customers, the number of buyers and the response rate to the offer by decile.

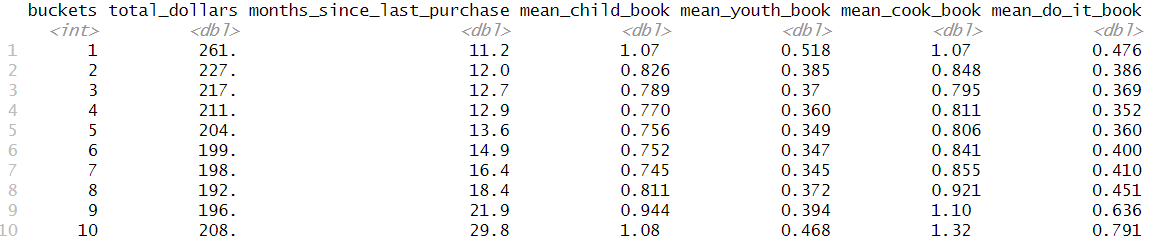




**4)**

Table showing the mean values of the variables by probability of purchase decile.

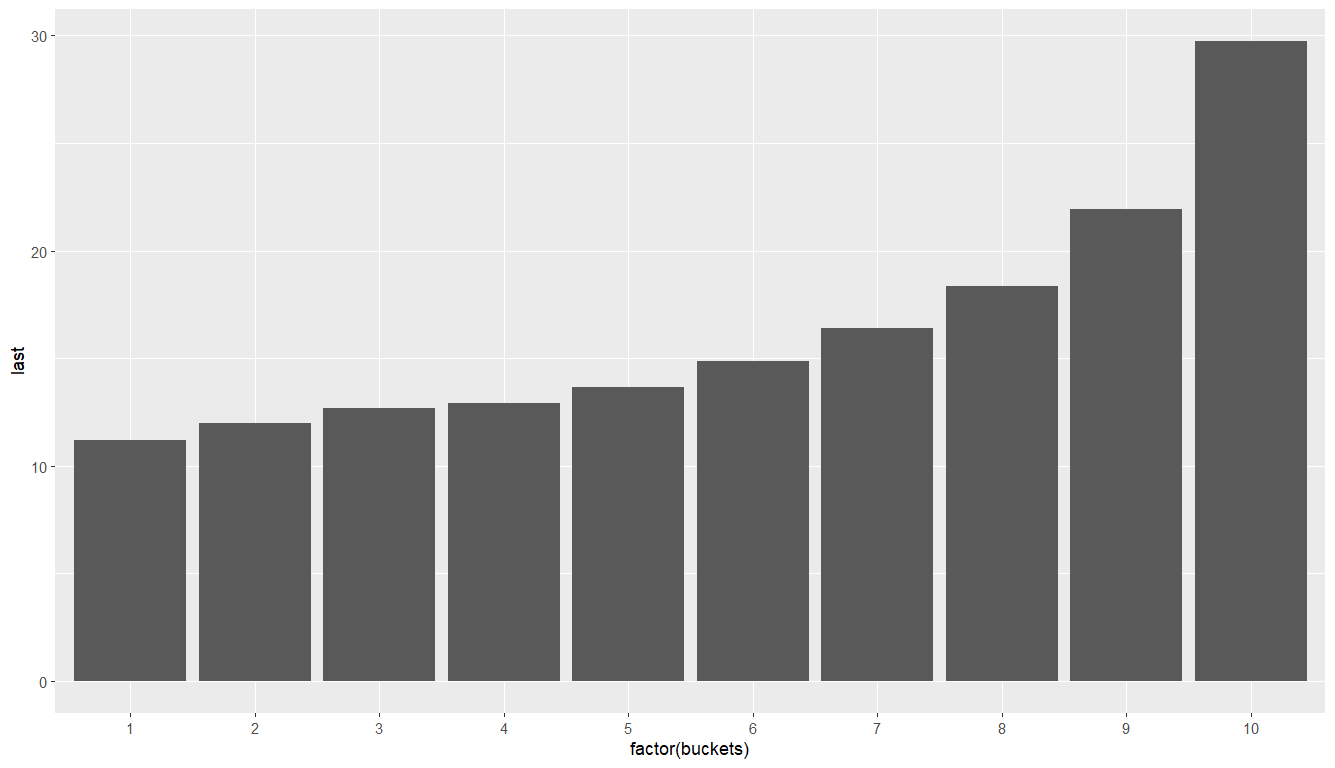




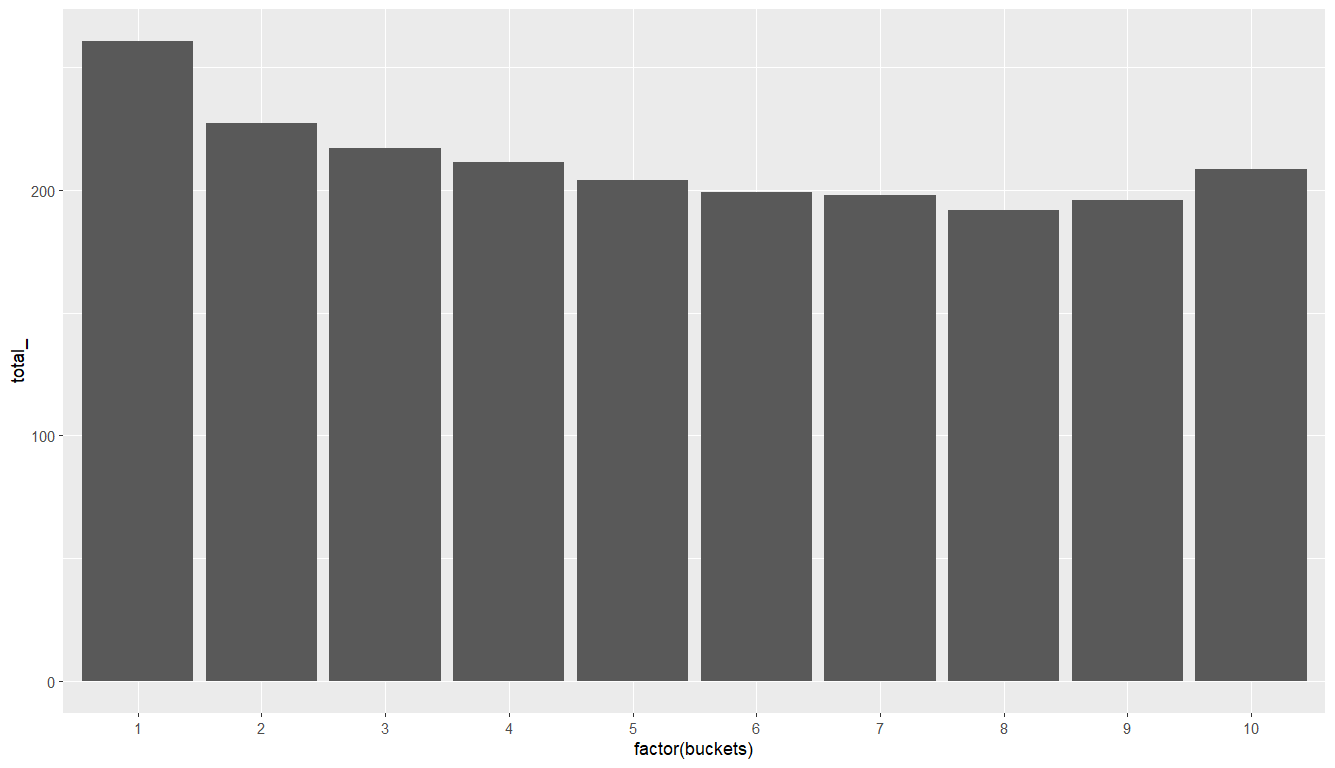
**5)**

Decile analysis results of different varaibles are -

last - As the number of months since the last purchase increases, the likelihood of making a purchase decrease across all decile buckets. This observation aligns with the findings from the logistic regression analysis.



total\_ , child, youth, cook: for total\_ within decile buckets 1-5, the probability of making a purchase decreases as the total dollars spent increases. However, the trend takes a different direction within buckets 6-10, where the probability of buying seems to rise as total dollars spent increases. This observation does not align with the findings from the logistic regression analysis. Similar pattern can be observed in child, youth and cook.



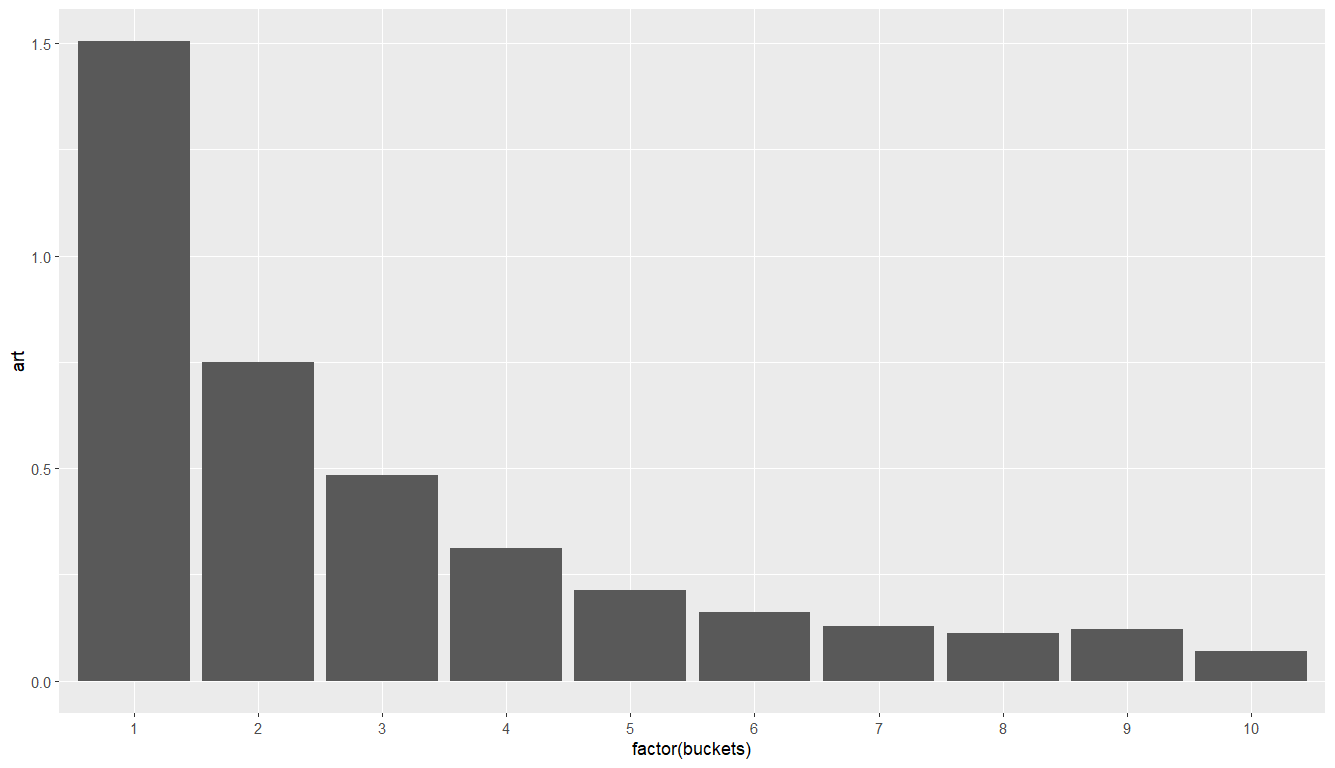


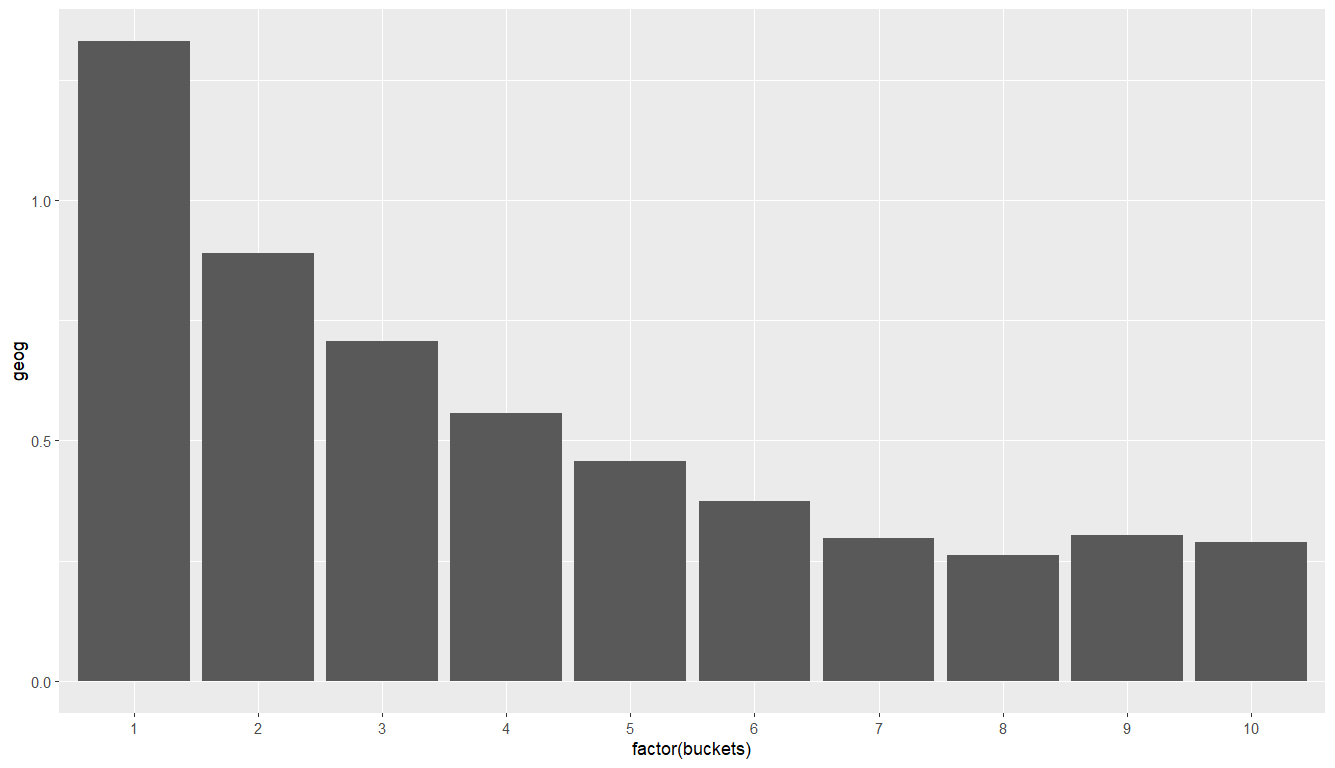




Reference, art, geog: These three features exhibit a direct correlation with the likelihood of making a purchase, closely resembling the output from logistic regression for the top 8 buckets. They consistently follow a similar pattern from buckets 1 to 8, but their behavior diverges in the case of buckets 9 and 10.

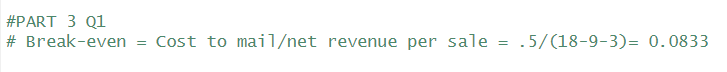






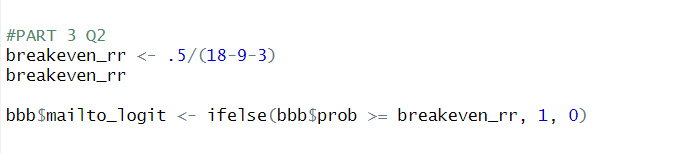
**Part III: Profitability Analysis**

**1)**

****

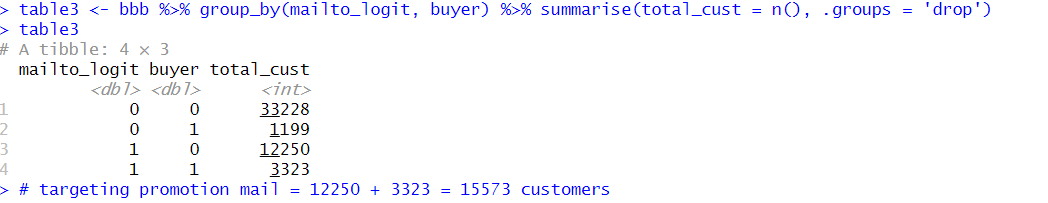
**2)**

creating a new variable (mailto\_logit)

****

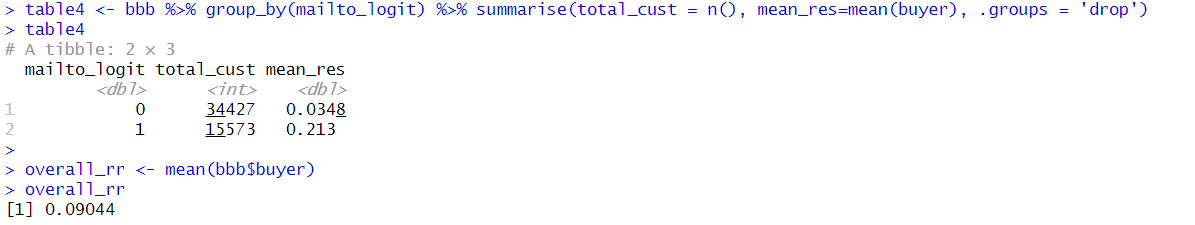
**3)**

Customers that should have received the targeting promotion mail based on the breakeven response rate (i.e., the number of mailto\_logit ==1 in the data) as per the below command.

****

Therefore, customers who should have received the targeting promotion mail – 12250 + 3323 = 15573

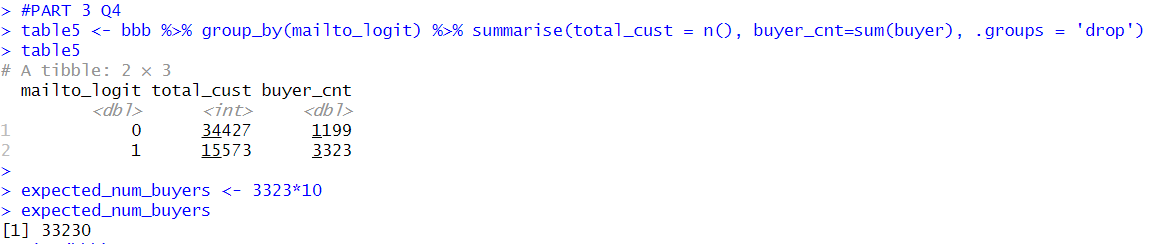
Response rate of people who have been targeted and overall



We can see that targeted response rate is 0.21 and it is significantly higher than overall response rate (by 0.12 or 12%).

**4)**

Expected number of buyers if we do targeted mailing based on the breakeven response rate.

****

Breakeven response rate = 33230