Q1**.** Based on the transactions, which product category dominates in terms of $ amount?

(Hint: You will need to merge Transactions and Product data sets and then look at Product category)

#Answer:

Product\_Category revenue no\_items

<fctr> <int> <int>

1 Entertainment 34986564 11636

2 MP3Players 23663229 227411

3 Hardware 11337175 85230

4 GameConsoles 6121417 6983

5 HiFi 5507965 11251

6 Telephony 3683800 12600

7 Software 1392347 20176

#So we can infer from the above observation, the product category **Entertainment** is dominating

Q2. Perform a suitable age grouping and find out contribution of each of the age group in terms of $ amount spent.

(Hint: A merge between Customer and Transaction table will be required)

#ANSWER

**Age** **Amount\_Spent**

1 0-41.36 8366567

2 41.36-46.35 9864597

3 46.35-49.83 9841517

4 49.83-52.63 9126069

5 52.63-55.02 8101500

6 55.02-57.57 8208014

7 57.57-60.46 8293880

8 60.46-64.36 8280924

9 64.36-70.20 8491086

10 70.20-115.08 8118343

Q3. Find the response rate to the campaign. Also identify the age group of customers where response rate is high. Is there a consistent trend.

(Hint: Add age information from customer file to the campaign file, compute response rate by age. This can be done by binning age either in deciles or quartiles and seeing if there is a consistent trend)

#ANSWER:

age\_group response\_rate

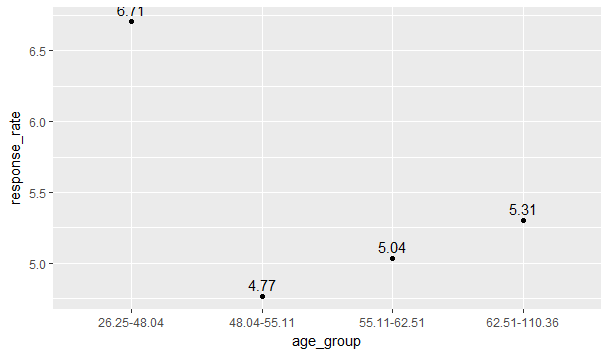
1 26.25-48.04 6.711409

2 48.04-55.11 4.768301

3 55.11-62.51 5.036938

4 62.51-110.36 5.305574

From the above table we can infer that for the age group **26.25-48.04** response rate is high.



It can be observed from the above graph that response rate is the highest for age group 26-48 and after that increase in age group leads to increase in response rate.

Q4. Repeat the analysis above with “Tenure” of customer. (Tenure will be defined as the time period between the Date of Registration and 31/12/2002)

#ANSWER:

tenure\_period response\_rate

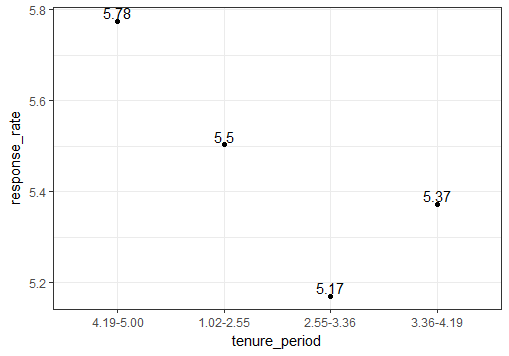
1 1.02-2.55 5.503356

2 2.55-3.36 5.171256

3 3.36-4.19 5.372733

4 4.19-5.00 5.775688

It can be observed that tenure between 4 and 5 is active in giving response.



It can be observed from the above chart tere is no specific trendz which can be inferred from the relationship of tenure and response rate.

Q5. Create a cross tab of response rate between Age and Tenure of customers. Do you observe anything?

Tenure\_Quantile

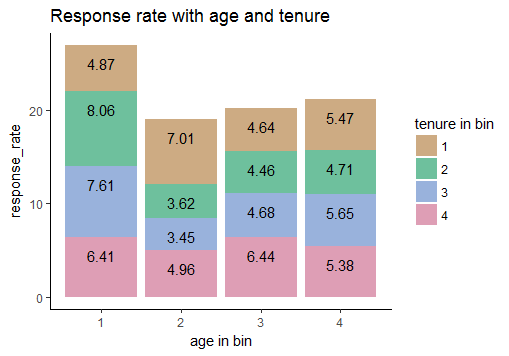
age\_quantile 1 2 3 4

1 4.871795 8.055556 7.611549 6.406685

2 7.008086 3.617571 3.448276 4.960836

3 4.637681 4.461942 4.679803 6.442577

4 4 5.468750 4.709141 5.649718 5.384615



From the above chart it can be observed that

1)The age group of 26.27 to 48.06 i.e quantil1 is more active in giving response as compared to other groups

2)there is no such trendz observed in the above chart between age in bin and response rate with respect to tenure in bin. It's randomly distributed.

Q6. Which mode of payment is most popular? Is mode of payment affected by the time of transaction?

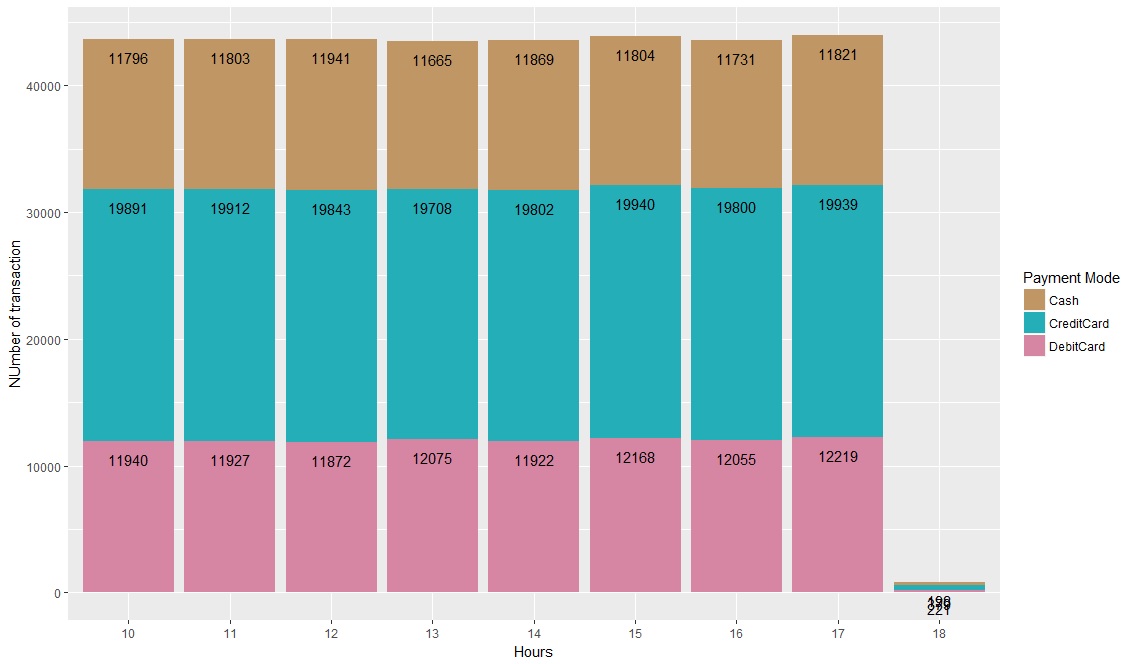
(Hint: Extract hour information from timestamp column by using appropriate date conversion function, based on the hour of the day extracted, you can do an appropriate classification and then look at the cross tab between payment mode and time of day)

#ANSWER:

Cash CreditCard DebitCard

Total\_Transaction 94628 159214 96399

It can be observed from the above table credit card is most popular



It can be observed that there is no such variation in payment method with respect to time,in all hours of the day credit card is most popular payment method.

Q7. Do you think, based on the data, that age and gender has any impact on $ amount spent?

(Hint: You’ll need to merge customer and transaction tables appropriately and then do an age classification, post that you can create a cross tab between gender and age to arrive at an opinion)

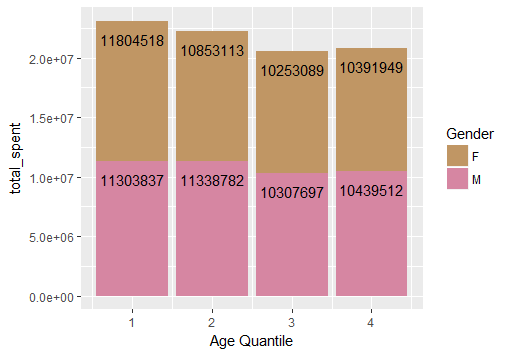
quantile F M

1 11804518 11303837

2 10853113 11338782

3 10253089 10307697

4 10391949 10439512



from this plot we can infer except quantile1 i.e age group of people between 25 and 48,in all other

age group from 48-115 ,men are spending more than women

Q8. Produce a histogram for “tenure of a customer” separately for male and female customers.

