This is a document containing the result of than business intelligence analysis of company AXV, a telecommunication company in country z. The focus of this business analysis is to explains what is responsible for customer transferring from our network to our competition, our clients contract types and what is responsible for them having the contract type they are on, how we can get them to buy into our services, retain them, and also reduce the amount of contract terminations and arrears from our clients.

Two datasets are provided with one tagged DDM\_22\_19\_customer with 13 columns and 7513 rows indicating the data for 7513 clients and 13 data points per client. The second data set tagged DDM\_22\_19\_score as 2 columns and 2078 meaning the joining of the two data sets will introduce missing data as some customers were not catered for in in the DDM\_22\_19\_score dataset. This analysis presents the company’s current standing with some visualization, tables, explanation of what they entail and some suggestions from the analytics angle to make the company more profitable and increase customer retention.

We got 7807 unique missing data which accounts for a ratio of 0.07993324 which means 7.99% of the data is missing and this means we have to do something about the missing dat points as deleting the rows in question might reduce our dataset to something we can’t infer good knowledge from. On further analysis it was discovered that click carries the highest percentage of the missing data with 7498 unique missing data point which accounts for 0.9980035 ratio of our number of rows and a removal or click from our dataset reduced our error to 0.009427393 which means we now have 0.943% of our data missing and so we can remove the row containing missing data.

The summary statistics of the dataset is as presented below

summary(DDM22\_19\_customer)

gender age citizenship tenure

Length:7210 Min. : 13.00 Length:7210 Min. : 39.00

Class :character 1st Qu.: 31.00 Class :character 1st Qu.: 79.00

Mode :character Median : 41.00 Mode :character Median : 89.00

Mean : 42.59 Mean : 89.36

3rd Qu.: 52.00 3rd Qu.:100.00

Max. :101.00 Max. :146.00

contract\_type data\_allowance payment\_method monthly\_charges

Length:7210 Length:7210 Length:7210 Min. : 10.00

Class :character Class :character Class :character 1st Qu.: 50.00

Mode :character Mode :character Mode :character Median : 50.00

Mean : 54.78

3rd Qu.: 62.00

Max. :109.00

payment\_arrears contract\_terminated number\_transferred cid

Min. :0.0000 Length:7210 Length:7210 Min. : 1

1st Qu.:0.0000 Class :character Class :character 1st Qu.:1876

Median :0.0000 Mode :character Mode :character Median :3770

Mean :0.1589 Mean :3760

3rd Qu.:0.0000 3rd Qu.:5639

Max. :4.0000 Max. :7513

The table above show the data type and length for the categorical variables while it shows Min as minimum, 1st Q as first quantile, median, mean, 3rd Qu as third quantile and max for the continuous variable (continuous variable are variables that are numbers).

gender age citizenship tenure

3 84 6 103

contract\_type data\_allowance payment\_method monthly\_charges

3 5 3 89

payment\_arrears contract\_terminated number\_transferred cid

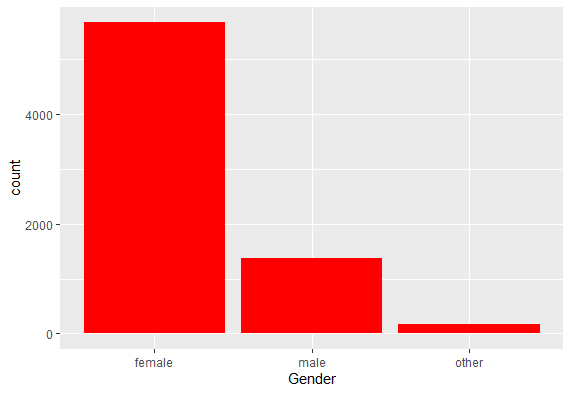
5 3 2 7210

The table above shows the number of unique data each data point as. After this we joint the score data and the customer data on the customer id column.

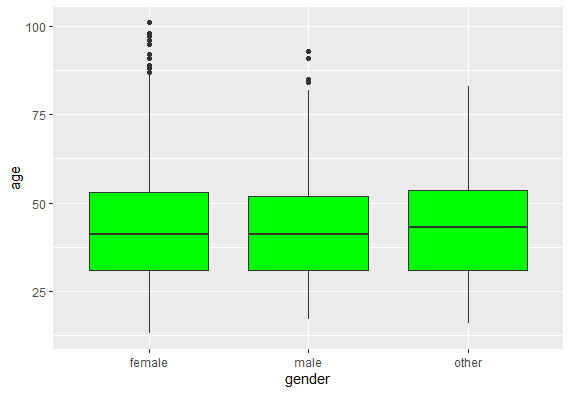
data <- left\_join(DDM22\_19\_customer, DDM22\_19\_score)

Joining, by = "cid"

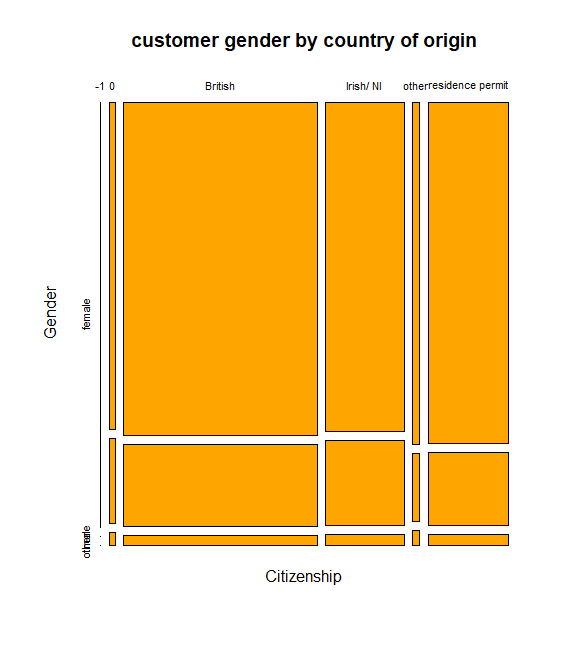
Number of distinct customer id we have is 7210 which is less than 7807 number of rows we have which indicates we have some repeated data point where we have a customer data repeated but as this will not necessarily affect our business analysis we can neglect it at the moment.

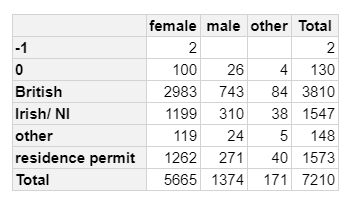


This is a plot of the distribution of the gender and it is indicative that over 50% of our clients are female.

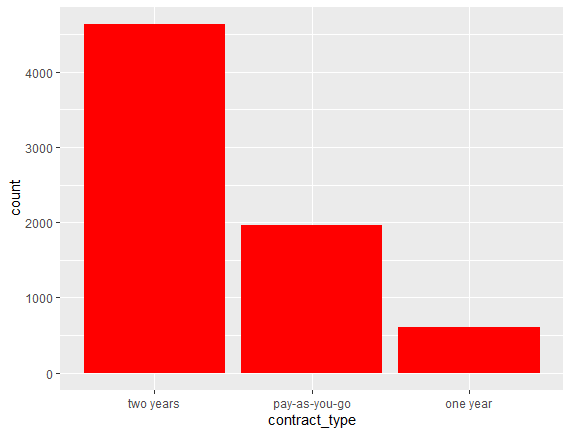


The plot above is a box plot of the age of the customer by their genders and we can see some outliers in the male and female category. Checking the pivot table generated with the code on the code.r script shoes the aga data is skewed and should be dropped so we dropped the age, cid and the click columns.

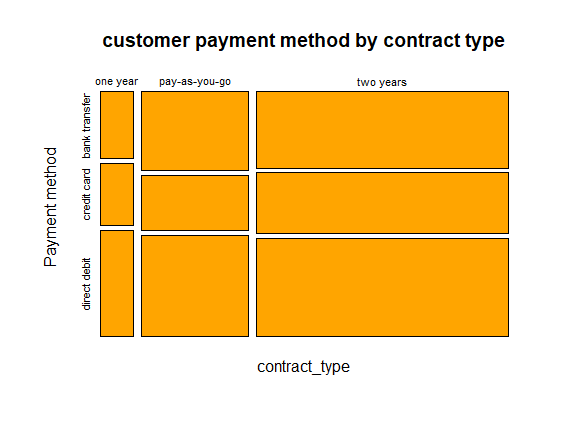


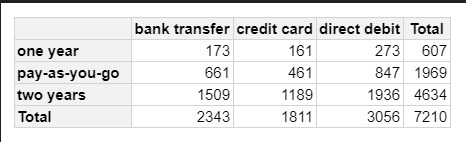


The table and plot above shows the distribution of the citizenship by gender and its indicative that most of our clients are British or have residence permit.

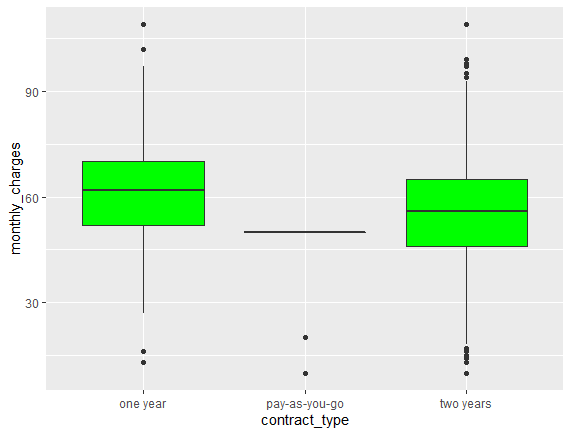


The plot above shows the distribution of the contract types of our clients and it is obvious that most of our clients are on a two year contract type which is a good sign of being with us for a long time and this is indicative of a higher customer lifetime value (CLV).

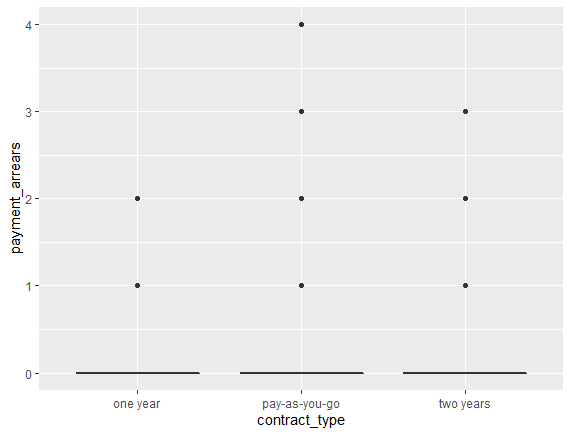


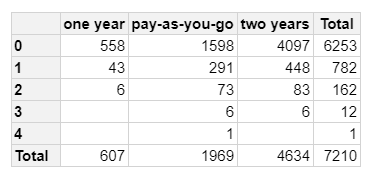


The plot above shows the distribution od our contract types against payment method and it shows a variation in the number of people using which payment either per category of contract type, but the general concept is that direct payment is higher than other forms of payment.

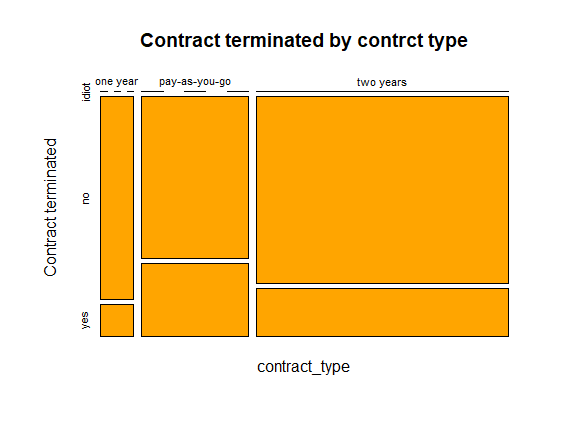


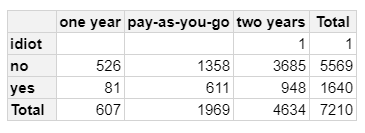
This plot above shows the box plot of monthly charges distributed by contract type with ay as you go almost inexistent the two year contract type seems to have outliers at both ends of the box plot and so is the one year contract type too.



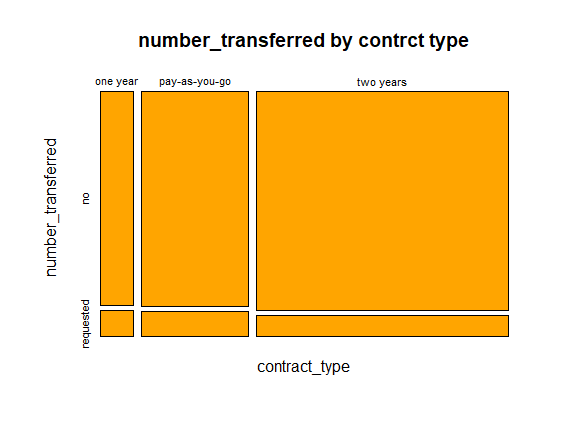


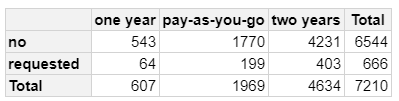
This where the data gets interesting as the lot and table above shows the distribution of payment arrears as against contract type and practically most o the customer have most to no payment arrears which is indicative of the fact that the accounting department has a good strategy in place to get the buyers of our services to pay for the services we are rendering to them.



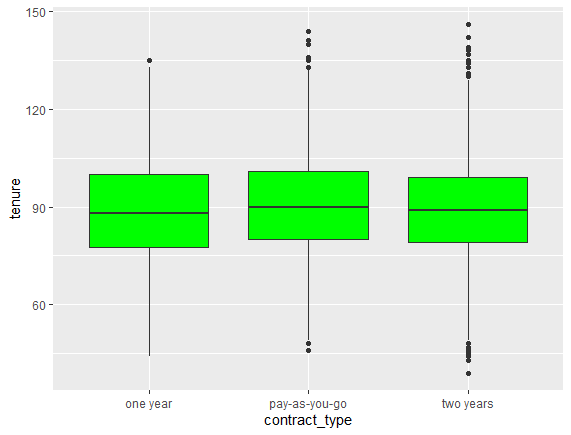


The table above shows the distribution of contracts terminated and contract types and there seems to be a misclassification as we have a single idiot but that said we have a higher percentage of our clients not terminating their contract but the ratio terminating their contract is quite alarming as around1 of every 5 customers seems to be terminating their contract. And this must be due to probably poor service delivery or better service from our comatitions.

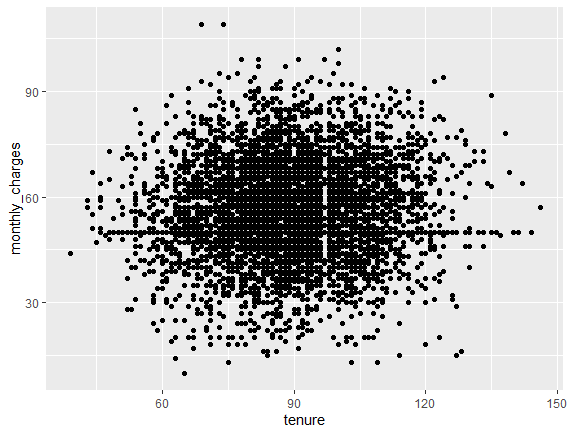


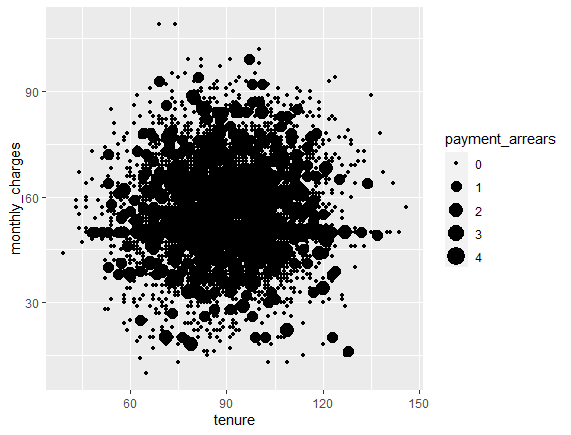


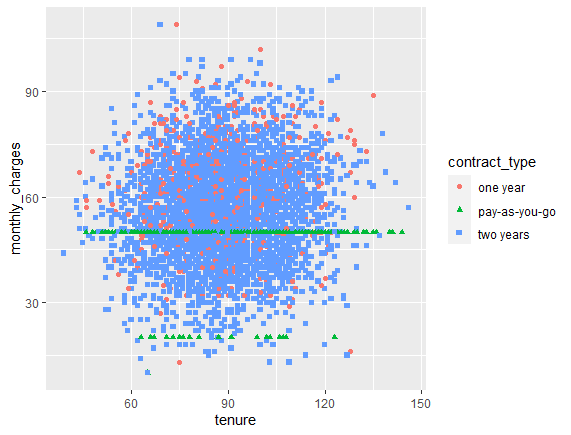
The table and plot above shows the distribution of number of transfers against contract types and from all indication 1 of every 10 clients we presently have as requested for a transfer this this mixed with the fact that 1 of every 5 clients as terminated their contract shows the fact that most of our clients are moving to our competition which can be due to a whole lot of factor with major being checking our service delivery and understand what is wrong or check our competition for what they are doing the=at we can put in place.



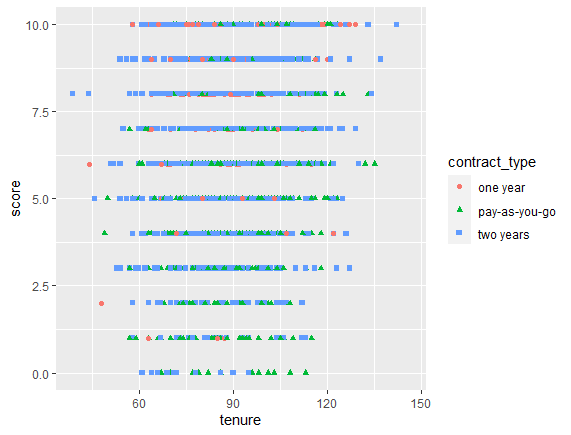
The plot above shows the box plot of the tenure as distributed into the various contract types with the two years plan having the highest outliers followed by the pay as you go.



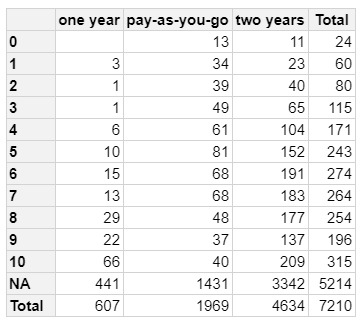


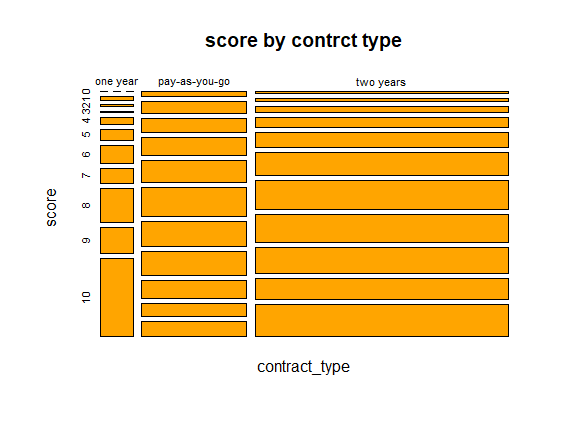


The plot above shows the plot of monthly charges against tenure and it’s indicative of the fact that monthly charges and tenure so not correlate wuth each other.



The graph above shoe a plot of score against tenure score acd hued with contract type and it is indicative that the number of clients increase ast the score increase which is expected cos a deline might indicate clients jumping ships, i.e., moving to other service providers.





The table above shows the distribution of clients into scores aaginst contract type and it solidify what was stated in the last plot asides the fact theat in also increases as yu move from pay as you go contract type to one year conract type to two years contract type. And most of the clients have not yet been asigned a score.