Data Exploration Assignment

### **Problem Statement:**

Consider yourself to be Sam, who is a Data Scientist at ‘Verizon’. In the past 6 months customers of your company have been churning out to other networks. So, you as a data scientist have to understand what are the factors leading to the customers churning out and what can you do to reduce the churn rate.

Sam starts off with data exploration. So, Sam wants to understand the structure of the data first

## Questions on Data Extraction:

1. Extract these individual columns with the ‘$’ symbol:
   1. Extract ‘InternetService’ column and store it in ‘customer\_internet\_service’
   2. Extract ‘PaperlessBilling’ column and store it in ‘customer\_Paperless\_Billing’
   3. Extract ‘StreamingTV’ column and store it in ‘customer\_Streaming\_TV’
2. Extract the 3rd, 6th and 9th columns from the ‘customer\_churn’ data.frame & store it in ‘customer\_random\_columns’
3. Extract all the columns from column number-10 to column-number 20 and store the result in ‘customer\_10\_20’
4. Extract only these row numbers: 65, 765, 3726 & 7000 and store the result in ‘customer\_random\_rows’
5. Extract all the rows starting from row number-655 to row number-6550 & store the result in ‘customer\_continuous\_rows’
6. Extract row numbers- 10, 100 & 1000 & column numbers- 5, 10, 15 & store the result in ‘customer\_random\_data’

Now, Sam decides to implement some flow control statements to explore the data

## Questions on Flow Control Statements:

1. Check if the value in the 6th cell of ‘PaymentMethod’ column is ‘Electronic check’. If yes, print “Yes, the payment method is Electronic check”
2. Check the value present in 12th cell of ‘Contract’ column.

If it’s ‘month-to-month’, print ‘The contract is on a month to month basis’

If it’s ‘One year’, print ‘The contract is on a yearly basis’

If it’s ‘Two year’, print ‘The contract is on a two-year basis’

1. Use switch to check the gender in 6th cell of ‘gender’ column.

If it’s ‘Male’, give a discount of 20% in ‘MonthlyCharges’

If it’s ‘Female’, give a discount of 50% in ‘MonthlyCharges’

1. Use for loop to get the count of customers whose ‘InternetService’ is ‘DSL’
2. Use while to find the number of customers whose tenure is exactly ‘2’ months

After this, Sam would be applying some inbuilt functions on the data to get some insights

## Questions on Inbuilt Functions:

1. Do these operations with the head() function:
   1. Get the first 4 records from ‘PhoneService’ column
   2. Get the first 8 records from ‘Contract’ column
2. Do these operations with the tail() function:
   1. Get the last record of ‘TotalCharges’ column
   2. Get the last 5 records of ‘tenure’ column
3. Find the average, minimum, maximum & range from the ‘tenure’ column
4. Get 10 random values from the ‘TotalCharges’ column using sample()
5. Find the count of different levels in ‘PaymentMethod’ & ‘Contract’ columns using table()