Analysis of Telco Churn Dataset (from Kaggle)

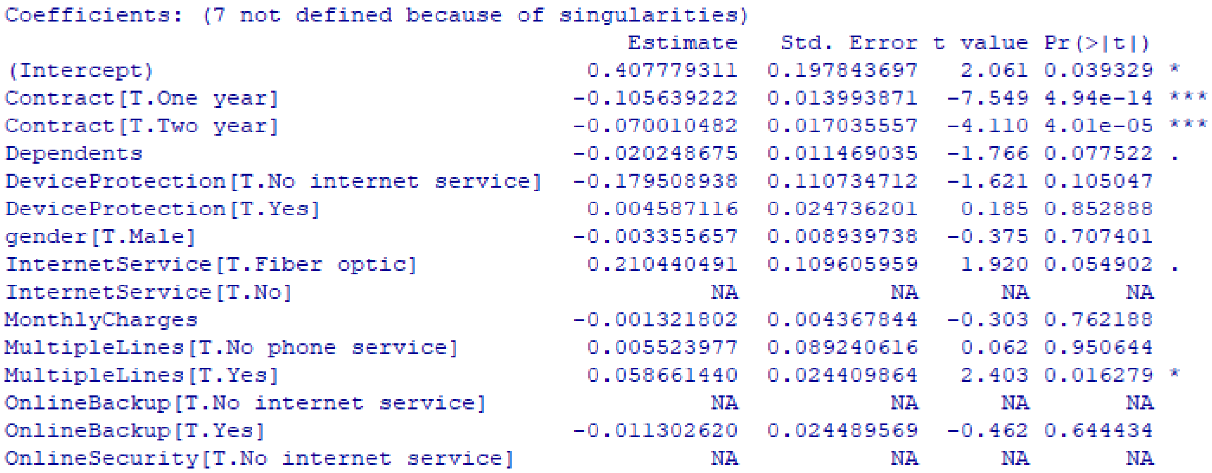
About the dataset:

The Dataset tells about whether or not a customer stays with the Telecom service provider. The factors affecting the Churn are:

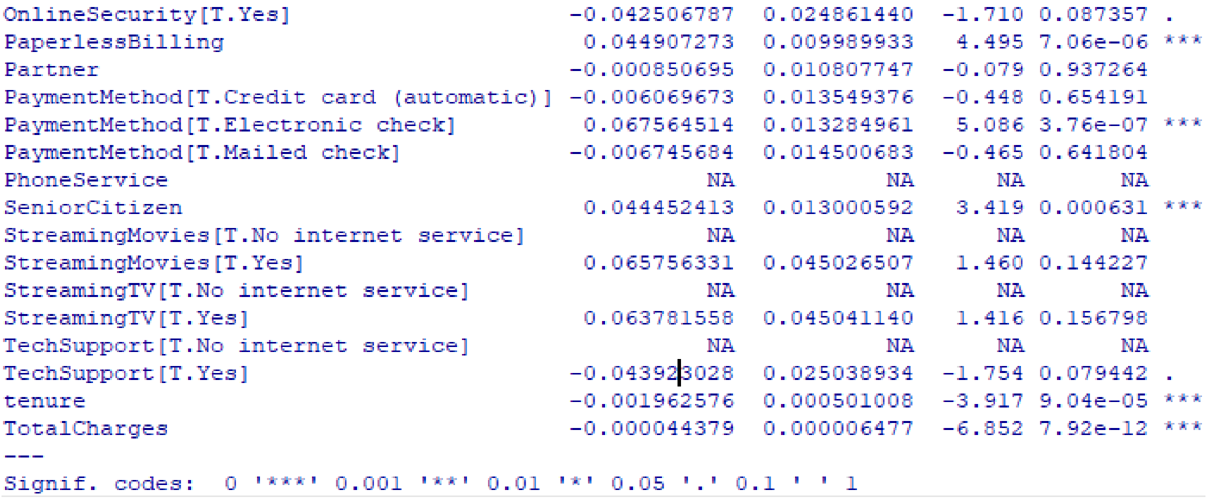
- Gender – male, female
- Senior citizen – yes / no
- Partner – yes / no
- Dependents – yes/no
- Tenure – continuous variable
- Phone Service – yes/no
- Multiple Lines – yes/no/No phone service
- Internet Service – No, DSL, Fiber Optic
- Online security – yes/no/No internet service
- online Backup – Yes/no/No internet service
- Device protection - Yes/no/No internet service
- Tech support - Yes/no/No internet service
- Streaming TV - Yes/no/No internet service
- Streaming Movies - Yes/no/No internet service
- Contract – Month-Month, One year, Two year
- Paperless billing – yes/no
- Monthly charges – Continuous variable
- Total Charges - continuous variable

Data Cleaning: Conversion of alphabetic values like yes/no to 1 and 0 is done. Here, the dominant factor, Yes is coded 1, and the non-dominant factor, No, is coded 0.

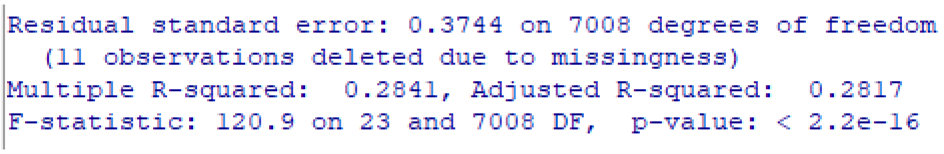
From the above data, when we do Linear Model analysis in R commander, we get the following result:



```
Coefficients: (7 not defined because of singularities)
                                            Estimate  Std. Error t value Pr(>|t|)    
(Intercept)                              0.407779311 0.197843697   2.061 0.039329 *  
Contract[T.One year]                    -0.105639222 0.013993871  -7.549 4.94e-14 ***
Contract[T.Two year]                    -0.070010482 0.017035557  -4.110 4.01e-05 ***
Dependents                              -0.020248675 0.011469035  -1.766 0.077522 .  
DeviceProtection[T.No internet service] -0.179508938 0.110734712  -1.621 0.105047    
DeviceProtection[T.Yes]                  0.004587116 0.024736201   0.185 0.852888    
gender[T.Male]                          -0.003355657 0.008939738  -0.375 0.707401    
InternetService[T.Fiber optic]           0.210440491 0.109605959   1.920 0.054902 .  
InternetService[T.No]                             NA          NA      NA       NA    
MonthlyCharges                          -0.001321802 0.004367844  -0.303 0.762188    
MultipleLines[T.No phone service]        0.005523977 0.089240616   0.062 0.950644    
MultipleLines[T.Yes]                     0.058661440 0.024409864   2.403 0.016279 *  
OnlineBackup[T.No internet service]               NA          NA      NA       NA    
OnlineBackup[T.Yes]                     -0.011302620 0.024489569  -0.462 0.644434    
OnlineSecurity[T.No internet service]             NA          NA      NA       NA    
```



```
OnlineSecurity[T.Yes]                     -0.042506787  0.024861440  -1.710 0.087357 .
PaperlessBilling                           0.044907273  0.009989933   4.495 7.06e-06 ***
Partner                                   -0.000850695  0.010807747  -0.079 0.937264
PaymentMethod[T.Credit card (automatic)] -0.006069673  0.013549376  -0.448 0.654191
PaymentMethod[T.Electronic check]          0.067564514  0.013284961   5.086 3.76e-07 ***
PaymentMethod[T.Mailed check]             -0.006745684  0.014500683  -0.465 0.641804
PhoneService                                        NA           NA      NA       NA
SeniorCitizen                              0.044452413  0.013000592   3.419 0.000631 ***
StreamingMovies[T.No internet service]              NA           NA      NA       NA
StreamingMovies[T.Yes]                     0.065756331  0.045026507   1.460 0.144227
StreamingTV[T.No internet service]                  NA           NA      NA       NA
StreamingTV[T.Yes]                         0.063781558  0.045041140   1.416 0.156798
TechSupport[T.No internet service]                  NA           NA      NA       NA
TechSupport[T.Yes]                        -0.043923028  0.025038934  -1.754 0.079442 .
tenure                                    -0.001962576  0.000501008  -3.917 9.04e-05 ***
TotalCharges                              -0.000044379  0.000006477  -6.852 7.92e-12 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```



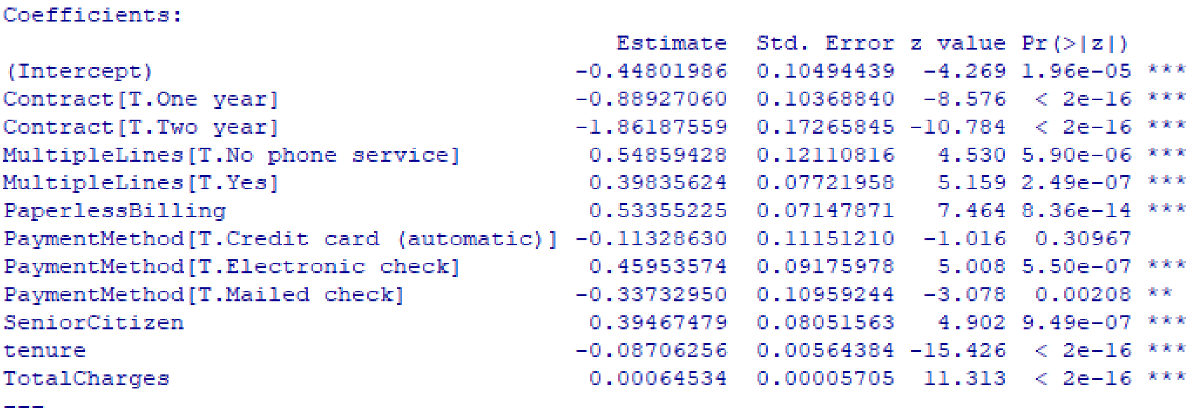
```
Residual standard error: 0.3744 on 7008 degrees of freedom
  (11 observations deleted due to missingness)
Multiple R-squared:  0.2841, Adjusted R-squared:  0.2817
F-statistic: 120.9 on 23 and 7008 DF,  p-value: < 2.2e-16
```

From here, we can the p-value is less than 2.2e-16. This represents that overall linear model is statistically significant for analysis.

Now, when we look into the coefficients of each Independent variables, we take the variables for which the p-value is less than 0.05. Only these variables are considered for the Generalized Linear model (Logistic Regression). The variables for which the p-value less than 0.05 are:

- Contract
- Multiple Lines
- Paperless billing
- Payment method
- Senior Citizen
- Tenure
- Total charges

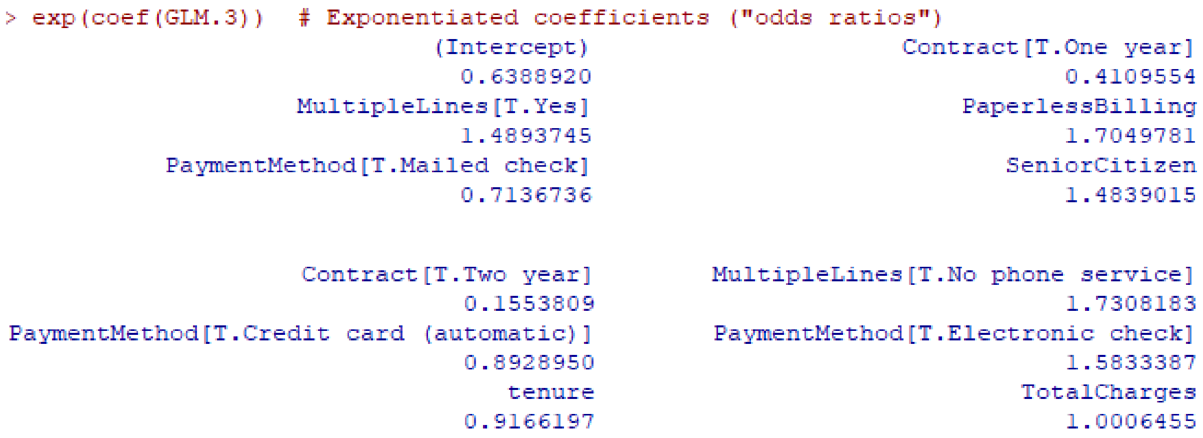
Now, let's go into Generalized Linear Model (GLM). The result of GLM from R commander is given below:



```
Coefficients:
                                            Estimate  Std. Error z value Pr(>|z|)
(Intercept)                              -0.44801986  0.10494439  -4.269 1.96e-05 ***
Contract[T.One year]                     -0.88927060  0.10368840  -8.576  < 2e-16 ***
Contract[T.Two year]                     -1.86187559  0.17265845 -10.784  < 2e-16 ***
MultipleLines[T.No phone service]         0.54859428  0.12110816   4.530 5.90e-06 ***
MultipleLines[T.Yes]                      0.39835624  0.07721958   5.159 2.49e-07 ***
PaperlessBilling                          0.53355225  0.07147871   7.464 8.36e-14 ***
PaymentMethod[T.Credit card (automatic)] -0.11328630  0.11151210  -1.016  0.30967
PaymentMethod[T.Electronic check]         0.45953574  0.09175978   5.008 5.50e-07 ***
PaymentMethod[T.Mailed check]            -0.33732950  0.10959244  -3.078  0.00208 **
SeniorCitizen                             0.39467479  0.08051563   4.902 9.49e-07 ***
tenure                                   -0.08706256  0.00564384 -15.426  < 2e-16 ***
TotalCharges                              0.00064534  0.00005705  11.313  < 2e-16 ***
---
```

From the GLM analysis, we can see that the p-value for all the variables selected are less than 0.05, except PaymentMethon[T.Credit card (automatic)] = 0.30967

Now, if we exponentiate the coefficient values that we got from the model, we get the exponential coefficients as given below:



```
> exp(coef(GLM.3))  # Exponentiated coefficients ("odds ratios")
                          (Intercept)                     Contract[T.One year] 
                            0.6388920                                0.4109554 
                     MultipleLines[T.Yes]                         PaperlessBilling 
                            1.4893745                                1.7049781 
              PaymentMethod[T.Mailed check]                            SeniorCitizen 
                            0.7136736                                1.4839015 

                    Contract[T.Two year]        MultipleLines[T.No phone service] 
                            0.1553809                                1.7308183 
 PaymentMethod[T.Credit card (automatic)]    PaymentMethod[T.Electronic check] 
                            0.8928950                                1.5833387 
                                   tenure                             TotalCharges 
                            0.9166197                                1.0006455 
```

From the values, the following inferences are made:

- For Contract [One year], the chances of “No Churn” is (1/0.411) 2.43 times more than the Contract Month-month and Two year.
- For Two-year Contract, the chance of “No Churn” is (1/0.1554) 6.43 times more than the month-month and one-year Contract. So, if the contract period is more, the chances of Churn are less.
- For Multiple Lines [yes], the chances of “Churn” (‘yes churn’) is 1.489 times more than the Multiple Lines [No and No phone service]
- For Multiple Lines [No phone service], the chances of “Churn” are 1.73 times more than the Multiple Lines [yes and no].
- For paperless Billing [1], the chances of “Churn” are 1.7 times more than the paperless billing [0]
- For Payment Method [Mailed check], the chances of “No churn” is (1/0.714) 1.4 times more than the Payment Method [Electronic, credit card, Bank transfer].
- For Credit card as the Payment Method, the chances for “Churn” or “No churn” are almost same. It means, the credit card payment method is not really a reason for the customers to Churn or Not to churn.
- For electronic Check Payment method, the chances of “Churn” are 1.5 times more than the payment method with Mailed check, Credit card, and bank transfer.
- For senior citizen [1], the chances of “Churn” are 1.48 times more than the Senior citizen [0] (who are not senior citizens). Means, youngsters’ chance of churn is less.
- For Tenure, the chances of “Churn” or “No churn” are almost same. The variable is not really contributing for a customer to Churn or Not churn.
- For Total Charges, we can see the Odds Ratio is 1. Which means, the chances for “Churn” or “No churn” because of Total charges are same. This variable doesn’t contribute to the decision of a customer to Churn or Not to churn.

From the above inferences, we can decide upon the factors that makes a customer to Churn or not Churn. Also, the variable – Tenure, Total Charges, and Credit card Payment method, is invariable for a customer for his/her decision point to stay or not stay with the Telcom service.