Telco customer churn analysis

A telecommunications company seeks to gain deeper insights into its customer churn rates. To achieve this, a comprehensive analysis of their customer data is conducted using R.

The objective is to uncover patterns and drivers behind customer attrition, enabling the company to develop targeted strategies for reducing churn and enhancing customer retention.

A shiny app was created to visualise the insights obtained and allow the stakeholders to interact with it.

Based on insights from the data, I recommend that the company

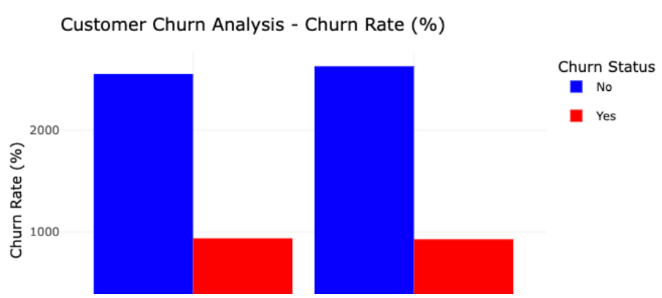
1. Promote one-year and two-year contracts by offering incentives to reduce churn.
2. Investigate issues with electronic checks and explore alternatives or enhancements to improve retention.
3. Implement strategies to engage and retain new customers, especially those with short tenures.
4. Assess and adjust pricing to ensure it meets customer expectations and reduces churn among high-charge customers.
5. Continuously gather and analyse feedback to address churn-related issues and improve service offerings.

These steps aim to address key factors driving churn and improve overall customer retention.

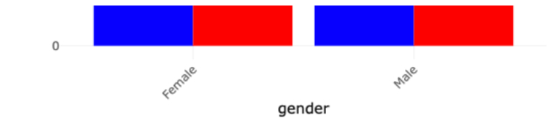
```
# A tibble: 2 × 3
  Churn count percentage
  <fct> <int>      <dbl>
1 No     5163       73.4
2 Yes    1869       26.6
```

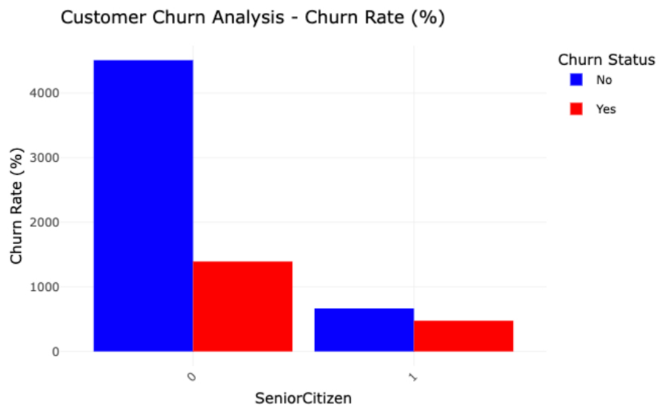
This data reveals that 73% of customers remained with the company, meaning they did not churn

Demographic analysis evaluates how gender and senior citizen status relate to customer churn.



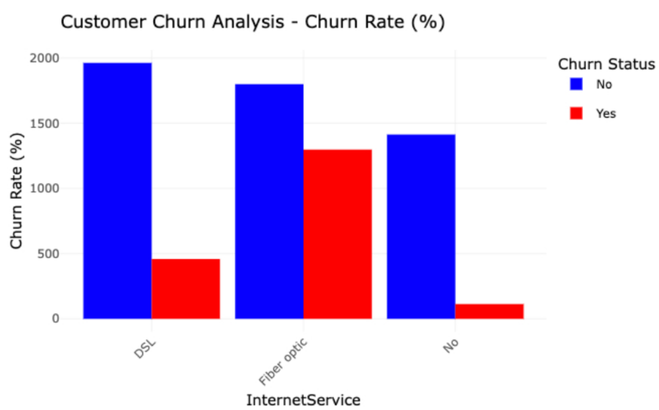
The churn rates are fairly similar between the genders. This insight can guide further investigation into the factors contributing to churn within each gender group.





Senior citizens have a higher churn rate compared to non-senior citizens. This suggests senior citizens have a higher likelihood of churn. This should be addressed in customer retention strategies.

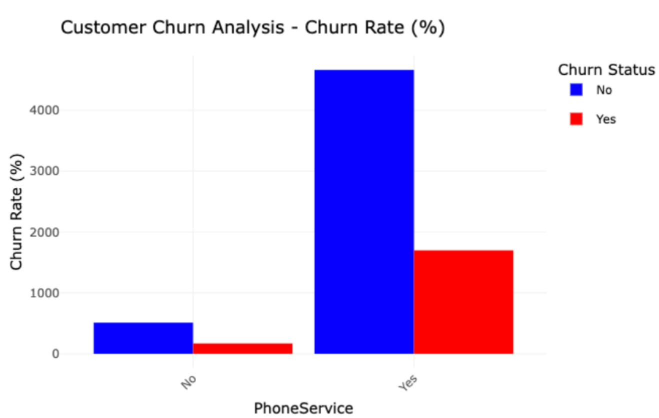
Service usage analysis assess whether customers without phone service exhibit higher churn rates and compare churn rates between customers using DSL and those using fiber optic internet.



2. DSL customers show a moderate churn rate while customers without internet service have the lowest churn rate.

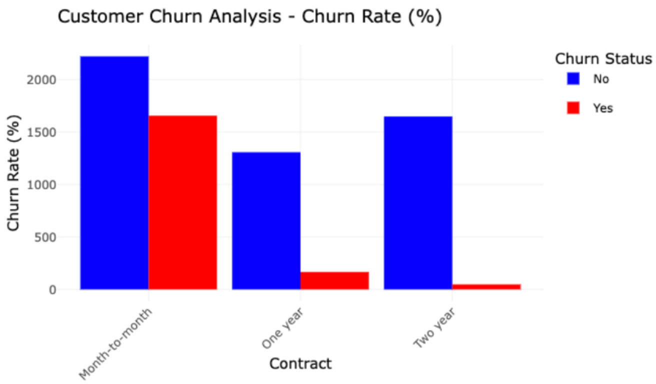
3. These findings highlight the need for targeted churn prevention strategies, especially for fiber optic users who are at a significantly higher risk of leaving compared to those using other service types.

1. Fiber optic customers have the highest churn rate which is a concern for retention. This is likely due to issues such as higher costs or dissatisfaction with service.



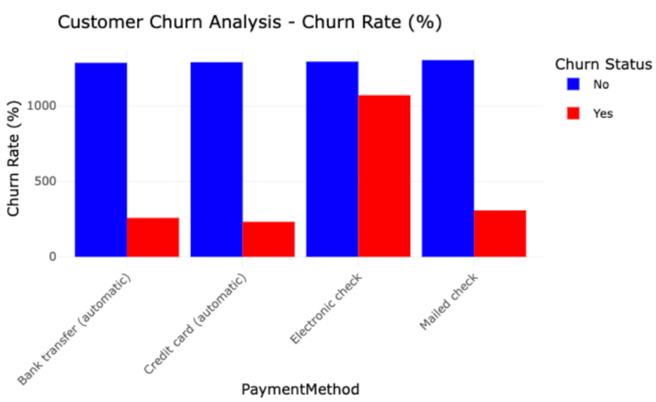
The churn rates for customers with phone service have higher churn rate compared to those without.

Contract and payment method analysis reveals the churn rates based on contract type, noting that month-to-month contracts often show higher churn, It also investigates whether certain payment methods, such as electronic checks, are linked to increased churn rates.



Customers with month-to-month contracts have the highest churn rate while those with one-year contracts and two-year contracts have significantly lower churn rates.

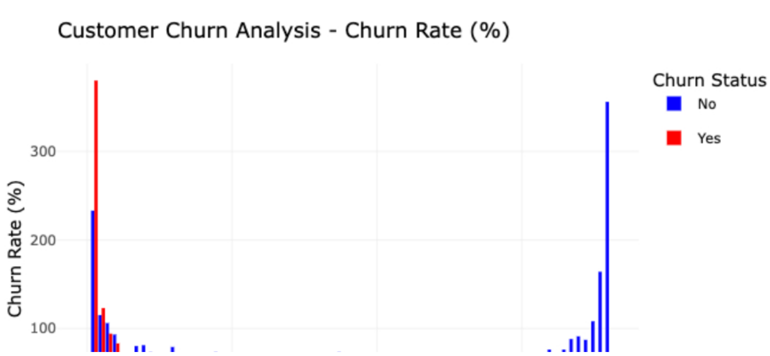
This suggests that longer contract terms are associated with lower churn, highlighting the potential benefits of encouraging customers to commit to longer contract periods to improve retention.



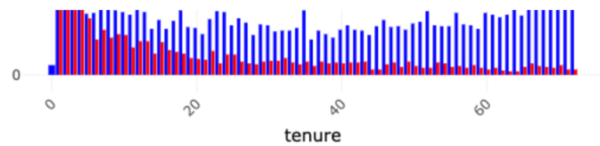
Customers using electronic checks have the highest churn rate suggesting that this payment method is associated with higher customer attrition.

In contrast, customers using bank transfers (automatic) and credit cards (automatic) have lower churn rates. Mailed checks fall in between them. These insights suggest that addressing issues related to electronic check payments could be key to improving overall customer retention.

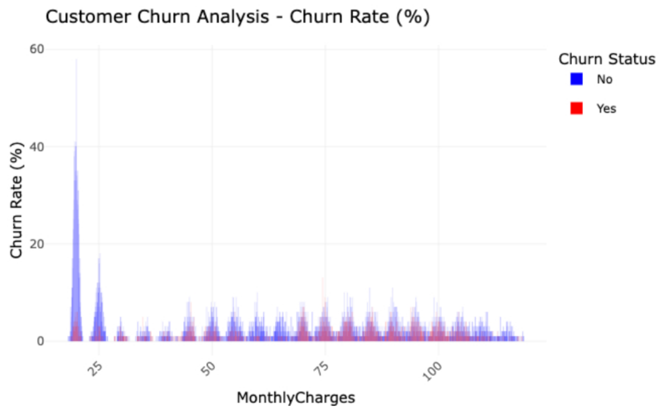
Financial analysis shows whether higher monthly charges are associated with an increased likelihood of churn. It analyses the relationship between customer tenure and churn.



3. This trend underscores the importance of early engagement and retention strategies to reduce churn in the initial months of customer relationships.



1. Customers with shorter tenures suggesting they are more likely to leave soon after starting service.
2. In contrast, those with longer tenures (67 months upwards) show significantly lower churn rates, indicating greater customer retention over time.



This indicates that as monthly charges increase, there may be a higher risk of churn, especially if the charges become too high or if customers perceive the service as not meeting their expectations relative to the cost.

The churn rate varies with monthly charges. Low monthly charges are associated with no churn, while higher monthly charges have a mix of churn rates including some extreme cases of 100% churn.

Predictive modeling uses logistic regression to predict churn based on features such as contract type, payment method, and monthly charges, and employ random forest or decision trees to identify the most significant factors driving customer churn. To decide which features to select, Recursive Feature Elimination (RFE) method was used to recursively removes features and build a model on those that remain. RFE ranks features by importance and selects the best subset. Then, the data is split into training and test samples before using logistic regression to predict customer churn.

```
 [1] "tenure"           "TotalCharges"     "MonthlyCharges"   "Contract"
 [5] "TechSupport"      "OnlineSecurity"   "InternetService"  "OnlineBackup"
 [9] "PaperlessBilling" "MultipleLines"    "PaymentMethod"    "SeniorCitizen"
[13] "StreamingMovies"  "StreamingTV"      "DeviceProtection" "PhoneService"
[17] "Dependents"       "Partner"          "customerID"       "gender"
```

Tenure, TotalCharges, MonthlyCharges were ranked higher Dependants, Partner, and Gender.

```
> summary(model)

Call:
glm(formula = Churn ~ tenure + TotalCharges + MonthlyCharges + 
    Contract + OnlineSecurity + TechSupport, family = binomial, 
    data = train)
```

```
Coefficients: (1 not defined because of singularities)
                                    Estimate Std. Error z value Pr(>|z|)
(Intercept)                       -5.308e-01  1.993e-01  -2.664 0.007730 **
tenure                            -5.819e-02  7.544e-03  -7.713 1.23e-14 ***
TotalCharges                       3.182e-04  8.625e-05   3.690 0.000224 ***
MonthlyCharges                     1.719e-02  2.757e-03   6.234 4.54e-10 ***
ContractOne year                  -8.258e-01  1.240e-01  -6.658 2.77e-11 ***
ContractTwo year                  -1.798e+00  2.214e-01  -8.122 4.57e-16 ***
OnlineSecurityNo internet service -1.042e+00  1.791e-01  -5.815 6.06e-09 ***
OnlineSecurityYes                 -7.040e-01  9.841e-02  -7.154 8.46e-13 ***
TechSupportNo internet service            NA         NA      NA       NA
TechSupportYes                    -5.331e-01  9.839e-02  -5.419 6.00e-08 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 5699.5  on 4921  degrees of freedom
Residual deviance: 4147.1  on 4913  degrees of freedom
AIC: 4165.1

Number of Fisher Scoring iterations: 6
```

The model indicates that customer **tenure**, contract type, and service options are crucial in predicting churn, helping to identify key areas for customer retention strategies. The key insights garnered include:

- Selected features that are predictors of churn are tenure, TotalCharges, MonthlyCharges, Contract type, OnlineSecurity, TechSupport).
- Having a longer-term contract (One year or Two year) greatly reduces churn likelihood.
- Higher charges (monthly and total) increase churn risk, though total charges have a smaller effect.
- Customers with online security and tech support are less likely to churn especially compared to those without these services.

In general, the deviance reduction and the AIC value suggest that your model is better than a model with no predictors. However, it’s crucial to interpret compare with alternative models to confirm that it is the best model.

```
> print(conf_matrix)
Confusion Matrix and Statistics

          Reference
Prediction   No  Yes
       No  1345  264
       Yes  204  297

               Accuracy : 0.7782
                 95% CI : (0.7599, 0.7958)
    No Information Rate : 0.7341
    P-Value [Acc > NIR] : 1.728e-06

                  Kappa : 0.4118

 Mcnemar's Test P-Value : 0.006386
```

```
            Sensitivity : 0.8683
            Specificity : 0.5294
         Pos Pred Value : 0.8359
         Neg Pred Value : 0.5928
             Prevalence : 0.7341
         Detection Rate : 0.6374
   Detection Prevalence : 0.7626
      Balanced Accuracy : 0.6989

       'Positive' Class : No
```

The confusion matrix provides counts of true positives, false positives, true negatives, and false negatives. The model has a good accuracy. 77.82% of the time, the model correctly predicts whether a customer will churn or not. Also, the model has high sensitivity, indicating it is good at detecting churn. However, it could benefit from further tuning to improve its ability to correctly predict non-churners and to balance performance across both classes.