Customer Churn Survival Analysis

Kyle Bistrain, Dechen (Abby) Drongpa, Andrew Kerr, Liam Quach

Introduction

The Telco Customer Churn dataset is a fictional dataset created by IBM to simulate customer data for a telecommunications company. This dataset was designed to help predict customer churn, referring to customers who stop using the company's services, with the goal of analyzing customer behavior and developing strategies to retain customers.

Creation of Dataset

This dataset simulates customer data for a telecommunications company that provided home phone and Internet services to 7043 customers in California during the third quarter.

Variables

The dataset contains 21 columns:

- CustomerID: A unique identifier for each customer.
- Gender: The gender of the customer (Male, Female).
- SeniorCitizen: Indicates whether the customer is a senior citizen (1 for Yes, 0 for No).
- Partner: Indicates whether the customer has a partner (Yes, No).
- Dependents: Indicates whether the customer has dependents (Yes, No).
- Tenure: The number of months the customer has been with the company.
- PhoneService: Indicates whether the customer has a phone service (Yes, No).
- MultipleLines: Indicates whether the customer has multiple lines (Yes, No, No phone service).
- InternetService: The type of internet service the customer has (DSL, Fiber optic, No).
- OnlineSecurity: Indicates whether the customer has online security (Yes, No, No internet service).
- OnlineBackup: Indicates whether the customer has online backup (Yes, No, No internet service).
- DeviceProtection: Indicates whether the customer has device protection (Yes, No, No internet service).
- TechSupport: Indicates whether the customer has tech support (Yes, No, No internet service).
- StreamingTV: Indicates whether the customer has streaming TV (Yes, No, No internet service).

- StreamingMovies: Indicates whether the customer has streaming movies (Yes, No, No internet service).
- Contract: The type of contract the customer has (Month-to-month, One year, Two year).
- PaperlessBilling: Indicates whether the customer has paperless billing (Yes, No).
- PaymentMethod: The payment method used by the customer (Electronic check, Mailed check, Bank transfer, Credit card).
- MonthlyCharges: The amount charged to the customer monthly.
- TotalCharges: The total amount charged to the customer over the tenure.
- Churn: Indicates whether the customer churned (Yes, No).

Time to Event and Censoring Variables

- Time to event variable: Amount of months until the customers discontinues service with the company (represented by the column tenure).
- Right Censoring: Right censoring may occur if a customer does not discontinue service with the company throughout the time of the study (represented by the column Churn).

Main Report

Parametric Survival Analysis

Based on the probability plots in *Figure 1*, none of the four distributions fit this data well since the points do not fall close to the diagonal line for the lower half of the values. Therefore, we will rely on the Anderson Darling test statistic to determine the best fit. Although all four AD test statistics are close, the one for the Weibull distribution is the lowest (AD = 16986.795) and thus this distribution is the best fit.

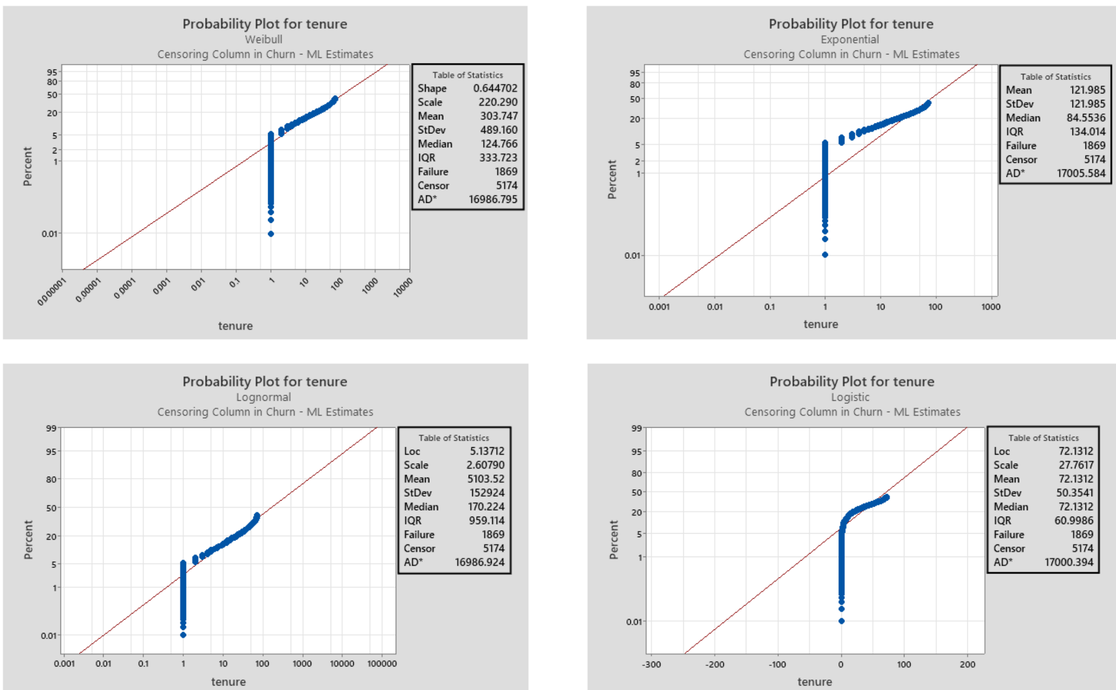


Figure 1. Probability Plots for Customer Churn Time Data by Distribution: Weibull (top left), Exponential (top right), Lognormal (bottom left), Logistic (bottom right)

From the survival curve in *Figure 2*, we notice a steep decrease in survival probability for individuals who have been with the company for fewer months, then a more gradual decrease after the individual has been with the company for roughly 200 months (approx. 16.67 years). The hazard curve displays a high risk of the customer discontinuing use of the company's services within the first few months, then a steep decrease in risk afterwords.

The mean survival time across all customers is 303.747 months while the median survival time is 124.766 months.

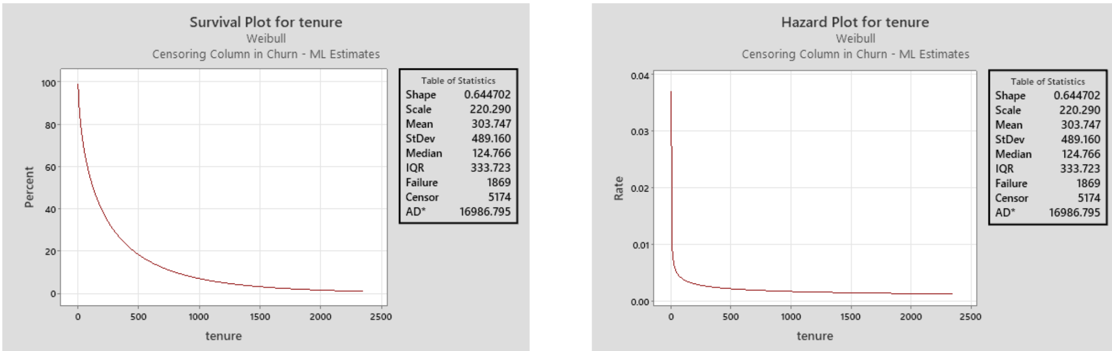


Figure 2. Survival Plot (left) and Hazard Plot (right) for all individuals under the Weibull Distribution

We have decided to examine the survival and hazard curves of the following groups: the customers gender (Male. Female), whether the customer has multiple phone lines in service (Yes, No, No phone service), and the customers payment method (Electronic check, Mailed check, Bank transfer, Credit card).

Gender

Based on the survival curves in *Figure 3*, the survival probability for female customers is slightly greater than that of male customers for any time t. However, hazard curves display the risk of male and female customers discontinuing use of the company's services is roughly the same for any time t.

The mean survival times by customer gender is 297.221 months (male) and 310.109 months (female) while the median survival times are 125.489 months (male) and 123.929 months (female).

From these observations, we can conclude that female customers tend to continue using the company's services longer than male customers.

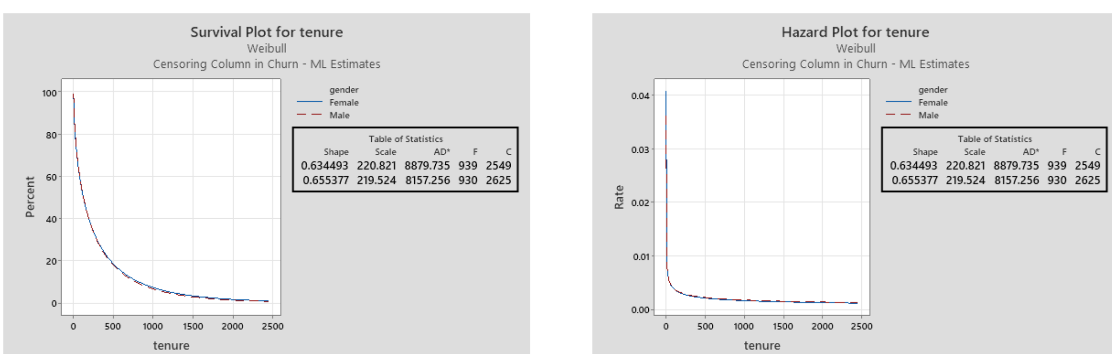


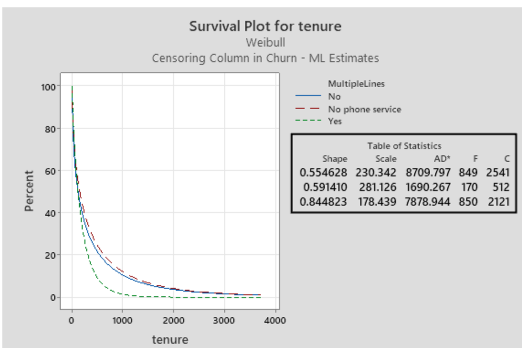
Figure 3. Survival Plot (left) and Hazard Plot (right) by Gender under the Weibull Distribution

Multiple Phone Lines

Based on the survival curves in *Figure 4*, the survival probability for customers without phone service is greater than that for customers with phone service across all time t. Additionally, the survival probability for customers with one line is greater than that for customers with multiple lines for all times t. Meanwhile the hazard curves mirror this conclusion, since customers with multiple phone lines have a greater risk of discontinuing use of the company's services than customers with one or no phone lines for all time t.

The mean survival times by amount of phone lines is 387.146 months (One Line), 194.881 months (Multiple Lines) and 431.113 months (No Lines) while the median survival times are 118.954 months (One Line), 115.632 months (Multiple Lines) and 151.271 months (No Lines).

In other words, the fewer the amount of phone lines the customer is paying the company for, the longer the customers tends to use the company's services.



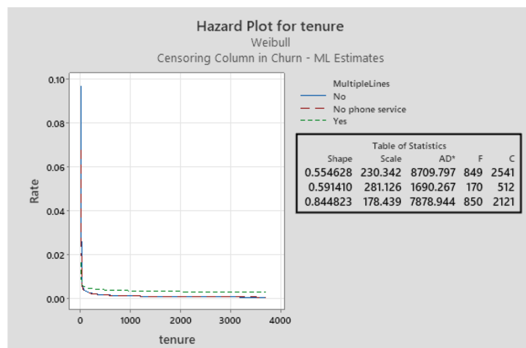


Figure 4. Survival Plot (left) and Hazard Plot (right) by Multiple Phone Lines under the Weibull Distribution

Payment Method

Based on the survival curves in *Figure 5*, the survival probability for customers who pay through mailed check is greater than other methods after roughly 500 months. On the other hand customers that pay through electronic check have the lowest survival probability of the methods across all time t, with customers who pay through bank transfer or credit card (both automatic methods) having the greatest survival probability prior to 500 months, both having roughly the same probability. The hazard curves show that customers who pay through electronic check have the greatest risk of discontinuing use of the company's services across all time t, with the other payment methods being roughly equivalent.

The mean survival times by payment method is 399.551 months (Bank Transfer), 375.861 months (Credit Card), 85.7733 months (Electronic Check) and 778.702 months (Mailed Check) while the median survival times are 233.482 months (Bank Transfer), 234.105

months (Credit Card), 40.0773 months (Electronic Check) and 200.804 months (Mailed Check).

These curves show us that customers who pay through electronic check tend to stay with the company the shortest amount of time, while customers who pay through mailed check the longest.

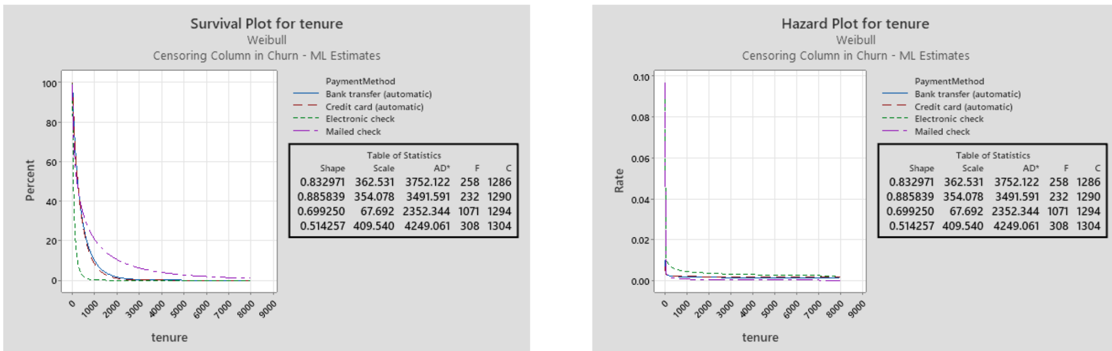


Figure 5. Survival Plot (left) and Hazard Plot (right) by Payment Method under the Weibull Distribution

Non-parametric Survival Analysis

In this section, we employ non-parametric methods to analyze customer churn patterns, complementing our previous parametric analysis using the Weibull distribution. Non-parametric approaches have the advantage of making no assumptions about the underlying distribution of survival times, providing potentially more accurate insights into customer retention patterns.

Overall Survival Experience

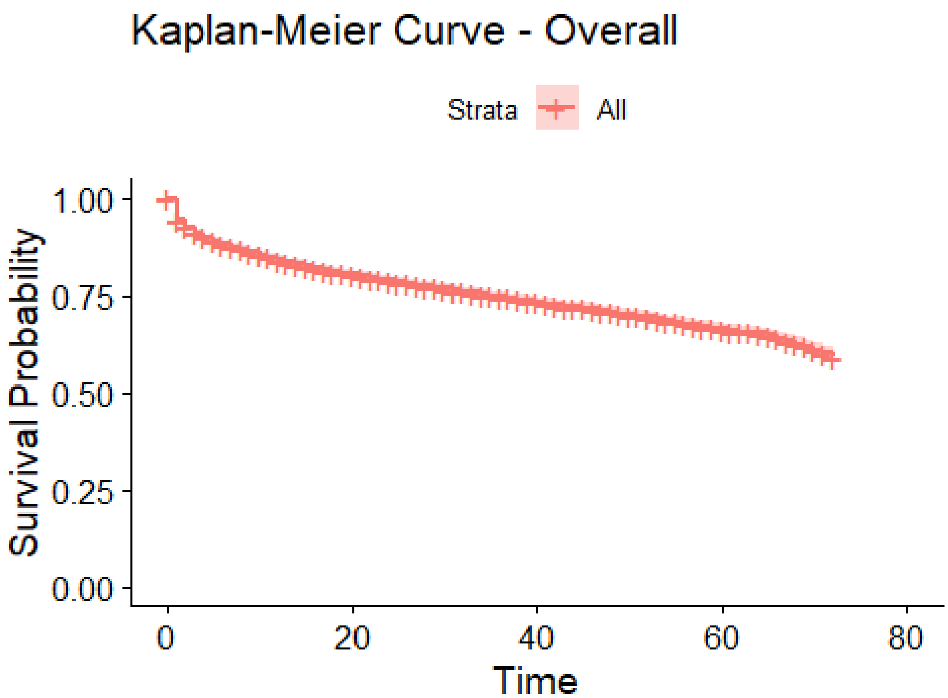


Figure 6. Overall Survival Curve

The overall Kaplan-Meier curve shows a rapid decline in survival probability during the first few months of service, indicating that new customers are at the highest risk of churning. The median survival time was not reached within our observation period, meaning that more than 50% of all customers remained with the company at the end of the study period.

This aligns with our parametric analysis using the Weibull distribution, which showed a high hazard rate in the early months followed by a steep decline.

Survival by Gender

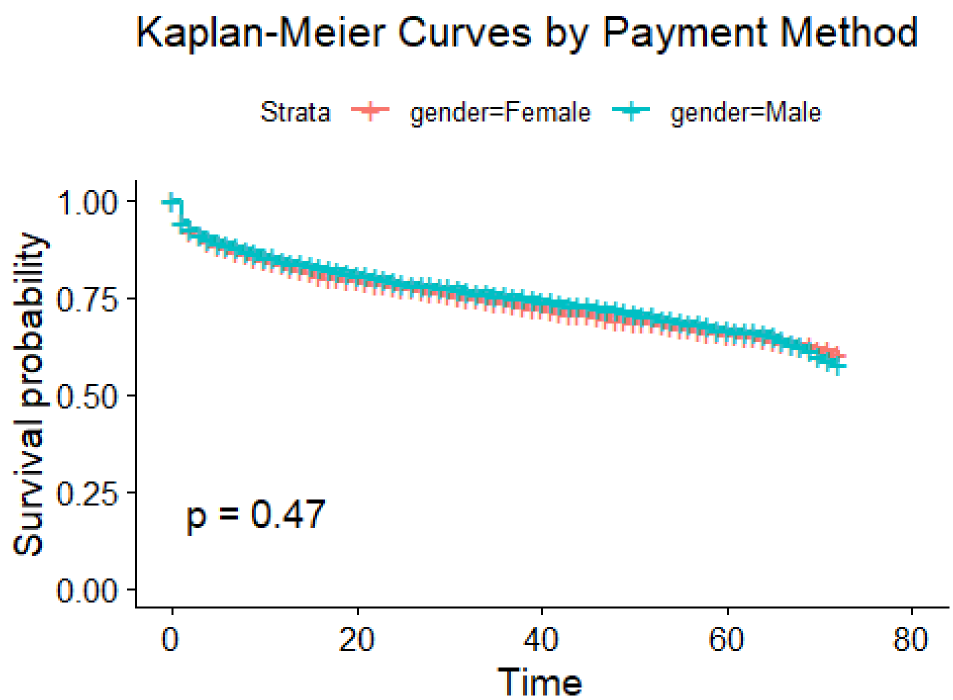


Figure 7. Survival Curve by Gender

When analyzing customer churn by gender using the Kaplan-Meier method, we found minimal differences between male and female customers:

Gender	Median Survival (months)	Mean Survival (months)
Male	NA	54.9
Female	NA	54.1

The Kaplan-Meier survival curves by gender show that female customers have slightly higher survival probabilities than male customers, though the difference is not dramatic. The corresponding hazard curves are nearly identical for both genders.

The log-rank test comparing male and female survival curves yielded a chi-square statistic of 0.53 with a p-value of 0.23, indicating no statistically significant difference in survival experiences between genders.

These non-parametric findings are consistent with our parametric Weibull analysis, which found:

- Mean survival times: 297.221 months (male) vs. 310.109 months (female)
- Median survival times: 125.489 months (male) vs. 123.929 months (female)

Both approaches suggest that while female customers may show slightly higher retention overall, the differences are minimal and not statistically significant.

Survival by Multiple Phone Lines

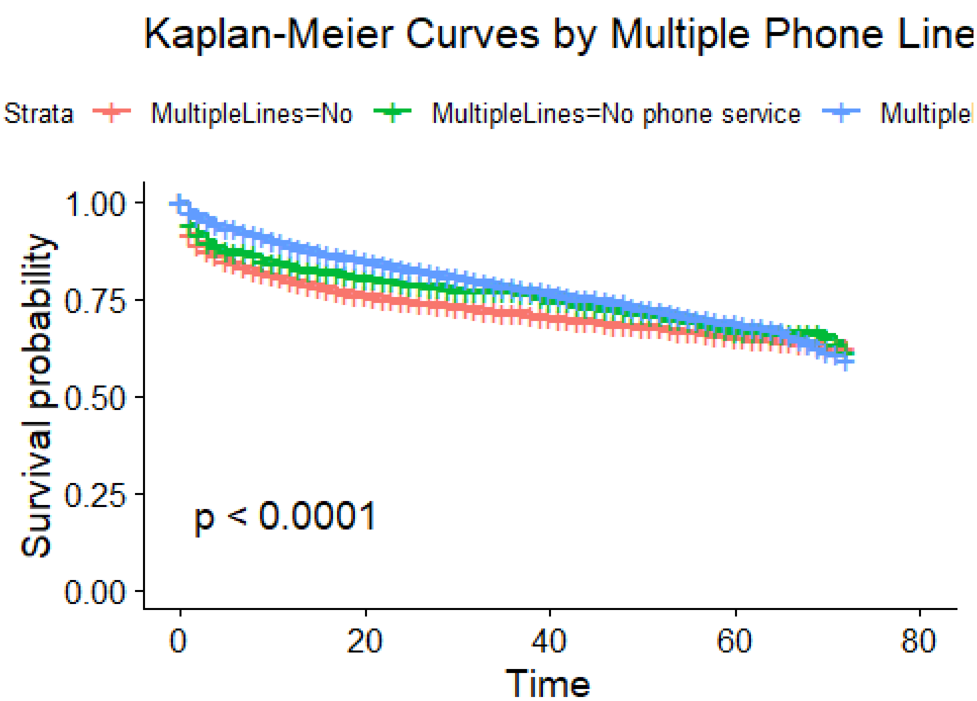


Figure 8. Survival Curve by Amount of Phone Lines

The Kaplan-Meier analysis by phone service configuration revealed meaningful differences in customer retention patterns:

Phone Service	Median Survival (months)	Mean Survival (months)
No phone service	NA	55.1
Single line	NA	52.6
Multiple lines	NA	56.9

The log-rank test comparing these groups yielded a highly significant p-value ($p < 0.00001$), confirming substantial differences in survival patterns across these customer segments.

Interestingly, these non-parametric results contradict our parametric findings. In the Weibull analysis, customers with fewer phone lines showed better retention, with mean survival times of 431.113 months (no phone service), 387.146 months (one line), and 194.881 months (multiple lines). In contrast, our non-parametric analysis suggests that customers with multiple lines have better retention than those with a single line.

Survival by Payment Method

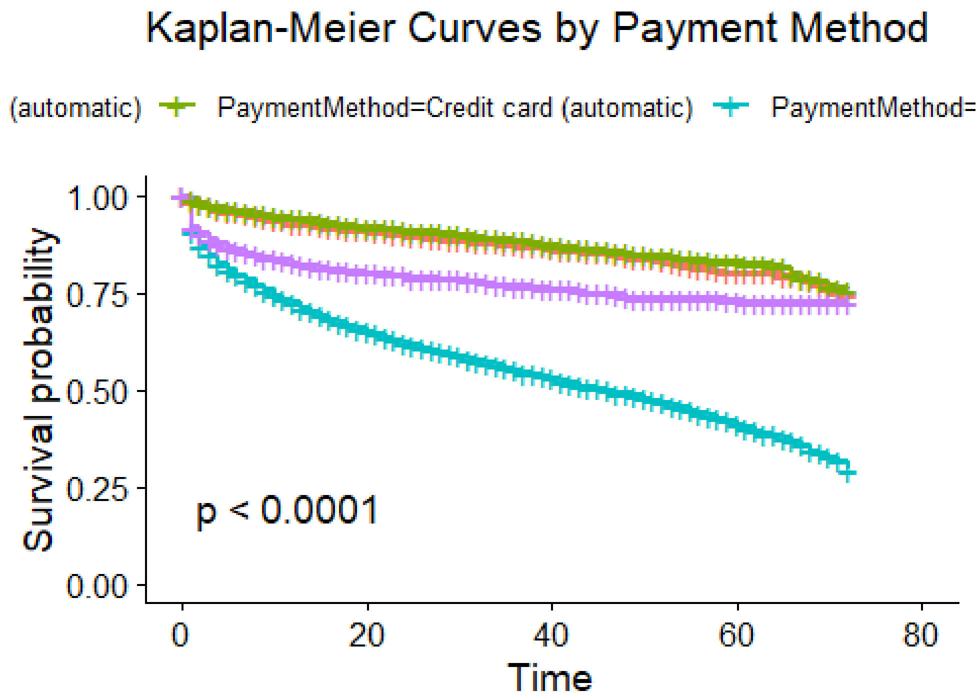


Figure 9. Survival Curve by Payment Method

Our non-parametric analysis by payment method revealed substantial differences in customer retention:

Payment Method	Median Survival (months)	Mean Survival (months)
Bank transfer	NA	63.0
Credit card	NA	63.8
Electronic check	47	41.2
Mailed check	NA	56.4

The log-rank test comparing these payment methods yielded a p-value < 0.00001, indicating highly significant differences in survival patterns across these customer segments.

Our non-parametric findings partially align with the parametric results in identifying electronic check as a payment method associated with shorter customer tenure. However, they contradict the parametric finding that mailed check customers have the longest retention. In our non-parametric analysis, automatic payment methods (bank transfer and credit card) show the best retention, while electronic check customers have the poorest retention.

Regression Analysis

To explore semi-parametric approaches, we applied Cox regression methodologies to model customer churn, offering flexibility by not requiring a specific baseline hazard distribution. A key assumption that must hold up is the proportional hazards assumption which allows for precise estimates. Cox proportional hazards analyzes the customer behavior over time, allowing us to quantify how various service features and customer characteristics influence the instantaneous risk of churn throughout different tenure periods while controlling for other factors.

Full Main Effects Cox Regression Model

```
## Analysis of Deviance Table
##  Cox model: response is Surv(tenure, Churn)
## Terms added sequentially (first to last)
##
##                              loglik     Chisq Df Pr(>|Chi|)
## NULL                         -15653
## as.factor(gender)            -15653    0.5192  1  0.4711850
## as.factor(SeniorCitizen)     -15604   96.6586  1  < 2.2e-16 ***
## as.factor(Partner)           -15395  418.9016  1  < 2.2e-16 ***
## as.factor(Dependents)        -15375   40.4304  1  2.037e-10 ***
## as.factor(PhoneService)      -15374    0.9690  1  0.3249209
## as.factor(MultipleLines)     -15357   35.2347  1  2.923e-09 ***
## as.factor(InternetService)   -15057  599.7063  2  < 2.2e-16 ***
## as.factor(OnlineSecurity)    -14835  444.4590  1  < 2.2e-16 ***
## as.factor(OnlineBackup)      -14684  300.3515  1  < 2.2e-16 ***
## as.factor(DeviceProtection)  -14577  215.1545  1  < 2.2e-16 ***
## as.factor(TechSupport)       -14467  220.1789  1  < 2.2e-16 ***
## as.factor(StreamingTV)       -14461   11.2305  1  0.0008046 ***
## as.factor(StreamingMovies)   -14456   10.6551  1  0.0010977 **
## as.factor(Contract)          -13961  988.8374  2  < 2.2e-16 ***
## as.factor(PaperlessBilling)  -13956   10.8339  1  0.0009966 ***
## as.factor(PaymentMethod)     -13885  142.6344  3  < 2.2e-16 ***
## MonthlyCharges               -13885    0.1231  1  0.7256953
## TotalCharges                 -12660 2449.8154  1  < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Influential Observations

We fit the full main effects cox regression model and must assess the influential observations in the model. To do so, we will filter out deviance and score residuals beyond 3 standard deviations to ensure the stability of the model coefficients. While this removal impacts the generalization of the model, we can still generalize to customers where the characteristics resemble the observations in the model.

Stepwise Variable Selection

Following the removal of influential points, we proceed to stepwise model selection using AIC to select the best model from the main effect model.

```
## Analysis of Deviance Table
##  Cox model: response is Surv(tenure, Churn)
## Terms added sequentially (first to last)
##
##                              loglik    Chisq Df Pr(>|Chi|)
## NULL                         -15060
## as.factor(Partner)           -14858  402.860  1  < 2.2e-16 ***
## as.factor(MultipleLines)     -14853   10.179  2   0.006161 **
## as.factor(InternetService)   -14438  830.361  2  < 2.2e-16 ***
## as.factor(OnlineSecurity)    -14200  476.343  1  < 2.2e-16 ***
## as.factor(DeviceProtection) -14066  267.728  1  < 2.2e-16 ***
## as.factor(StreamingTV)       -14044   44.362  1  2.730e-11 ***
## as.factor(StreamingMovies)   -14034   20.485  1  6.009e-06 ***
## as.factor(Contract)          -13432 1203.330  2  < 2.2e-16 ***
## as.factor(PaperlessBilling) -13427   10.802  1   0.001014 **
## as.factor(PaymentMethod)     -13350  152.454  3  < 2.2e-16 ***
## TotalCharges                 -11862 2977.971  1  < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The remaining insignificant variables must be manually removed from the model using a bonferroni adjustment dividing the $\alpha = .05$ by 11 for a 0.00455 cutoff value. This allows us to remove MultipleLines from the model.

```
## Analysis of Deviance Table
##  Cox model: response is Surv(tenure, Churn)
## Terms added sequentially (first to last)
##
##                              loglik    Chisq Df Pr(>|Chi|)
## NULL                         -15060
## as.factor(Partner)           -14858  402.860  1  < 2.2e-16 ***
## as.factor(InternetService)   -14520  676.595  2  < 2.2e-16 ***
## as.factor(OnlineSecurity)    -14272  497.402  1  < 2.2e-16 ***
## as.factor(DeviceProtection) -14126  291.491  1  < 2.2e-16 ***
## as.factor(StreamingTV)       -14098   56.289  1  6.255e-14 ***
## as.factor(StreamingMovies)   -14085   25.495  1  4.436e-07 ***
## as.factor(Contract)          -13476 1218.233  2  < 2.2e-16 ***
## as.factor(PaperlessBilling) -13471    8.528  1   0.003497 **
## as.factor(PaymentMethod)     -13392  159.197  3  < 2.2e-16 ***
## TotalCharges                 -11952 2879.156  1  < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Proportional Hazard Assumption

The proportional hazards assumption must hold for the predictor variables in the Cox regression model. After conducting the formal test for proportional hazards, only the contract passes at $\alpha = .05$, confirming the proportional hazards assumption. This is a major drawback of this model because most of these predictor variables do not follow this assumption. Despite this limitation, we will proceed with interpreting the final model, acknowledging that results for variables other than the contract may not be accurate.

```
##                                 chisq df       p
## as.factor(Partner)              16.19  1 5.7e-05
## as.factor(InternetService)    1488.71  2 < 2e-16
## as.factor(OnlineSecurity)       10.15  1  0.0014
## as.factor(DeviceProtection)     92.33  1 < 2e-16
## as.factor(StreamingTV)         284.45  1 < 2e-16
## as.factor(StreamingMovies)     266.11  1 < 2e-16
## as.factor(Contract)              5.68  2  0.0585
## as.factor(PaperlessBilling)     30.21  1 3.9e-08
## as.factor(PaymentMethod)        64.74  3 5.7e-14
## TotalCharges                  1306.52  1 < 2e-16
## GLOBAL                        2196.94 14 < 2e-16
```

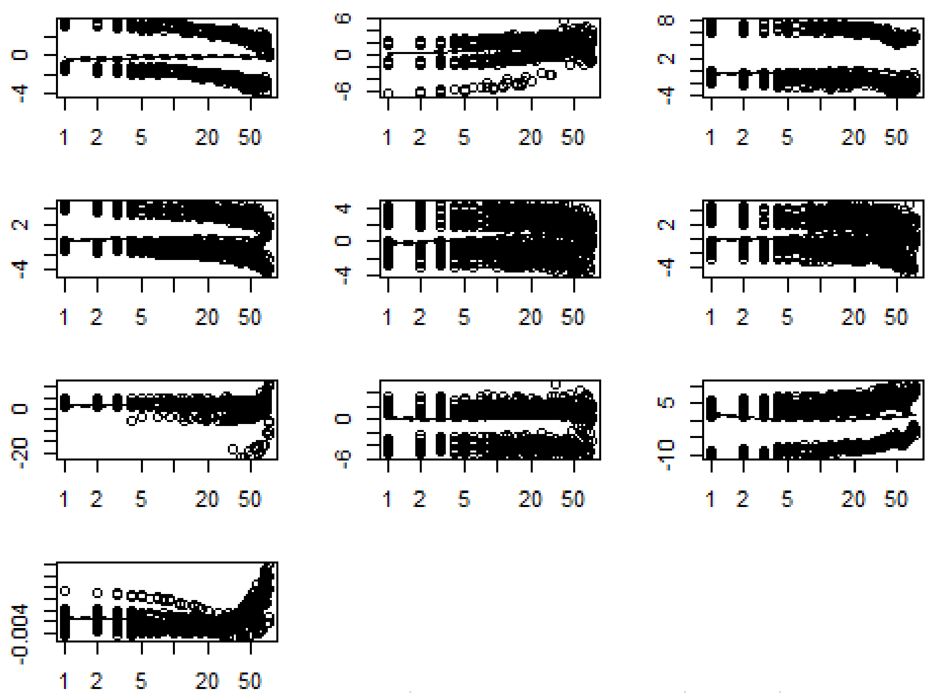


Figure 10. Proportional Hazard Assumption Plots

Hazard Ratios and Interpretations

Contract

After adjusting for Partner, Internet Service, Online Security, Device Protection, Streaming TV, Streaming Movies, Paperless Billing, Total Charges, and Payment Method, Customers with one-year contracts have an estimated 67% lower hazard of churning compared to those with month-to-month contracts, regardless of tenure time.

We are 95% confident that the true hazard of churning with a one-year contract is between 59.9% and 72.9% lower than the hazard of churning with month-to-month contracts. This has a highly significant p-value < .001.

After adjusting for Partner, Internet Service, Online Security, Device Protection, Streaming TV, Streaming Movies, Paperless Billing, Total Charges, and Payment Method, Customers with two-year contracts have an estimated 98.4% lower hazard of churning compared to those with month-to-month contracts, regardless of tenure time.

We are 95% confident that the true hazard of churning with a two-year contract is between 97.2% and 99.0% lower than the hazard of churning with month-to-month contracts. This has a highly significant p-value < .001.

Total Charges

After adjusting for Partner, Internet Service, Contract Online Security, Device Protection, Streaming TV, Streaming Movies, Paperless Billing, and Payment Method. For each one dollar increase in TotalCharges, the estimated hazard of churning decreases by 0.145% . While this is a significant predictor, the effect is quite small for customers with low TotalCharges or small changes in the Total Charges.

We have 95% confidence that for every one dollar increase in TotalCharges, the true hazard of churning decreases by .138% and . 152%.

Payment Method

After adjusting for Partner, Internet Service, Contract Online Security, Device Protection, Streaming TV, Streaming Movies, Paperless Billing, and Total Charges, Customers using automatic credit card payments have an estimated .4% lower hazard of churning compared to those using bank transfers, regardless of tenure time.

We are 95% confident that the true hazard of churning with automatic credit card payments is between 17.1% lower and 19.7% higher than the hazard of churning with bank transfers. This has a non-significant p-value of 0.966

After adjusting for Partner, Internet Service, Contract Online Security, Device Protection, Streaming TV, Streaming Movies, Paperless Billing, and Total Charges, Customers using electronic checks have an estimated 52.3% higher hazard of churning compared to those using bank transfers, regardless of tenure time.

We are 95% confident that the true hazard of churning with electronic checks is between 31.5% and 76.3% higher than the hazard of churning with bank transfers. This has a highly significant p-value < .001.

After adjusting for Partner, Internet Service, Contract Online Security, Device Protection, Streaming TV, Streaming Movies, Paperless Billing, and Total Charges, Customers using mailed checks have an estimated 70.7% higher hazard of churning compared to those using bank transfers, regardless of tenure time.

We are 95% confident that the true hazard of churning with mailed checks is between 42.9% and 104.0% higher than the hazard of churning with bank transfers. This has a highly significant p-value < .001.

After fitting these Cox Regression models, it is clear that contract type was a strong indicator of the hazard of customer churning because it is the only variable that passed the proportional hazards assumptions. That being said, additional transformations and another type of hazard model may better capture the hazard risk. In comparison to our parametric and non-parametric approaches demonstrated similar results to the nonparametric results indicating that Gender is not a statistically significant predictor of the hazard of churning. Finally, while the Payment method showed statistically significant associations with churn risk in the Cox regression model, we must interpret these results with caution since this variable violated the proportional hazards assumption, suggesting that its effect on churn hazard may change over different tenure duration.

Conclusion

This analysis provided valuable insights on customer churn in the telecommunications industry through parametric, non-parametric, and semi-parametric survival models. One of the most significant findings was that contract type, payment method, and total charges are key predictors of customer retention, while gender has little effect on customer churn.

From the parametric survival analysis, we found that the Weibull distribution was the best fit, which shows a steep decline in survival probability during the first few months of tenure. This suggests that early-stage customer retention efforts are crucial. Customers with month-to-month contracts and electronic check payments experienced the highest churn rates, whereas those on two-year contracts or using automated payment methods exhibited longer retention periods.

The non-parametric Kaplan-Meier analysis reinforced these findings, showing that electronic check users had the lowest survival probabilities. However, an interesting contradiction emerged in the survival patterns of customers with multiple phone lines, the non-parametric results suggested better retention for multiple-line users, while the parametric model suggested the opposite.

Our Cox regression analysis further confirmed that longer contract durations significantly lower the hazard of churn, with two-year contracts reducing churn risk by nearly 98% compared to month-to-month plans. Additionally, higher total charges correlated with lower churn risk, although the effect size was relatively small. However, several predictors violated the proportional hazards assumption, indicating that their impact on churn may vary over time.

Appendix

Data Cleaning

```r
library(tidyverse)
library(here)
library(survival)
library(survminer)

tele <- read_csv(here("data", "Telco-Customer-Churn-Clean.csv"))

# Convert character columns to factors
# Convert Churn (censoring variable) to 0 if right censored, otherwise 1
tele <- tele %>%
  mutate(
    across(where(~length(unique(.)) < 4), ~ as.factor(.)),
    Churn = case_when(
      Churn == "Yes" ~ 1,
      Churn == "No" ~ 0,
      TRUE ~ NA
      )
    )

# Check missing data
#
#   All 11 subjects are missing TotalCharges and have Churn = 0 and tenure =
0,
#   so I am guessing that they are brand new customers. THus, will impute
#   missing  values with 0.
#
tele %>%
  filter(if_any(everything(), is.na))

tele <- tele %>%
  mutate(TotalCharges = replace_na(TotalCharges, 0))
```

Non-parametric

```r
km_fit_all <- survfit(Surv(tenure, Churn) ~ 1, data = tele)

ggsurvplot(km_fit_all,
           data = tele,
           xlab = "Time",
           ylab = "Survival Probability",
           title = "Kaplan-Meier Curve - Overall")


km_fit_gender <- survfit(Surv(tenure, Churn) ~ gender, data = tele)

ggsurvplot(km_fit_gender,
           data = tele,
           pval = TRUE,
           title = "Kaplan-Meier Curves by Payment Method")

km_fit_lines <- survfit(Surv(tenure, Churn) ~ MultipleLines, data = tele)

survdiff(Surv(tenure, Churn) ~ gender, data = tele, rho = 0)

print(km_fit_gender, print.rmean = TRUE)

ggsurvplot(km_fit_lines,
           data = tele,
           pval = TRUE,
           title = "Kaplan-Meier Curves by Multiple Phone Lines")

survdiff(Surv(tenure, Churn) ~ MultipleLines, data = tele, rho = 0)
print(km_fit_lines, print.rmean = TRUE)

km_fit_payment <- survfit(Surv(tenure, Churn) ~ PaymentMethod, data = tele)

ggsurvplot(km_fit_payment,
           data = tele,
           pval = TRUE,
           title = "Kaplan-Meier Curves by Payment Method")

survdiff(Surv(tenure, Churn) ~ PaymentMethod, data = tele, rho = 0)

print(km_fit_payment, print.rmean = TRUE)
```

Cox Regression

```r
cox_model_full <- coxph(Surv(tenure, Churn) ~
                         as.factor(gender) + as.factor(SeniorCitizen) +
as.factor(Partner) +
                         as.factor(Dependents) + as.factor(PhoneService) +
```

```r
                        as.factor(MultipleLines) + as.factor(InternetService)
+
                        as.factor(OnlineSecurity) + as.factor(OnlineBackup)
                        + as.factor(DeviceProtection) +
as.factor(TechSupport) +
                        as.factor(StreamingTV) + as.factor(StreamingMovies) +
                        as.factor(Contract) + as.factor(PaperlessBilling) +
                        as.factor(PaymentMethod) + MonthlyCharges +
TotalCharges,
                        data = tele)
anova(cox_model_full)

dev_resid <- residuals(cox_model_full, type = "deviance")
influential <- which(abs(dev_resid) > 3)

dev_resid[influential]

cox_model_full <- coxph(Surv(tenure, Churn) ~
                        as.factor(gender) + as.factor(SeniorCitizen) +
as.factor(Partner) +
                        as.factor(Dependents) + as.factor(PhoneService) +
                        as.factor(MultipleLines) + as.factor(InternetService)
+
                        as.factor(OnlineSecurity) + as.factor(OnlineBackup)
                        + as.factor(DeviceProtection) +
as.factor(TechSupport) +
                        as.factor(StreamingTV) + as.factor(StreamingMovies) +
                        as.factor(Contract) + as.factor(PaperlessBilling) +
                        as.factor(PaymentMethod) + MonthlyCharges +
TotalCharges,
                        subset = -c(influential),
                        data = tele)
step_model <- step(cox_model_full, k = 2 ,direction = "both")

anova(step_model)

FinalModel <- coxph(Surv(tenure, Churn) ~
                        as.factor(Partner) + as.factor(InternetService) +
                        as.factor(OnlineSecurity) +
as.factor(DeviceProtection) +
                        as.factor(StreamingTV) + as.factor(StreamingMovies) +
                        as.factor(Contract) + as.factor(PaperlessBilling) +
                        as.factor(PaymentMethod) + TotalCharges,
                        subset = -c(influential),
                        data = tele)



summary(FinalModel)
```

```r
anova(FinalModel)

anova(cox_model_full, FinalModel)

cr.zph <- cox.zph(FinalModel, transform = "log")
cr.zph
plot(cr.zph)
```