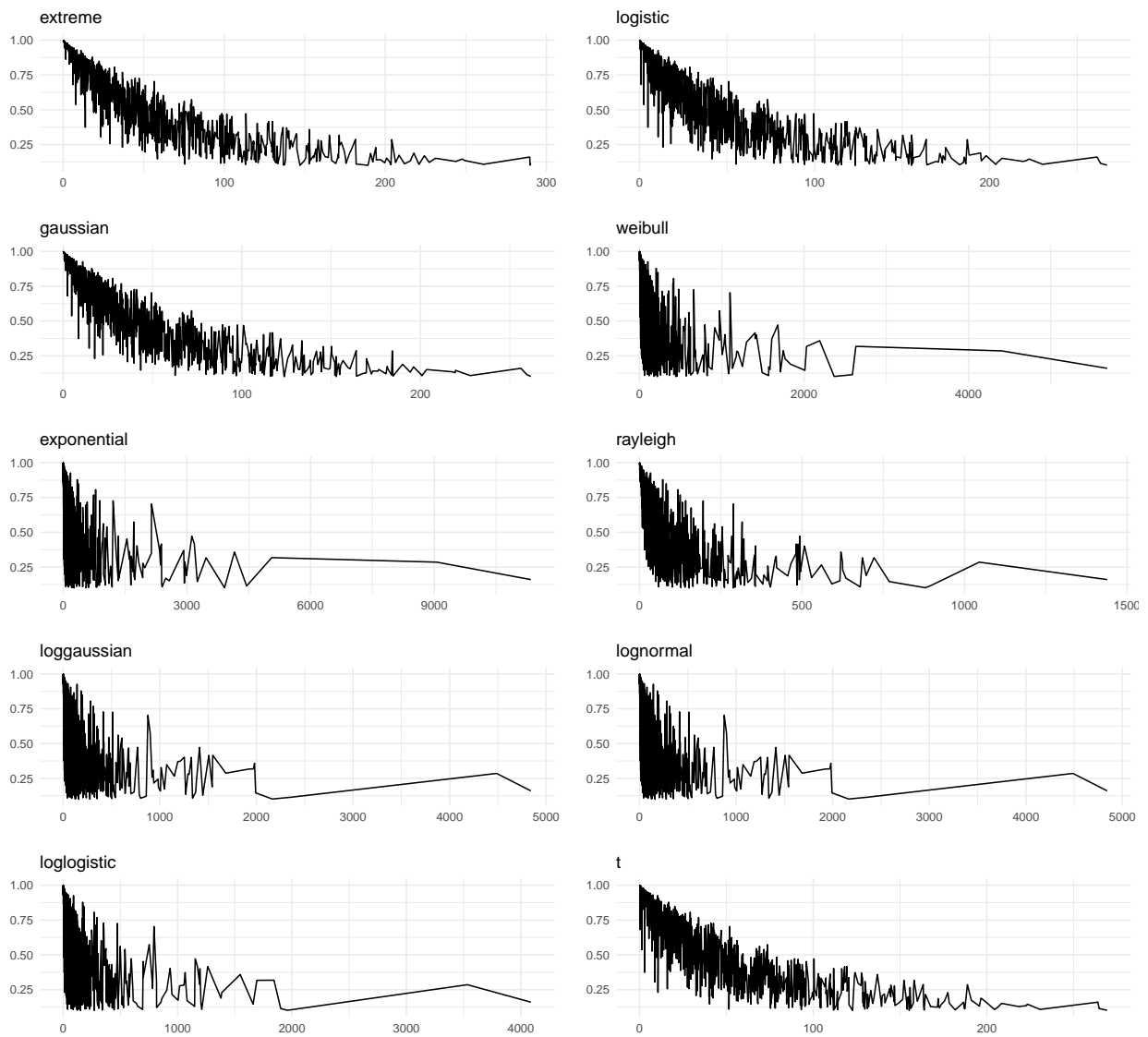


Survival Analysis

Gayane Ohanjanyan

2024-05-01

Probability of Survival for Different Distributions - AFT Model



Results of AIC and BIC statistical measures.

```
## Distribution with minimum AIC: loggaussian - 2950.3
```

```
## Distribution with minimum BIC: loggaussian - 3053.363
```

All the models have very small p values, which indicates that regardless of the chosen distributions, the models are statistically significant. Moreover, the models clearly highlight many statistically significant features, also there are some features which do not play huge role (have small p values) and can be dropped to have better and more accurate models. However, as we look closely we will find that the Loggaussian distribution have the smallest p value, and can be considered the best choice. In addition, the AIC and BIC statistical measures have used to find the most accurate model, and according to the results, Loggaussian performs the best.

AFT Model with Loggaussian distribution

```
##
## Call:
## survreg(formula = formula, data = telco, dist = "loggaussian")
##               Value Std. Error      z      p
## (Intercept)    2.53948    0.24325 10.44 < 2e-16
## age            0.03606    0.00640  5.63 1.8e-08
## maritalUnmarried -0.45526    0.11466 -3.97 7.2e-05
## address        0.04444    0.00886  5.02 5.2e-07
## internetYes    -0.88795    0.13814 -6.43 1.3e-10
## custcatE-service 0.99741    0.16918  5.90 3.7e-09
## custcatPlus service 0.73065    0.16546  4.42 1.0e-05
## custcatTotal service 0.65726    0.16595  3.96 7.5e-05
## Log(scale)      0.28755    0.04607  6.24 4.3e-10
##
## Scale= 1.33
##
## Log Normal distribution
## Loglik(model)= -1466   Loglik(intercept only)= -1602.5
##  Chisq= 273.09 on 7 degrees of freedom, p= 3.3e-55
## Number of Newton-Raphson Iterations: 5
## n= 1000
```

Interpretation of Coefficients:

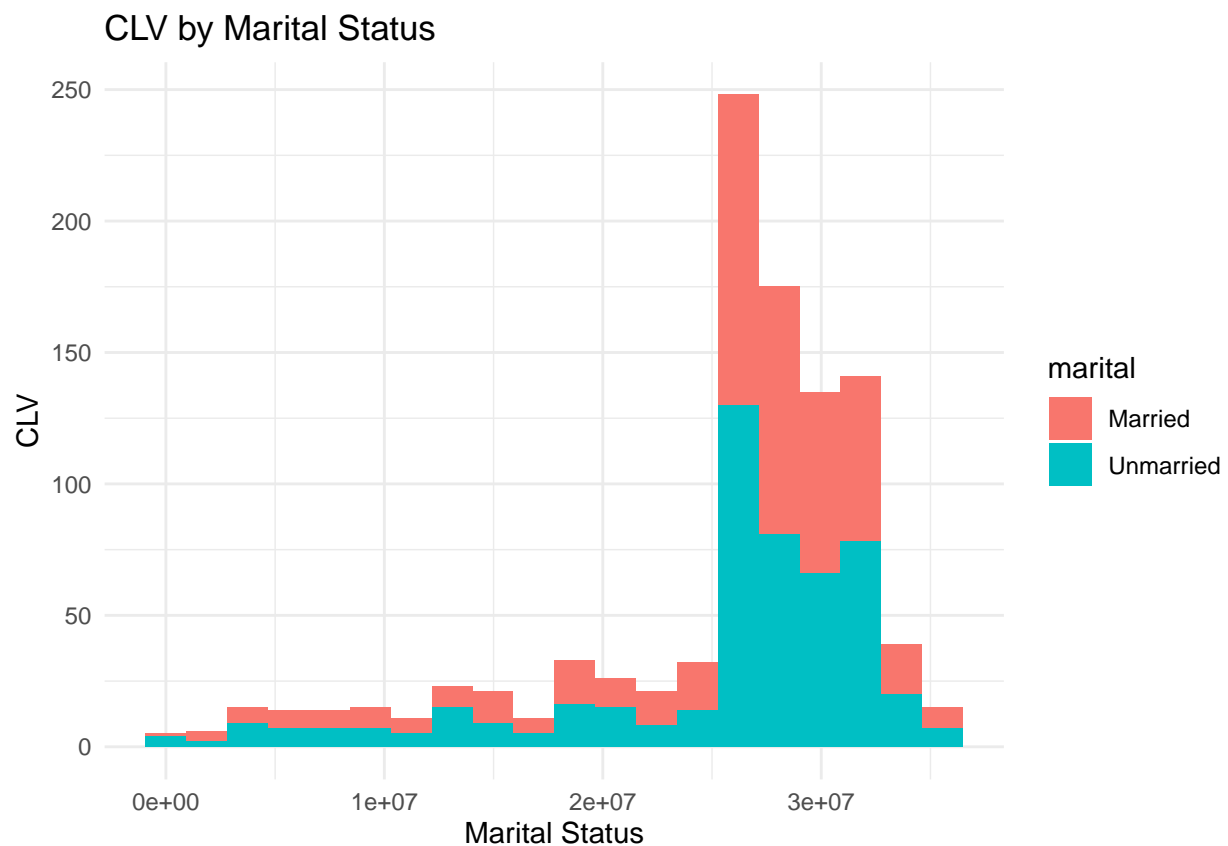
- Intercept: This represents the baseline value of the survival time when all other predictors are zero. In this case, it's 2.53948.
- Age: For each one-unit increase in age, the log of the survival time increases by 0.03606. This suggests that older individuals tend to have longer survival times.
- MaritalUnmarried: Being unmarried (compared to being married) is associated with a decrease in the log of survival time by 0.45526 units. This suggests that unmarried individuals may have shorter survival times compared to married individuals.
- Address: Each one-unit increase in the number of addresses is associated with an increase in the log of survival time by 0.04444 units.

- InternetYes: Having internet service (compared to not having internet service) is associated with a decrease in the log of survival time by 0.88795 units. This suggests that individuals with internet service may have shorter survival times.
- CustcatE-service, custcatPlus service, custcatTotal service: These coefficients represent the effect of different customer categories compared to a reference category on the log of survival time.

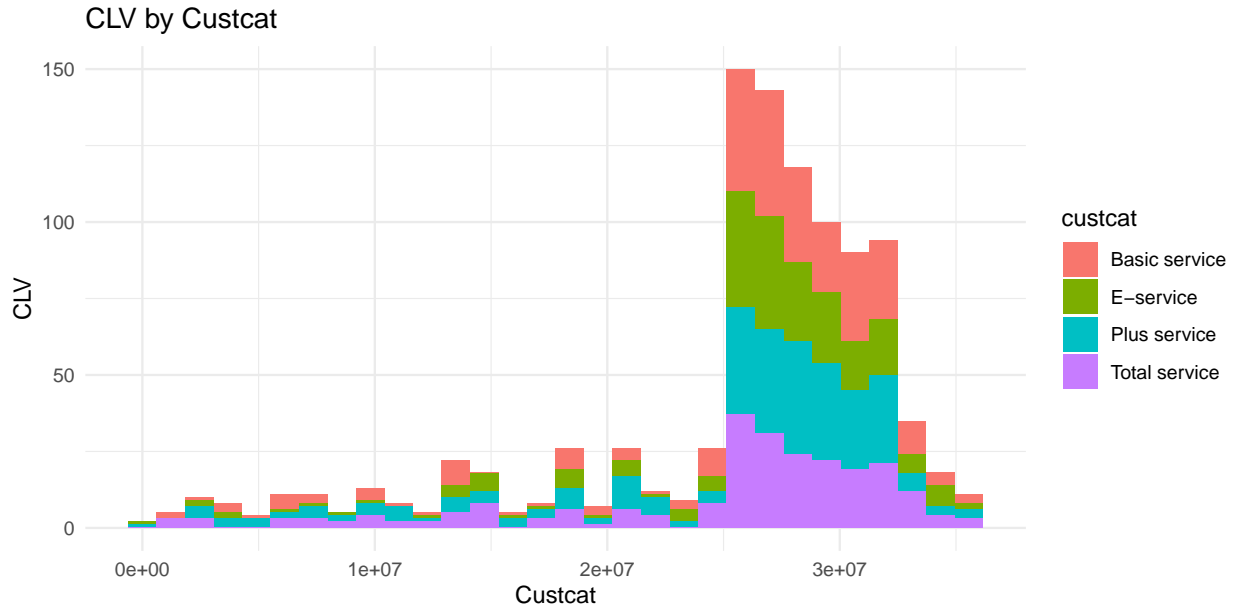
Model Fit:

The model's goodness-of-fit is assessed using likelihood-based statistics such as log-likelihood, chi-square, and p-values. In this case, the chi-square value is significant ($p = 3.3e-55$), indicating that the model significantly improves the fit compared to an intercept-only model.

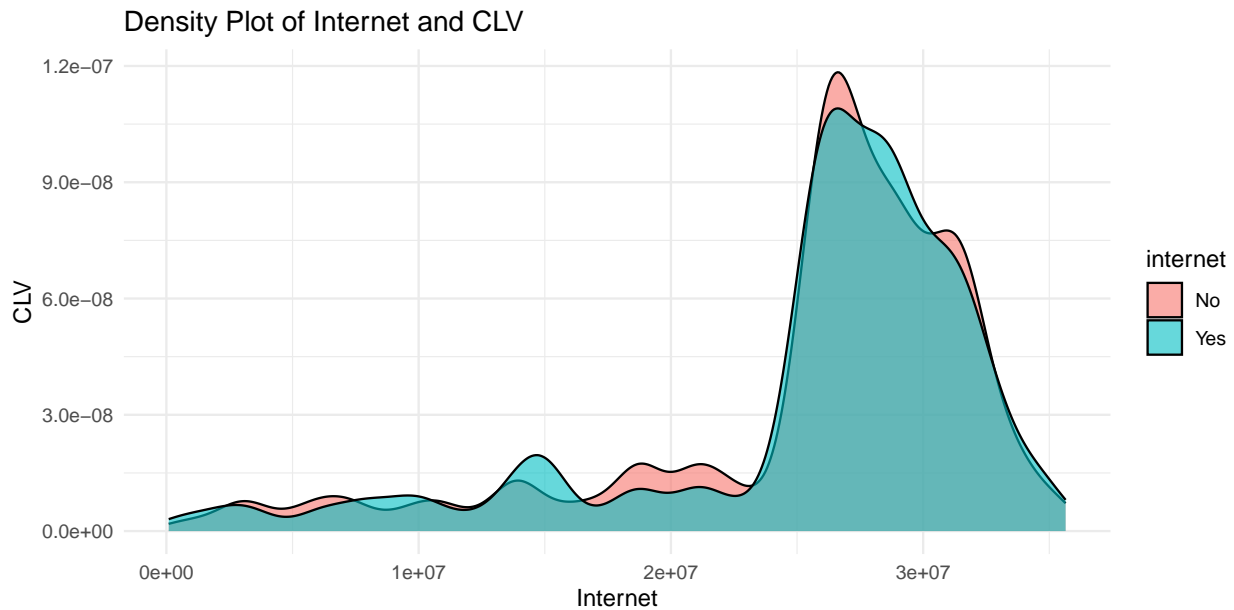
CLV Analysis



In the Telco dataset, the Customer Lifetime Value (CLV) tends to be higher for married individuals compared to unmarried individuals. This observation suggests that marital status may be a significant factor influencing CLV within the Telco customer base. A possible reason can be that married couples may be more likely to subscribe to bundled services such as family plans, which can increase their overall CLV.



In the CLV plot for different customer categories (Custcat) in the Telco dataset, it's evident that customers categorized under "Basic service" have the highest CLV, followed by those categorized under "E-service," "Plus service," and "Total service" in descending order. This ordering may suggest that Basic service subscribers may be enrolled in long-term contracts or higher-priced plans, resulting in higher CLV compared to other customer categories.



The density plot illustrates the distribution of Customer Lifetime Value (CLV) across different internet subscription statuses in the Telco dataset. While the CLV distributions for customers with and without internet subscriptions are largely similar, there appears to be a slightly higher density of CLV values for customers without internet subscriptions. This observation suggests that customers without internet subscriptions may have a slightly higher CLV compared to those with internet subscriptions in the Telco dataset.

Conclusion

Total CLV: 25521176086

Number of At-Risk Subscribers within a year: 412

Annual Retention Budget: 41200

Improving customer retention involves various strategies aimed at enhancing customer satisfaction, engagement, and loyalty. Here are some suggestions for retention strategies:

- Customer Education: Educate customers about the full range of your products or services and how they can derive maximum value from them. Offer tutorials, guides, or webinars to help customers make the most of their purchase.
- Personalized Communication: Tailor communication and offers based on individual customer preferences, behavior, and lifecycle stage. Use personalized emails, messages, or targeted advertising to make customers feel valued and understood.
- Community Building: Foster a sense of community among your customers by creating forums, social media groups, or online communities where they can connect, share experiences, and support each other.