Homework 3 | Survival Analysis

Professor: Karen Hovhannisyan

Student: Elen Petrosyan

Parametric Models

```
# Import libraries
import pandas as pd
from lifelines import WeibullAFTFitter, LogNormalAFTFitter,
LogLogisticAFTFitter
import matplotlib.pyplot as plt

df = pd.read_csv('telco.csv')

print(df.head())
```

```
   ID  region  tenure  age    marital  address  income  \
0   1  Zone 2      13   44    Married        9      64
1   2  Zone 3      11   33    Married        7     136
2   3  Zone 3      68   52    Married       24     116
3   4  Zone 2      33   33  Unmarried       12      33
4   5  Zone 2      23   30    Married        9      30

                             ed retire  gender voice internet forward
\
0                College degree     No    Male    No       No     Yes

1     Post-undergraduate degree     No    Male   Yes       No     Yes

2  Did not complete high school     No  Female    No       No      No

3             High school degree     No  Female    No       No      No

4  Did not complete high school     No    Male    No       No     Yes


         custcat churn
0  Basic service   Yes
1  Total service   Yes
2   Plus service    No
3  Basic service   Yes
4   Plus service    No
```

```
def new_data(input_data):
    processed_data = input_data.copy()

    # Drop 'ID' column
```

```
    processed_data.drop(['ID'], axis=1, inplace=True)

    # Convert 'churn' to categorical
    processed_data['churn'] =
processed_data['churn'].astype('category')

    # Identify categorical columns
    categorical_cols = ['region', 'retire', 'marital', 'ed', 'gender',
'voice', 'internet', 'custcat', 'forward']

    # Handle categorical columns using one-hot encoding
    processed_data = pd.get_dummies(processed_data,
columns=categorical_cols, drop_first=True)

    return processed_data

data = new_data(df)

data

     tenure  age  address  income churn  region_Zone 2  region_Zone 3
\
0        13   44        9      64   Yes              1              0

1        11   33        7     136   Yes              0              1

2        68   52       24     116    No              0              1

3        33   33       12      33   Yes              1              0

4        23   30        9      30    No              1              0

..      ...  ...      ...     ...   ...            ...            ...

995      10   39        0      27    No              0              1

996       7   34        2      22    No              0              0

997      67   59       40     944    No              0              1

998      70   49       18      87    No              0              1

999      50   36        7      39   Yes              0              1


     retire_Yes  marital_Unmarried  ed_Did not complete high school  \
0             0                  0                                0
1             0                  0                                0
2             0                  0                                1
3             0                  1                                0
4             0                  0                                1
```

```
..          ...                ...                                ...
995           0                  1                                  0
996           0                  1                                  0
997           0                  1                                  0
998           0                  1                                  0
999           0                  0                                  0

     ed_High school degree  ed_Post-undergraduate degree  ed_Some
college  \
0                        0                             0
0
1                        0                             1
0
2                        0                             0
0
3                        1                             0
0
4                        0                             0
0
..                     ...                           ...
...
995                      0                             0
1
996                      0                             1
0
997                      0                             1
0
998                      1                             0
0
999                      0                             0
1

     gender_Male  voice_Yes  internet_Yes  custcat_E-service  \
0              1          0             0                  0
1              1          1             0                  0
2              0          0             0                  0
3              0          0             0                  0
4              1          0             0                  0
..           ...        ...           ...                ...
995            0          0             0                  0
996            0          0             0                  0
997            0          1             1                  0
998            0          1             0                  0
999            0          0             1                  1

     custcat_Plus service  custcat_Total service  forward_Yes
0                       0                      0            1
1                       0                      1            1
2                       1                      0            0
3                       0                      0            0
```

```
4                        1                       0             1
..                     ...                     ...           ...
995                      0                       0             0
996                      0                       0             0
997                      0                       1             1
998                      1                       0             1
999                      0                       0             0

[1000 rows x 20 columns]

# mapping 'Yes' to 1 and 'No' to 0 in the 'churn' column
data['churn'] = data['churn'].map({'Yes': 1, 'No': 0})

def build_aft_models(data):
    distributions = ['weibull', 'lognormal', 'loglogistic']

    for distribution in distributions:
        if distribution == 'weibull':
            model = WeibullAFTFitter()
        elif distribution == 'lognormal':
            model = LogNormalAFTFitter()
        elif distribution == 'loglogistic':
            model = LogLogisticAFTFitter()

        # Fit the AFT model
        model.fit(data, duration_col='tenure', event_col='churn')

        # Print the summary of the model
        print(f"\nSummary for {distribution} distribution:")
        print(model.summary)

# Assuming 'data' is your processed dataset
build_aft_models(data)


Summary for weibull distribution:
                                             coef  exp(coef)  se(coef)
\
param   covariate

lambda_ address                          0.041363   1.042230  0.008821

        age                              0.027802   1.028192  0.006748

        custcat_E-service                0.977597   2.658060  0.155731

        custcat_Plus service             0.739767   2.095447  0.192965

        custcat_Total service            0.995856   2.707040  0.213172

        ed_Did not complete high school  0.437873   1.549409  0.194276
```

```
        ed_High school degree              0.319959   1.377071  0.145973
        ed_Post-undergraduate degree       0.223556   1.250516  0.190606
        ed_Some college                    0.253833   1.288957  0.144665
        forward_Yes                       -0.098678   0.906034  0.148202
        gender_Male                        0.004320   1.004329  0.103055
        income                             0.001035   1.001035  0.000926
        internet_Yes                      -0.773505   0.461393  0.138356
        marital_Unmarried                 -0.346694   0.707022  0.104357
        region_Zone 2                     -0.062115   0.939775  0.127965
        region_Zone 3                      0.115448   1.122376  0.127037
        retire_Yes                         0.170056   1.185371  0.522097
        voice_Yes                         -0.335197   0.715197  0.148444
        Intercept                          2.781115  16.137001  0.271337
rho_    Intercept                          0.174823   1.191035  0.051073

                                           coef lower 95%  coef upper
95%  \
param   covariate

lambda_ address                                  0.024074
0.058652
        age                                      0.014575
0.041028
        custcat_E-service                        0.672369
1.282824
        custcat_Plus service                     0.361562
1.117972
        custcat_Total service                    0.578046
1.413666
        ed_Did not complete high school          0.057099
0.818648
        ed_High school degree                    0.033856
0.606061
        ed_Post-undergraduate degree            -0.150024
0.597137
        ed_Some college                         -0.029706
```

```
0.537372
        forward_Yes                                   -0.389149
0.191793
        gender_Male                                   -0.197665
0.206305
        income                                        -0.000781
0.002850
        internet_Yes                                  -1.044678       -
0.502331
        marital_Unmarried                             -0.551230       -
0.142157
        region_Zone 2                                 -0.312921
0.188692
        region_Zone 3                                 -0.133539
0.364436
        retire_Yes                                    -0.853235
1.193348
        voice_Yes                                     -0.626143       -
0.044252
        Intercept                                      2.249305
3.312925
rho_    Intercept                                      0.074721
0.274924

                                               exp(coef) lower 95%  \
param   covariate
lambda_ address                                               1.024366
        age                                                   1.014682
        custcat_E-service                                     1.958872
        custcat_Plus service                                  1.435570
        custcat_Total service                                 1.782552
        ed_Did not complete high school                       1.058760
        ed_High school degree                                 1.034436
        ed_Post-undergraduate degree                          0.860687
        ed_Some college                                       0.970731
        forward_Yes                                           0.677633
        gender_Male                                           0.820644
        income                                                0.999219
        internet_Yes                                          0.351805
        marital_Unmarried                                     0.576240
        region_Zone 2                                         0.731308
        region_Zone 3                                         0.874993
        retire_Yes                                            0.426034
        voice_Yes                                             0.534650
        Intercept                                             9.481142
rho_    Intercept                                             1.077583

                                               exp(coef) upper 95%  cmp
to  \
```

```
param   covariate

lambda_ address                                  1.060407    0.0

        age                                      1.041881    0.0

        custcat_E-service                        3.606813    0.0

        custcat_Plus service                     3.058646    0.0

        custcat_Total service                    4.110997    0.0

        ed_Did not complete high school          2.267433    0.0

        ed_High school degree                    1.833196    0.0

        ed_Post-undergraduate degree             1.816909    0.0

        ed_Some college                          1.711503    0.0

        forward_Yes                              1.211420    0.0

        gender_Male                              1.229127    0.0

        income                                   1.002854    0.0

        internet_Yes                             0.605118    0.0

        marital_Unmarried                        0.867485    0.0

        region_Zone 2                            1.207669    0.0

        region_Zone 3                            1.439702    0.0

        retire_Yes                               3.298104    0.0

        voice_Yes                                0.956713    0.0

        Intercept                               27.465341    0.0

rho_    Intercept                                1.316431    0.0


                                         z              p   -
log2(p)
param   covariate

lambda_ address                   4.689016  2.745224e-06
18.474645
        age                       4.119853  3.791136e-05
14.687010
```

```
        custcat_E-service                    6.277452  3.441659e-10
31.436177
        custcat_Plus service                 3.833678  1.262413e-04
12.951529
        custcat_Total service                4.671604  2.988570e-06
18.352113
        ed_Did not complete high school      2.253869  2.420443e-02
5.368585
        ed_High school degree                2.191898  2.838689e-02
5.138632
        ed_Post-undergraduate degree         1.172871  2.408474e-01
2.053809
        ed_Some college                      1.754620  7.932427e-02
3.656094
        forward_Yes                         -0.665833  5.055180e-01
0.984166
        gender_Male                          0.041915  9.665667e-01
0.049059
        income                               1.116943  2.640186e-01
1.921288
        internet_Yes                        -5.590672  2.261928e-08
25.397872
        marital_Unmarried                   -3.322180  8.931700e-04
10.128778
        region_Zone 2                       -0.485404  6.273899e-01
0.672566
        region_Zone 3                        0.908778  3.634674e-01
1.460102
        retire_Yes                           0.325718  7.446380e-01
0.425389
        voice_Yes                           -2.258066  2.394152e-02
5.384342
        Intercept                           10.249684  1.187312e-24
79.478575
rho_    Intercept                            3.422980  6.193858e-04
10.656874

Summary for lognormal distribution:
                                            coef  exp(coef)  se(coef)
\
param  covariate

mu_    address                          0.042538   1.043456  0.008904

       age                              0.032670   1.033209  0.007254

       custcat_E-service                1.066401   2.904907  0.170532

       custcat_Plus service             0.924929   2.521689  0.215751
```

```
        custcat_Total service             1.198617   3.315528  0.250452
        ed_Did not complete high school  0.373624   1.452990  0.201587
        ed_High school degree             0.315938   1.371546  0.163183
        ed_Post-undergraduate degree    -0.034399   0.966186  0.223172
        ed_Some college                   0.272306   1.312989  0.165346
        forward_Yes                      -0.198139   0.820256  0.180037
        gender_Male                       0.051869   1.053238  0.114287
        income                            0.001396   1.001397  0.000920
        internet_Yes                     -0.771490   0.462324  0.143483
        marital_Unmarried                -0.455134   0.634363  0.115430
        region_Zone 2                    -0.097068   0.907494  0.142769
        region_Zone 3                     0.048258   1.049441  0.141538
        retire_Yes                        0.022532   1.022788  0.444076
        voice_Yes                        -0.433787   0.648050  0.168953
        Intercept                         2.362263  10.614950  0.292629
sigma_  Intercept                         0.275772   1.317548  0.045998

                                          coef lower 95%  coef upper 95%
\
param   covariate
mu_     address                                 0.025088        0.059989
        age                                     0.018452        0.046887
        custcat_E-service                       0.732165        1.400637
        custcat_Plus service                    0.502065        1.347793
        custcat_Total service                   0.707741        1.689493
        ed_Did not complete high school        -0.021480        0.768727
        ed_High school degree                  -0.003894        0.635771
        ed_Post-undergraduate degree           -0.471808        0.403009
```

```
       ed_Some college                            -0.051766        0.596379
       forward_Yes                                -0.551005        0.154728
       gender_Male                                -0.172130        0.275868
       income                                     -0.000408        0.003200
       internet_Yes                               -1.052710       -0.490269
       marital_Unmarried                          -0.681372       -0.228896
       region_Zone 2                              -0.376891        0.182755
       region_Zone 3                              -0.229152        0.325667
       retire_Yes                                 -0.847842        0.892906
       voice_Yes                                  -0.764929       -0.102646
       Intercept                                   1.788721        2.935805
sigma_ Intercept                                   0.185617        0.365928

                                             exp(coef) lower 95%  \
param  covariate
mu_    address                                              1.025405
       age                                                  1.018624
       custcat_E-service                                    2.079579
       custcat_Plus service                                 1.652129
       custcat_Total service                                2.029401
       ed_Did not complete high school                      0.978749
       ed_High school degree                                0.996113
       ed_Post-undergraduate degree                         0.623873
       ed_Some college                                      0.949551
       forward_Yes                                          0.576370
       gender_Male                                          0.841870
       income                                               0.999592
       internet_Yes                                         0.348991
       marital_Unmarried                                    0.505922
       region_Zone 2                                        0.685991
       region_Zone 3                                        0.795208
       retire_Yes                                           0.428338
       voice_Yes                                            0.465367
       Intercept                                            5.981799
sigma_ Intercept                                            1.203961

                                             exp(coef) upper 95%  cmp to
z  \
```

```
param  covariate

mu_    address                                              1.061825    0.0
4.777676
       age                                                  1.048004    0.0
4.503760
       custcat_E-service                                    4.057784    0.0
6.253394
       custcat_Plus service                                 3.848923    0.0
4.287021
       custcat_Total service                                5.416735    0.0
4.785820
       ed_Did not complete high school                      2.157019    0.0
1.853409
       ed_High school degree                                1.888478    0.0
1.936099
       ed_Post-undergraduate degree                         1.496321    0.0 -
0.154139
       ed_Some college                                      1.815533    0.0
1.646885
       forward_Yes                                          1.167340    0.0 -
1.100543
       gender_Male                                          1.317674    0.0
0.453847
       income                                               1.003205    0.0
1.517008
       internet_Yes                                         0.612462    0.0 -
5.376880
       marital_Unmarried                                    0.795411    0.0 -
3.942958
       region_Zone 2                                        1.200520    0.0 -
0.679895
       region_Zone 3                                        1.384954    0.0
0.340953
       retire_Yes                                           2.442217    0.0
0.050739
       voice_Yes                                            0.902447    0.0 -
2.567503
       Intercept                                           18.836666    0.0
8.072559
sigma_ Intercept                                            1.441851    0.0
5.995251

                                                            p   -log2(p)
param  covariate
mu_    address                                   1.773326e-06  19.105111
       age                                       6.676160e-06  17.192550
       custcat_E-service                         4.016273e-10  31.213423
       custcat_Plus service                      1.810853e-05  15.752971
```

```
        custcat_Total service             1.702906e-06  19.163569
        ed_Did not complete high school  6.382375e-02   3.969763
        ed_High school degree             5.285561e-02   4.241800
        ed_Post-undergraduate degree      8.775004e-01   0.188528
        ed_Some college                   9.958166e-02   3.327976
        forward_Yes                       2.710956e-01   1.883126
        gender_Male                       6.499389e-01   0.621624
        income                            1.292646e-01   2.951600
        internet_Yes                      7.578744e-08  23.653466
        marital_Unmarried                 8.048267e-05  13.600962
        region_Zone 2                     4.965707e-01   1.009929
        region_Zone 3                     7.331390e-01   0.447841
        retire_Yes                        9.595331e-01   0.059595
        voice_Yes                         1.024338e-02   6.609164
        Intercept                         6.883999e-16  50.367603
sigma_  Intercept                         2.031709e-09  28.874659

Summary for loglogistic distribution:
                                              coef  exp(coef)  se(coef)
\
param   covariate

alpha_  address                           0.038920   1.039687  0.008814

        age                               0.032441   1.032973  0.006942

        custcat_E-service                 1.040155   2.829656  0.165176

        custcat_Plus service              0.863530   2.371517  0.209202

        custcat_Total service             1.202978   3.330019  0.240610

        ed_Did not complete high school  0.434034   1.543471  0.198774

        ed_High school degree             0.335415   1.398521  0.154862

        ed_Post-undergraduate degree     -0.023335   0.976936  0.215783

        ed_Some college                   0.240953   1.272461  0.155660

        forward_Yes                      -0.194778   0.823018  0.170370

        gender_Male                       0.040035   1.040847  0.110132

        income                            0.001040   1.001041  0.000886

        internet_Yes                     -0.795452   0.451377  0.142620

        marital_Unmarried                -0.445134   0.640738  0.111082

        region_Zone 2                    -0.047518   0.953593  0.135294
```

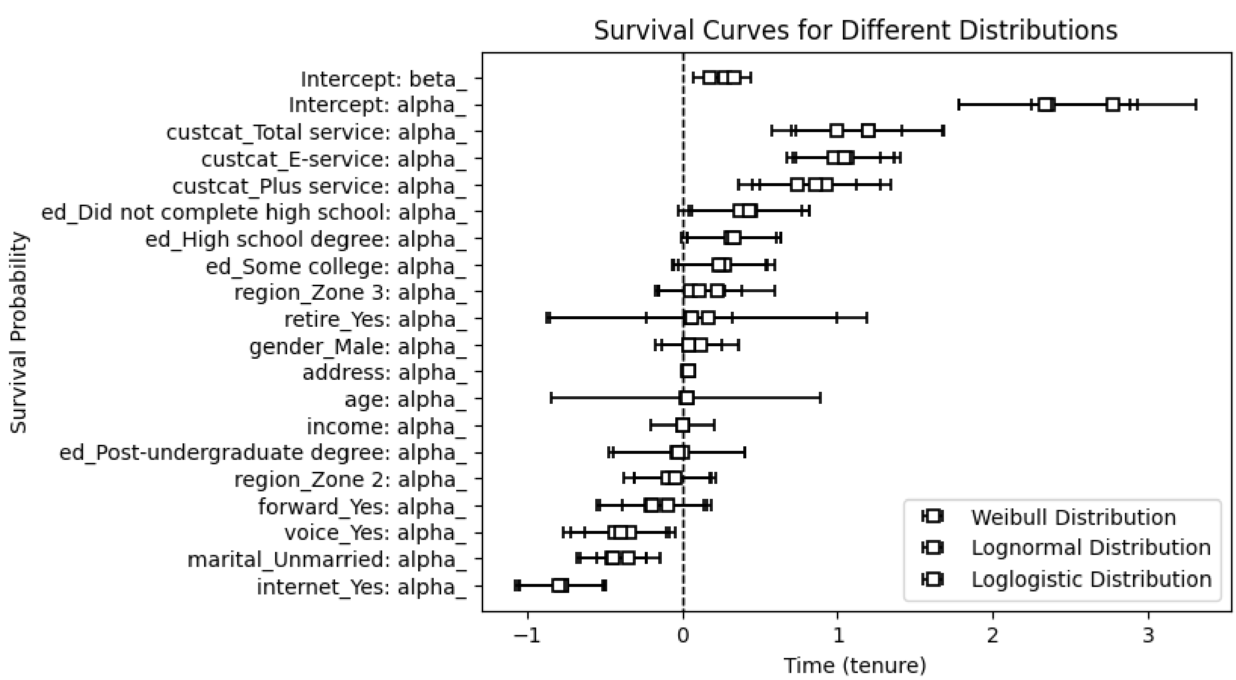
```
       region_Zone 3                       0.112707   1.119304  0.135959
       retire_Yes                          0.062707   1.064715  0.477516
       voice_Yes                          -0.398956   0.671020  0.162806
       Intercept                           2.334917  10.328607  0.281351
beta_  Intercept                           0.338380   1.402673  0.051017

                                           coef lower 95%  coef upper 95%
\
param  covariate
alpha_ address                                   0.021645        0.056195
       age                                       0.018834        0.046047
       custcat_E-service                         0.716416        1.363894
       custcat_Plus service                      0.453501        1.273559
       custcat_Total service                     0.731391        1.674565
       ed_Did not complete high school           0.044443        0.823624
       ed_High school degree                     0.031890        0.638940
       ed_Post-undergraduate degree             -0.446262        0.399593
       ed_Some college                          -0.064136        0.546042
       forward_Yes                              -0.528697        0.139142
       gender_Male                              -0.175819        0.255889
       income                                   -0.000696        0.002777
       internet_Yes                             -1.074982       -0.515921
       marital_Unmarried                        -0.662850       -0.227418
       region_Zone 2                            -0.312690        0.217653
       region_Zone 3                            -0.153767        0.379182
       retire_Yes                               -0.873208        0.998621
       voice_Yes                                -0.718049       -0.079863
```

```
       Intercept                                    1.783479        2.886356

beta_  Intercept                                    0.238389        0.438371


                                              exp(coef) lower 95%  \
param  covariate
alpha_ address                                                1.021881
       age                                                    1.019012
       custcat_E-service                                      2.047082
       custcat_Plus service                                   1.573812
       custcat_Total service                                  2.077968
       ed_Did not complete high school                        1.045445
       ed_High school degree                                  1.032404
       ed_Post-undergraduate degree                           0.640016
       ed_Some college                                        0.937878
       forward_Yes                                            0.589372
       gender_Male                                            0.838770
       income                                                 0.999304
       internet_Yes                                           0.341304
       marital_Unmarried                                      0.515380
       region_Zone 2                                          0.731477
       region_Zone 3                                          0.857471
       retire_Yes                                             0.417610
       voice_Yes                                              0.487703
       Intercept                                              5.950522
beta_  Intercept                                              1.269202

                                              exp(coef) upper 95%  cmp to
z  \
param  covariate

alpha_ address                                                1.057804     0.0
4.415717
       age                                                    1.047124     0.0
4.672846
       custcat_E-service                                      3.911396     0.0
6.297244
       custcat_Plus service                                   3.573549     0.0
4.127724
       custcat_Total service                                  5.336476     0.0
4.999696
       ed_Did not complete high school                        2.278743     0.0
2.183549
       ed_High school degree                                  1.894472     0.0
2.165891
       ed_Post-undergraduate degree                           1.491218     0.0 -
0.108139
       ed_Some college                                        1.726406     0.0
1.547940
```

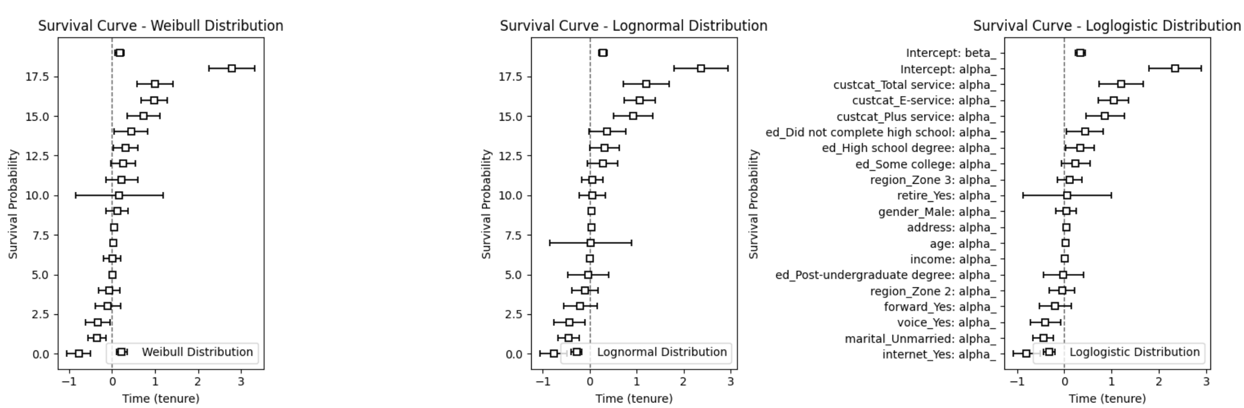
```
        forward_Yes                                      1.149287     0.0 -
1.143261
        gender_Male                                      1.291610     0.0
0.363519
        income                                           1.002781     0.0
1.174160
        internet_Yes                                     0.596950     0.0 -
5.577411
        marital_Unmarried                                0.796588     0.0 -
4.007270
        region_Zone 2                                    1.243156     0.0 -
0.351223
        region_Zone 3                                    1.461089     0.0
0.828980
        retire_Yes                                       2.714537     0.0
0.131319
        voice_Yes                                        0.923243     0.0 -
2.450504
        Intercept                                       17.927860     0.0
8.298938
beta_   Intercept                                        1.550179     0.0
6.632712

                                                    p   -log2(p)
param  covariate
alpha_  address                          1.006755e-05  16.599928
        age                              2.970545e-06  18.360841
        custcat_E-service                3.029835e-10  31.620042
        custcat_Plus service             3.663721e-05  14.736331
        custcat_Total service            5.742067e-07  20.731926
        ed_Did not complete high school  2.899543e-02   5.108031
        ed_High school degree            3.031951e-02   5.043610
        ed_Post-undergraduate degree     9.138856e-01   0.129914
        ed_Some college                  1.216368e-01   3.039349
        forward_Yes                      2.529303e-01   1.983188
        gender_Male                      7.162169e-01   0.481531
        income                           2.403309e-01   2.056906
        internet_Yes                     2.441248e-08  25.287806
        marital_Unmarried                6.142457e-05  13.990825
        region_Zone 2                    7.254207e-01   0.463110
        region_Zone 3                    4.071156e-01   1.296489
        retire_Yes                       8.955233e-01   0.159197
        voice_Yes                        1.426566e-02   6.131310
        Intercept                        1.050459e-16  53.079829
beta_   Intercept                        3.295739e-11  34.820607
```

```python
def visualize_survival_curves(data):
    distributions = ['weibull', 'lognormal', 'loglogistic']

    #creating subplots
```

```
    fig, ax = plt.subplots()

    for distribution in distributions:
        if distribution == 'weibull':
            model = WeibullAFTFitter()
        elif distribution == 'lognormal':
            model = LogNormalAFTFitter()
        elif distribution == 'loglogistic':
            model = LogLogisticAFTFitter()

        # Fitting the AFT model
        model.fit(data, duration_col='tenure', event_col='churn')

        # Ploting the survival function
        model.plot(ax=ax, label=f"{distribution.capitalize()}
Distribution")

    # Set plot labels and legend
    ax.set_xlabel("Time (tenure)")
    ax.set_ylabel("Survival Probability")
    ax.set_title("Survival Curves for Different Distributions")
    ax.legend()

    plt.show()

visualize_survival_curves(data)
```



```
def visualize_survival_curves(data):
    distributions = ['weibull', 'lognormal', 'loglogistic']

    # separate subplots for each distribution
    fig, axes = plt.subplots(1, len(distributions), figsize=(15, 5))

    for i, distribution in enumerate(distributions):
        if distribution == 'weibull':
            model = WeibullAFTFitter()
        elif distribution == 'lognormal':
            model = LogNormalAFTFitter()
        elif distribution == 'loglogistic':
            model = LogLogisticAFTFitter()

        # Fitting the AFT model
        model.fit(data, duration_col='tenure', event_col='churn')

        # Ploting the survival function on the i-th subplot
        model.plot(ax=axes[i], label=f"{distribution.capitalize()}
Distribution")

        # plot labels and title for each subplot
        axes[i].set_xlabel("Time (tenure)")
        axes[i].set_ylabel("Survival Probability")
        axes[i].set_title(f"Survival Curve -
{distribution.capitalize()} Distribution")
        axes[i].legend()

    plt.tight_layout()

    plt.show()

visualize_survival_curves(data)
```



CLV

```
significant_columns = ["address", "age", "internet_Yes",
"marital_Unmarried", "tenure", "churn",
```

```
                       "custcat_E-service", "custcat_Plus service",
"custcat_Total service", "voice_Yes"]

new_data = data[significant_columns]
new_data
```

```
     address  age  internet_Yes  marital_Unmarried  tenure churn  \
0          9   44             0                  0      13     1
1          7   33             0                  0      11     1
2         24   52             0                  0      68     0
3         12   33             0                  1      33     1
4          9   30             0                  0      23     0
..       ...  ...           ...                ...     ...   ...
995        0   39             0                  1      10     0
996        2   34             0                  1       7     0
997       40   59             1                  1      67     0
998       18   49             0                  1      70     0
999        7   36             1                  0      50     1

     custcat_E-service  custcat_Plus service  custcat_Total service
voice_Yes
0                    0                     0                      0
0
1                    0                     0                      1
1
2                    0                     1                      0
0
3                    0                     0                      0
0
4                    0                     1                      0
0
..                 ...                   ...                    ...
...
995                  0                     0                      0
0
996                  0                     0                      0
0
997                  0                     0                      1
1
998                  0                     1                      0
1
999                  1                     0                      0
0

[1000 rows x 10 columns]
```

```
new_data.columns
```

```
Index(['address', 'age', 'internet_Yes', 'marital_Unmarried',
'tenure',
       'churn', 'custcat_E-service', 'custcat_Plus service',
       'custcat_Total service', 'voice_Yes'],
      dtype='object')
```

```
# Monthly revenue per customer
monthly_revenue_per_customer = 100


discount_rate = 0.1

# CLV for each segment
def calculate_clv(segment_data):
    segment_data['CLV'] = (monthly_revenue_per_customer * (1 -
discount_rate) * segment_data['tenure']) / discount_rate
    return segment_data

# CLV for the entire dataset
all_data = new_data.copy()
all_data = calculate_clv(all_data)

# List of columns
columns_for_clv = ['address', 'age', 'internet_Yes',
'marital_Unmarried', 'tenure',
                   'churn', 'custcat_E-service', 'custcat_Plus
service',
                   'custcat_Total service', 'voice_Yes']

# CLV grouped by each specified column
for column in columns_for_clv:
    clv_by_column = all_data.groupby(column)['CLV'].mean()
    print(f"\nCLV Results for {column}:")
    print(clv_by_column)
```

```
CLV Results for address:
address
0     20201.785714
1     21917.647059
2     23686.363636
3     21496.721311
4     23459.016393
5     26298.000000
6     24250.000000
7     23654.716981
8     29792.307692
9     25770.731707
10    35455.263158
11    31147.826087
```

```
12    33325.000000
13    31036.363636
14    35475.000000
15    37612.500000
16    33235.714286
17    38372.727273
18    45540.000000
19    39568.965517
20    40545.000000
21    46800.000000
22    44520.000000
23    46320.000000
24    48225.000000
25    44925.000000
26    45825.000000
27    47418.750000
28    48075.000000
29    53100.000000
30    51136.363636
31    50550.000000
32    48780.000000
33    51840.000000
34    51750.000000
35    49050.000000
36    58500.000000
37    60900.000000
38    51480.000000
39    61200.000000
40    62100.000000
41    64800.000000
42    52200.000000
43    53700.000000
44    47700.000000
45    64800.000000
46    58500.000000
48    64800.000000
49    64800.000000
55    60300.000000
Name: CLV, dtype: float64

CLV Results for age:
age
18     1800.000000
19    12600.000000
20    10890.000000
21    10462.500000
22    20700.000000
23    19125.000000
24    16290.000000
```

```
25    25904.347826
26    16457.142857
27    23775.000000
28    22622.727273
29    26194.736842
30    21600.000000
31    20081.250000
32    25757.142857
33    23515.384615
34    24300.000000
35    24432.352941
36    23688.000000
37    33822.580645
38    27557.142857
39    32631.428571
40    31982.142857
41    28425.000000
42    35067.857143
43    29512.500000
44    29340.000000
45    39750.000000
46    37012.500000
47    36814.285714
48    37950.000000
49    35900.000000
50    36978.260870
51    36900.000000
52    45072.000000
53    48789.473684
54    37912.500000
55    40671.428571
56    39845.454545
57    45180.000000
58    45720.000000
59    46992.857143
60    56700.000000
61    51054.545455
62    40900.000000
63    44250.000000
64    45450.000000
65    46500.000000
66    52714.285714
67    54000.000000
68    49500.000000
69    34800.000000
70    58200.000000
71    34200.000000
72    58500.000000
73    54900.000000
```

```
74    27000.000000
75    64800.000000
76    48000.000000
77    60300.000000
Name: CLV, dtype: float64

CLV Results for internet_Yes:
internet_Yes
0    34039.082278
1    28425.815217
Name: CLV, dtype: float64

CLV Results for marital_Unmarried:
marital_Unmarried
0    34965.454545
1    29040.594059
Name: CLV, dtype: float64

CLV Results for tenure:
tenure
1       900.0
2      1800.0
3      2700.0
4      3600.0
5      4500.0
        ...
68    61200.0
69    62100.0
70    63000.0
71    63900.0
72    64800.0
Name: CLV, Length: 72, dtype: float64

CLV Results for churn:
churn
0    36421.487603
1    20187.591241
Name: CLV, dtype: float64

CLV Results for custcat_E-service:
custcat_E-service
0    30022.988506
1    39011.059908
Name: CLV, dtype: float64

CLV Results for custcat_Plus service:
custcat_Plus service
0    30372.183588
1    36070.462633
Name: CLV, dtype: float64
```

```
CLV Results for custcat_Total service:
custcat_Total service
0    32079.581152
1    31629.661017
Name: CLV, dtype: float64

CLV Results for voice_Yes:
voice_Yes
0    32044.396552
1    31810.855263
Name: CLV, dtype: float64
```

The parametric models using Weibull, LogNormal, and LogLogistic distributions were applied to investigate factors influencing churn risk in a telco dataset. For instance, positive coefficients suggest an increased risk of churn, while negative coefficients indicate a reduced risk. In exploring valuable segments, CLV was calculated for various factors such as 'address,' 'age,' ... These CLV outcomes offer a monetary perspective on different customer segments to identify high-value customer groups.

Assuming the dataset represents the entire population, estimating the annual retention budget involves a holistic assessment of CLV, survival probabilities, and identifying at-risk subscribers within a year. This entails a thorough examination of customer tenure, churn probabilities, and corresponding CLV to facilitate efficient resource allocation. To enhance retention strategies, it is recommended to delve deeper into variables significantly impacting churn risk and implement targeted initiatives such as personalized communication strategies, customer surveys, and addressing identified pain points for a more effective customer retention approach. Visualizing time-dependent churn probabilities through survival curves for various distributions provides additional insights into customer behavior, contributing to a more nuanced understanding of the retention landscape.

Implementing customer retention strategies and overall enhancing retention through can be done by personalized communication, proactive support, loyalty programs, and continuous data analysis. Optimize onboarding, foster community engagement, and analyze exit interviews for a comprehensive approach.