DAR F21 Project Status

DeFi

Soumya Mishra

10/28/2021

Contents


```r
#load Rds (binary version of csv file) into dataframe
# Assumes this notebook is in: ~/IDEA-Blockchain/DefiResearch/StudentNotebooks/Assignment02
df<-read_rds('../../Data/transactions.Rds')

# Let's take a quick look at the first few observation
head(df)
```

```
##       amount borrowRate borrowRateMode   onBehalfOf         pool reserve
## 1   41501.63   6.274937       Variable 8.502518e+47 1.034668e+48     DAI
## 2 7000000.00   2.589628       Variable 4.635974e+47 1.034668e+48    USDT
## 3   15000.00   8.802541       Variable 3.735263e+47 1.034668e+48    USDC
## 4    8193.19  48.747052         Stable 6.896232e+47 1.034668e+48    USDC
## 5   11000.00   3.225055       Variable 1.089455e+48 1.034668e+48    USDT
## 6   40000.00   5.739208       Variable 2.178337e+47 1.034668e+48    USDT
##    timestamp         user   type reservePriceETH reservePriceUSD  amountUSD
## 1 1621340435 8.502518e+47 borrow    2.852900e+14       0.9948044   41286.00
## 2 1622477822 4.635974e+47 borrow    3.812835e+14       1.0000000 7000000.00
## 3 1619775984 3.735263e+47 borrow    3.611000e+14       1.0043389   15065.08
## 4 1615481632 6.896232e+47 borrow    5.562201e+14       0.9993909    8188.20
## 5 1626914745 1.089455e+48 borrow    4.971100e+14       1.0000000   11000.00
## 6 1620936688 2.178337e+47 borrow    2.725248e+14       1.0000000   40000.00
##   collateralAmount collateralReserve principalAmount principalReserve
## 1               NA                                NA
## 2               NA                                NA
## 3               NA                                NA
## 4               NA                                NA
## 5               NA                                NA
## 6               NA                                NA
##   reservePriceETHPrincipal reservePriceUSDPrincipal reservePriceETHCollateral
## 1                       NA                       NA                        NA
## 2                       NA                       NA                        NA
## 3                       NA                       NA                        NA
## 4                       NA                       NA                        NA
## 5                       NA                       NA                        NA
## 6                       NA                       NA                        NA
##   reservePriceUSDCollateral amountUSDPincipal amountUSDCollateral
```

```
## 1                        NA                 NA                   NA
## 2                        NA                 NA                   NA
## 3                        NA                 NA                   NA
## 4                        NA                 NA                   NA
## 5                        NA                 NA                   NA
## 6                        NA                 NA                   NA
##   borrowRateModeFrom borrowRateModeTo stableBorrowRate variableBorrowRate
## 1                                                   NA                 NA
## 2                                                   NA                 NA
## 3                                                   NA                 NA
## 4                                                   NA                 NA
## 5                                                   NA                 NA
## 6                                                   NA                 NA
```

#Analysis of Borrows to Deposits:

```
borrows <- df %>%
  filter(type=="borrow")

deposits <- df %>%
  filter(type=="deposit")

depositBorrow <- left_join(deposits,borrows,by="user") %>%
  dplyr::rename(depositTime=timestamp.x) %>%
  dplyr::rename(borrowTime=timestamp.y) %>%
  group_by(user) %>%
  dplyr::summarise(timeDiff=case_when(min(borrowTime)-min(depositTime)>0 ~ min(borrowTime)-min(depositTi
  mutate(status=case_when(timeDiff==as.integer(21294796) ~ 0, timeDiff<=0 ~ 0, timeDiff>0 ~ 1)) %>%
  select(user,timeDiff,status)
```

```
km <- with(depositBorrow, Surv(timeDiff/86400, status))
head(km,80)
```

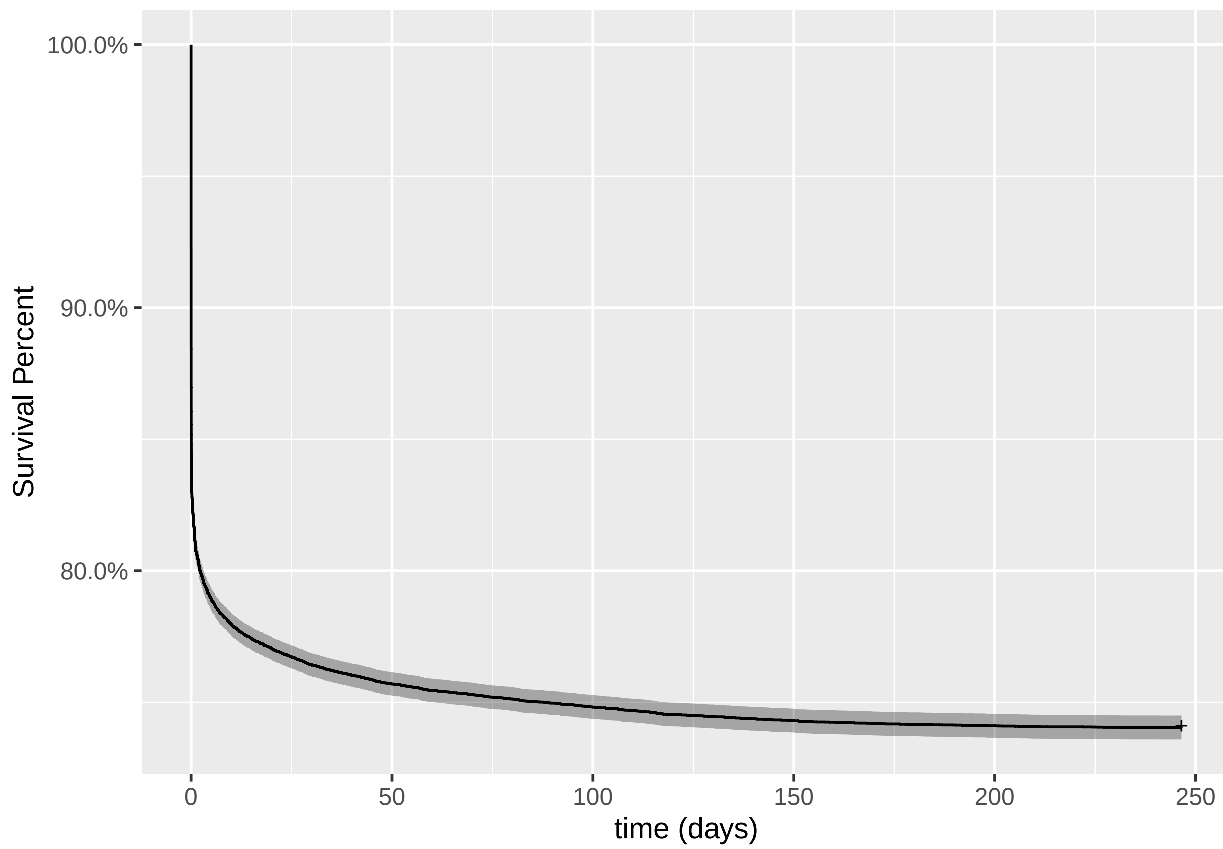
```
##  [1] 2.464675e+02+ 2.604167e-03  2.464675e+02+ 2.464675e+02+ 2.464675e+02+
##  [6] 2.464675e+02+ 2.464675e+02+ 2.436806e-01  2.464675e+02+ 2.464675e+02+
## [11] 3.503472e-02  2.464675e+02+ 2.464675e+02+ 2.464675e+02+ 2.464675e+02+
## [16] 2.464675e+02+ 8.360292e+01  2.464675e+02+ 2.464675e+02+ 2.464675e+02+
## [21] 2.464675e+02+ 7.664352e-01  2.464675e+02+ 9.375000e-04  2.464675e+02+
## [26] 6.759259e-03  2.464675e+02+ 2.464675e+02+ 6.670035e+00  2.464675e+02+
## [31] 2.464675e+02+ 2.464675e+02+ 2.464675e+02+ 2.464675e+02+ 2.464675e+02+
## [36] 2.464675e+02+ 5.081019e-03  2.464675e+02+ 2.464675e+02+ 7.222222e-03
## [41] 2.464675e+02+ 2.464675e+02+ 2.234954e-02  2.464675e+02+ 2.464675e+02+
## [46] 2.464675e+02+ 2.464675e+02+ 2.464675e+02+ 6.595439e+01  2.464675e+02+
## [51] 2.464675e+02+ 2.662037e-03  2.464675e+02+ 2.464675e+02+ 2.464675e+02+
## [56] 2.464675e+02+ 6.025463e-02  2.464675e+02+ 2.464675e+02+ 2.410880e-02
## [61] 2.464675e+02+ 2.464675e+02+ 2.464675e+02+ 2.464675e+02+ 2.464675e+02+
## [66] 2.464675e+02+ 2.464675e+02+ 2.464675e+02+ 1.844907e-02  2.464675e+02+
## [71] 2.464675e+02+ 2.418981e-02  2.464675e+02+ 2.464675e+02+ 1.003472e-02
## [76] 2.464675e+02+ 1.509082e+02  2.464675e+02+ 2.464675e+02+ 2.464675e+02+
```

```
km_fit <- survfit(Surv(timeDiff/86400, status) ~ 1, data=depositBorrow)
summary(km_fit, times = c(1,30,60,90*(1:10)))
```

```
## Call: survfit(formula = Surv(timeDiff/86400, status) ~ 1, data = depositBorrow)
##
##  time n.risk n.event survival std.err lower 95% CI upper 95% CI
```

```
##     1  28949     6750     0.811 0.00207         0.807         0.815
##    30  27285     1664     0.764 0.00225         0.760         0.769
##    60  26935      350     0.755 0.00228         0.750         0.759
##    90  26764      171     0.750 0.00229         0.745         0.754
##   180  26476      288     0.742 0.00232         0.737         0.746
```

```r
autoplot(km_fit,xlab="time (days)",ylab="Survival Percent",title="Survival Analysis of Borrow to Deposi
```



Deposits are the steepest of the survival analysis graphs.

#Analysis of Borrows to Repays:

```r
borrows <- df %>%
  filter(type=="borrow")

repays <- df %>%
  filter(type=="repay")

borrowRepay <- left_join(repays,borrows,by="user") %>%
  dplyr::rename(repayTime=timestamp.x) %>%
  dplyr::rename(borrowTime=timestamp.y) %>%
  group_by(user) %>%
  dplyr::summarise(timeDiff=case_when(min(borrowTime)-min(repayTime)>0 ~ min(borrowTime)-min(repayTime)
  mutate(status=case_when(timeDiff==as.integer(21294796) ~ 0, timeDiff<=0 ~ 0, timeDiff>0 ~ 1)) %>%
  select(user,timeDiff,status)
```

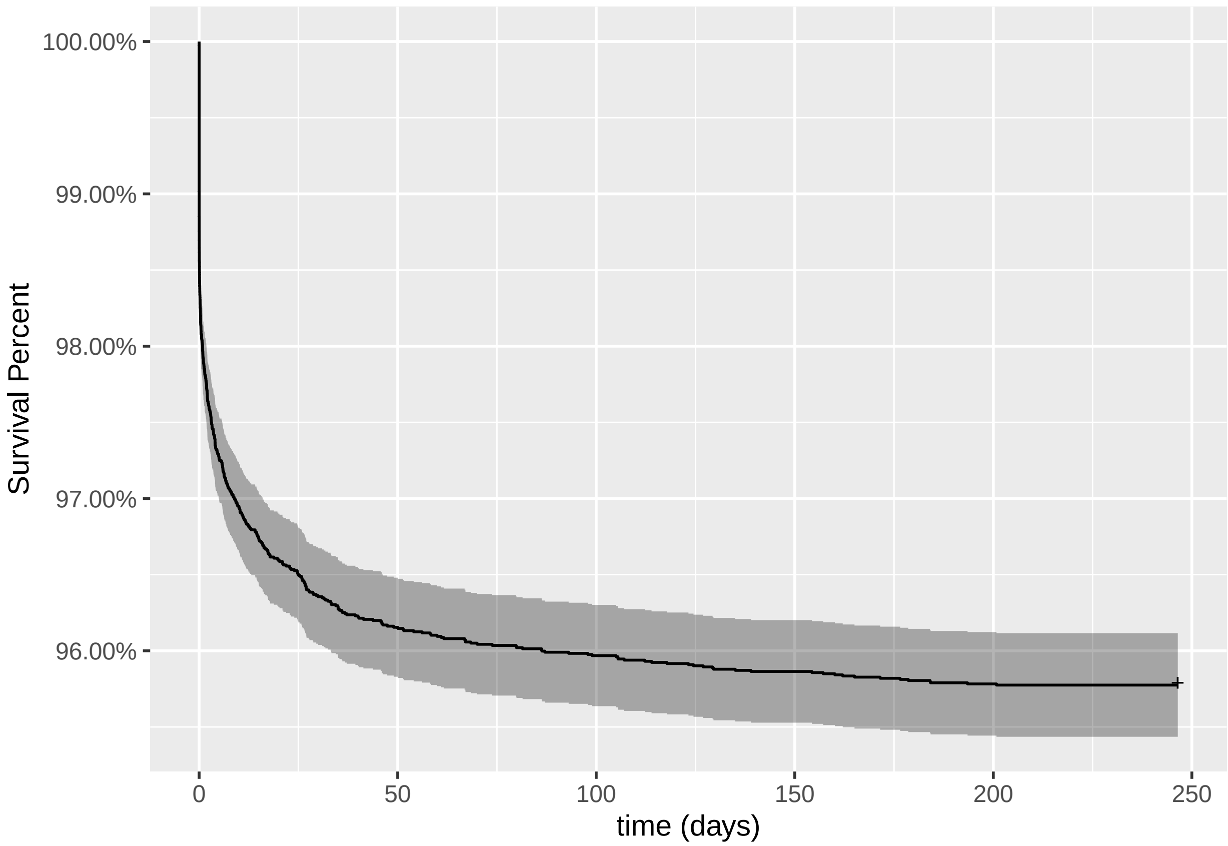
```r
km <- with(borrowRepay, Surv(timeDiff/86400, status))
head(km,80)
```

```
##  [1] 2.464675e+02+ 2.464675e+02+ 2.464675e+02+ 2.464675e+02+ 2.464675e+02+
##  [6] 2.464675e+02+ 2.464675e+02+ 2.464675e+02+ 2.464675e+02+ 2.464675e+02+
## [11] 2.464675e+02+ 2.464675e+02+ 3.638160e+00  2.464675e+02+ 2.464675e+02+
## [16] 2.464675e+02+ 2.464675e+02+ 2.464675e+02+ 2.464675e+02+ 2.847222e-03
## [21] 2.464675e+02+ 2.464675e+02+ 2.464675e+02+ 2.464675e+02+ 2.464675e+02+
## [26] 2.464675e+02+ 2.464675e+02+ 2.464675e+02+ 2.464675e+02+ 2.464675e+02+
## [31] 2.464675e+02+ 2.464675e+02+ 2.652416e+01  2.464675e+02+ 2.464675e+02+
## [36] 2.464675e+02+ 2.464675e+02+ 2.464675e+02+ 2.464675e+02+ 2.464675e+02+
## [41] 2.464675e+02+ 2.464675e+02+ 7.717257e+00  2.464675e+02+ 2.464675e+02+
## [46] 2.464675e+02+ 2.464675e+02+ 2.464675e+02+ 2.464675e+02+ 2.464675e+02+
## [51] 2.464675e+02+ 2.464675e+02+ 2.464675e+02+ 2.464675e+02+ 2.464675e+02+
## [56] 2.464675e+02+ 2.464675e+02+ 2.464675e+02+ 2.464675e+02+ 2.464675e+02+
## [61] 2.464675e+02+ 2.464675e+02+ 2.464675e+02+ 2.464675e+02+ 2.464675e+02+
## [66] 2.464675e+02+ 2.464675e+02+ 2.464675e+02+ 2.464675e+02+ 2.464675e+02+
## [71] 2.464675e+02+ 2.464675e+02+ 2.464675e+02+ 2.464675e+02+ 2.464675e+02+
## [76] 2.464675e+02+ 2.464675e+02+ 2.464675e+02+ 2.464675e+02+ 2.464675e+02+
```

```
km_fit <- survfit(Surv(timeDiff/86400, status) ~ 1, data=borrowRepay)
summary(km_fit, times = c(1,30,60,90*(1:10)))
```

```
## Call: survfit(formula = Surv(timeDiff/86400, status) ~ 1, data = borrowRepay)
##
##  time n.risk n.event survival std.err lower 95% CI upper 95% CI
##     1  13165     279    0.979 0.00123        0.977        0.982
##    30  12954     211    0.964 0.00162        0.960        0.967
##    60  12919      35    0.961 0.00167        0.958        0.964
##    90  12905      14    0.960 0.00169        0.957        0.963
##   180  12880      25    0.958 0.00173        0.955        0.961
```

```
autoplot(km_fit,xlab="time (days)",ylab="Survival Percent",title="Survival Analysis of Borrow to Repay")
```



Repayments occur at a much steeper curve before plateauing.

#Analysis of Deposits to Redeems

```
dep <- df %>%
  filter(type=="deposit")%>%
  arrange(user)

deposits<-dep[1:50000,]

red <- df %>%
  filter(type=="redeem")%>%
  arrange(user)

redeems<-red[1:50000,]

depositRedeem <- left_join(redeems,deposits,by="user") %>%
  dplyr::rename(depositTime=timestamp.x) %>%
  dplyr::rename(redeemTime=timestamp.y) %>%
  group_by(user) %>%
  dplyr::summarise(timeDiff=case_when(min(redeemTime)-min(depositTime)>0 ~ min(depositTime)-min(redeemTi
  mutate(status=case_when(timeDiff==as.integer(21294796) ~ 0, timeDiff<=0 ~ 0, timeDiff>0 ~ 1)) %>%
  select(user,timeDiff,status)
```

```
km <- with(depositRedeem, Surv(timeDiff/86400, status))
head(km,80)
```

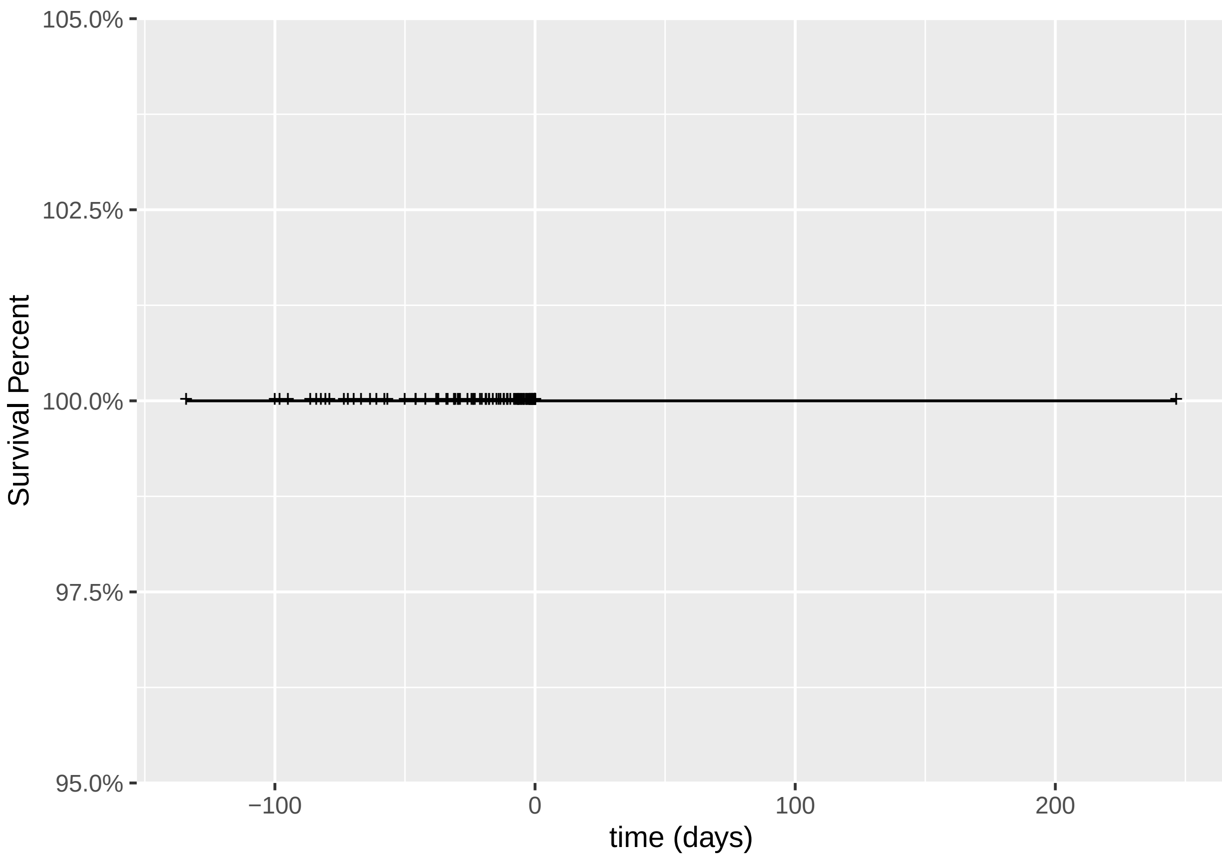
```
##  [1] -31.08388+ 246.46755+ 246.46755+ 246.46755+ 246.46755+ 246.46755+
```

```
##  [7] -25.97152+ 246.46755+ 246.46755+ 246.46755+ 246.46755+ 246.46755+
## [13] 246.46755+ 246.46755+ 246.46755+ 246.46755+ 246.46755+ 246.46755+
## [19] 246.46755+ 246.46755+ 246.46755+ 246.46755+ 246.46755+ 246.46755+
## [25] 246.46755+ 246.46755+ 246.46755+ 246.46755+ 246.46755+ 246.46755+
## [31] 246.46755+ 246.46755+ 246.46755+ 246.46755+ 246.46755+ 246.46755+
## [37] 246.46755+ 246.46755+ 246.46755+ 246.46755+ 246.46755+ 246.46755+
## [43] 246.46755+ 246.46755+ 246.46755+ 246.46755+ 246.46755+ 246.46755+
## [49] 246.46755+ 246.46755+ 246.46755+ 246.46755+ 246.46755+ 246.46755+
## [55] 246.46755+ 246.46755+ 246.46755+ 246.46755+ 246.46755+ 246.46755+
## [61] 246.46755+ 246.46755+ 246.46755+ 246.46755+ 246.46755+ 246.46755+
## [67] 246.46755+ 246.46755+ 246.46755+ 246.46755+ 246.46755+ 246.46755+
## [73] -86.43015+ 246.46755+ 246.46755+ 246.46755+ 246.46755+ 246.46755+
## [79] 246.46755+ 246.46755+
```

```r
km_fit <- survfit(Surv(timeDiff/86400, status) ~ 1, data=depositRedeem)
summary(km_fit, times = c(1,30,60,90*(1:10)))
```

```
## Call: survfit(formula = Surv(timeDiff/86400, status) ~ 1, data = depositRedeem)
##
##  time n.risk n.event survival std.err lower 95% CI upper 95% CI
##     1   8962       0        1       0            1            1
##    30   8962       0        1       0            1            1
##    60   8962       0        1       0            1            1
##    90   8962       0        1       0            1            1
##   180   8962       0        1       0            1            1
```

```r
autoplot(km_fit,xlab="time (days)",ylab="Survival Percent",title="Survival Analysis of Deposit to Redeem
```



It seems redeeems and deposits do not function anywhere similar to how borrowing does.

#Analysis of Borrows to Redeems

```r
borrows <- df %>%
  filter(type=="borrow")

redeems <- df %>%
  filter(type=="redeem")

borrowRedeem <- left_join(redeems,borrows,by="user") %>%
  dplyr::rename(redeemTime=timestamp.x) %>%
  dplyr::rename(borrowTime=timestamp.y) %>%
  group_by(user) %>%
  dplyr::summarise(timeDiff=case_when(min(borrowTime)-min(redeemTime)>0 ~ min(borrowTime)-min(redeemTim
  mutate(status=case_when(timeDiff==as.integer(21294796) ~ 0, timeDiff<=0 ~ 0, timeDiff>0 ~ 1)) %>%
  select(user,timeDiff,status)
```

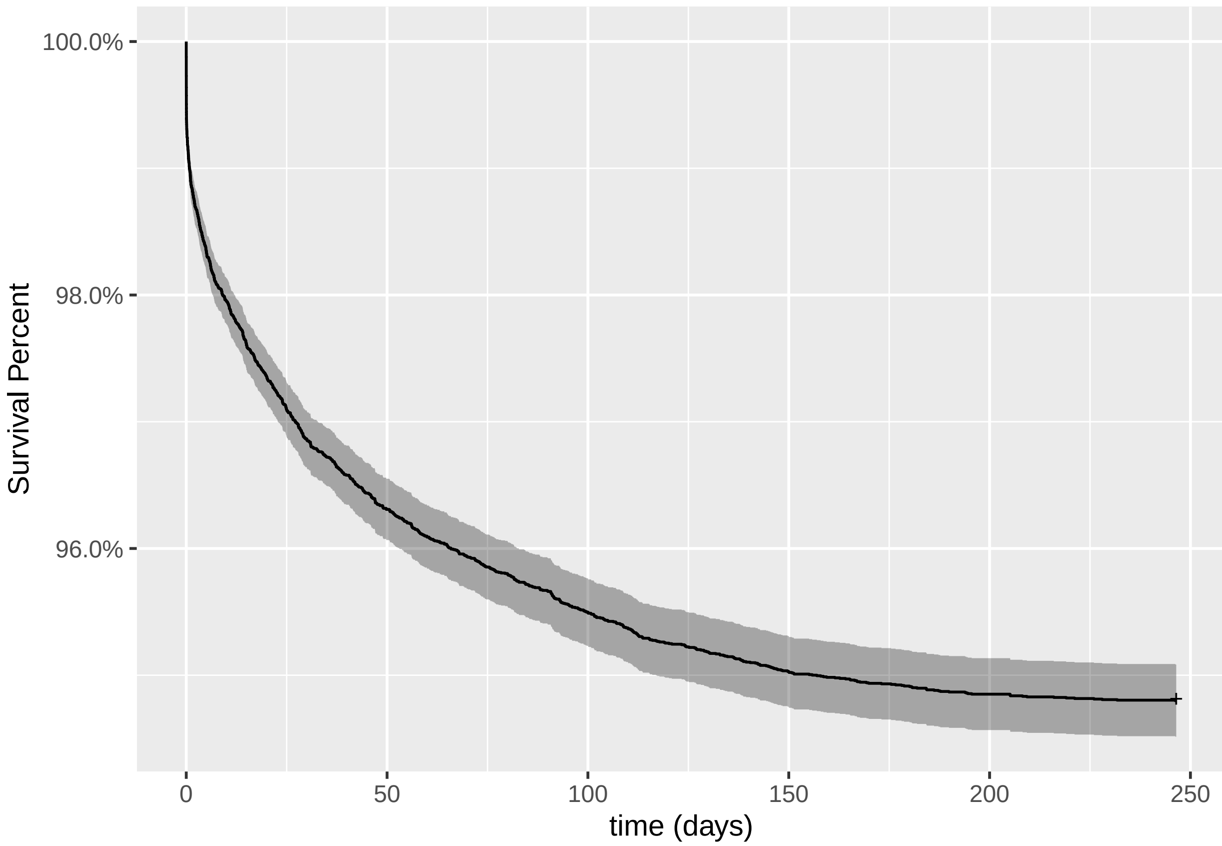
```r
km <- with(borrowRedeem, Surv(timeDiff/86400, status))
head(km,80)
```

```
##  [1] 246.4675463+ 246.4675463+ 246.4675463+ 246.4675463+ 246.4675463+
##  [6] 246.4675463+ 246.4675463+ 246.4675463+ 246.4675463+ 246.4675463+
## [11] 246.4675463+  81.5437731  246.4675463+ 246.4675463+ 246.4675463+
## [16] 246.4675463+ 246.4675463+   3.8297454  246.4675463+ 246.4675463+
## [21] 246.4675463+ 246.4675463+ 246.4675463+ 246.4675463+ 246.4675463+
## [26] 246.4675463+ 246.4675463+  54.9925116  246.4675463+ 246.4675463+
## [31] 246.4675463+ 246.4675463+ 246.4675463+ 246.4675463+ 246.4675463+
## [36] 246.4675463+ 246.4675463+ 246.4675463+ 246.4675463+ 246.4675463+
## [41] 246.4675463+ 111.2771181  246.4675463+ 246.4675463+ 246.4675463+
## [46] 246.4675463+ 246.4675463+  36.4634491  246.4675463+ 246.4675463+
## [51] 246.4675463+ 246.4675463+ 246.4675463+ 246.4675463+   0.1929282
## [56] 246.4675463+ 246.4675463+ 246.4675463+ 246.4675463+ 246.4675463+
## [61] 246.4675463+ 246.4675463+ 246.4675463+ 246.4675463+ 246.4675463+
## [66] 246.4675463+ 246.4675463+ 246.4675463+ 246.4675463+ 246.4675463+
## [71] 246.4675463+ 246.4675463+ 246.4675463+ 246.4675463+ 246.4675463+
## [76] 246.4675463+  36.4045949  246.4675463+ 246.4675463+ 246.4675463+
```

```r
km_fit <- survfit(Surv(timeDiff/86400, status) ~ 1, data=borrowRedeem)
summary(km_fit, times = c(1,30,60,90*(1:10)))
```

```
## Call: survfit(formula = Surv(timeDiff/86400, status) ~ 1, data = borrowRedeem)
##
##  time n.risk n.event survival  std.err lower 95% CI upper 95% CI
##     1  23076     246    0.989 0.000669        0.988        0.991
##    30  22590     486    0.969 0.001142        0.966        0.971
##    60  22410     180    0.961 0.001269        0.958        0.963
##    90  22311      99    0.957 0.001333        0.954        0.959
##   180  22136     175    0.949 0.001439        0.946        0.952
```

```r
autoplot(km_fit,xlab="time (days)",ylab="Survival Percent",title="Survival Analysis of Borrows to Redee
```



There is not as steep of a curve here which reflects redeeming an amount from the pool borrowed is slower.

#Analysis of borrows to liquidation

```
borrows <- df %>%
  filter(type=="borrow")

liquidations <- df %>%
  filter(type=="liquidation")

borrowLiquidation <- left_join(liquidations,borrows,by="user") %>%
  dplyr::rename(liquidationTime=timestamp.x) %>%
  dplyr::rename(borrowTime=timestamp.y) %>%
  group_by(user) %>%
  dplyr::summarise(timeDiff=case_when(min(borrowTime)-min(liquidationTime)>0 ~ min(borrowTime)-min(liqu
  mutate(status=case_when(timeDiff==as.integer(21294796) ~ 0, timeDiff<=0 ~ 0, timeDiff>0 ~ 1)) %>%
  select(user,timeDiff,status)
```

```
km <- with(borrowLiquidation, Surv(timeDiff/86400, status))
head(km,80)
```

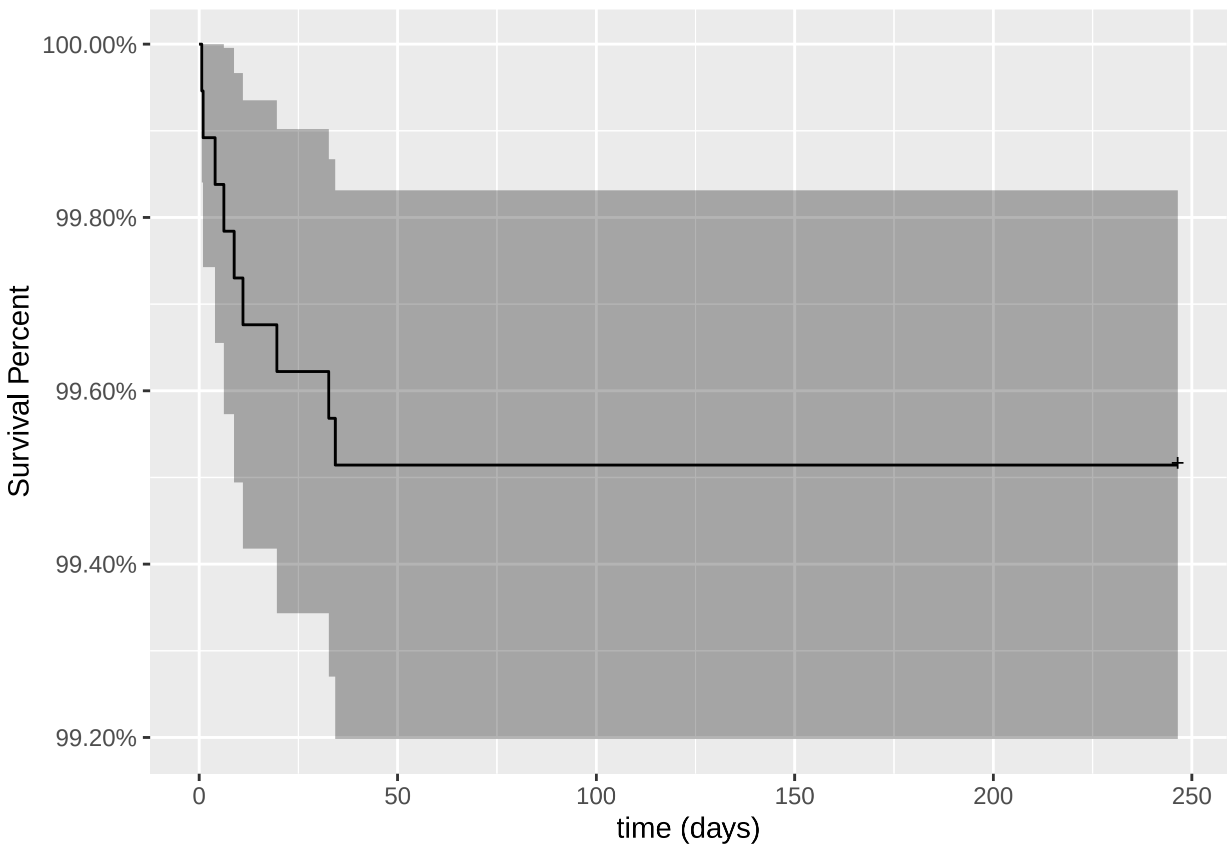
```
##  [1] 246.46755+ 246.46755+ 246.46755+ 246.46755+   4.00912  246.46755+
##  [7] 246.46755+ 246.46755+ 246.46755+ 246.46755+ 246.46755+ 246.46755+
## [13] 246.46755+ 246.46755+ 246.46755+ 246.46755+ 246.46755+ 246.46755+
## [19] 246.46755+ 246.46755+ 246.46755+ 246.46755+ 246.46755+ 246.46755+
## [25] 246.46755+ 246.46755+ 246.46755+ 246.46755+ 246.46755+ 246.46755+
## [31] 246.46755+ 246.46755+ 246.46755+ 246.46755+ 246.46755+ 246.46755+
## [37] 246.46755+ 246.46755+ 246.46755+ 246.46755+ 246.46755+ 246.46755+
```

```
## [43] 246.46755+ 246.46755+ 246.46755+ 246.46755+ 246.46755+ 246.46755+
## [49] 246.46755+ 246.46755+ 246.46755+ 246.46755+ 246.46755+ 246.46755+
## [55] 246.46755+ 246.46755+ 246.46755+ 246.46755+ 246.46755+ 246.46755+
## [61] 246.46755+ 246.46755+ 246.46755+ 246.46755+ 246.46755+ 246.46755+
## [67] 246.46755+ 246.46755+ 246.46755+ 246.46755+ 246.46755+ 246.46755+
## [73] 246.46755+ 246.46755+ 246.46755+ 246.46755+ 246.46755+ 246.46755+
## [79] 246.46755+ 246.46755+
```

```r
km_fit <- survfit(Surv(timeDiff/86400, status) ~ 1, data=borrowLiquidation)
summary(km_fit, times = c(1,30,60,90*(1:10)))
```

```
## Call: survfit(formula = Surv(timeDiff/86400, status) ~ 1, data = borrowLiquidation)
##
##  time n.risk n.event survival  std.err lower 95% CI upper 95% CI
##     1   1851       2    0.999 0.000763        0.997        1.000
##    30   1846       5    0.996 0.001425        0.993        0.999
##    60   1844       2    0.995 0.001615        0.992        0.998
##    90   1844       0    0.995 0.001615        0.992        0.998
##   180   1844       0    0.995 0.001615        0.992        0.998
```

```r
autoplot(km_fit,xlab="time (days)",ylab="Survival Percent",title="Survival Analysis of Borrows to Liqui
```



This graph is by far the most jagged and I believe it is due to the lack of data for liquidation. I previously analyzed only 1-2% of the data ended in liquidation, thus a less smooth curve here. However I believe since liquidation is automatically not survival, the following graph is a more accurate without the survival package. For the upcoming week I may try to adjust the parameters for the survival percentage when it comes to liquidation.

```r
#load liquidation data
liquidation.df<-df%>% filter(type=="liquidation")%>%
  select(user, timestamp)
head(liquidation.df)
```

```
##           user  timestamp
## 1 2.976865e+47 1626124715
## 2 3.748214e+47 1619145033
## 3 1.130833e+48 1621319875
## 4 9.560356e+45 1614324006
## 5 6.451374e+45 1621788289
## 6 1.460589e+48 1621429473
```

```r
#load borrows
borrow.df<-df%>% filter(type=="borrow")%>%
  select(onBehalfOf, timestamp)%>%
  rename(user = onBehalfOf)
head(borrow.df)
```

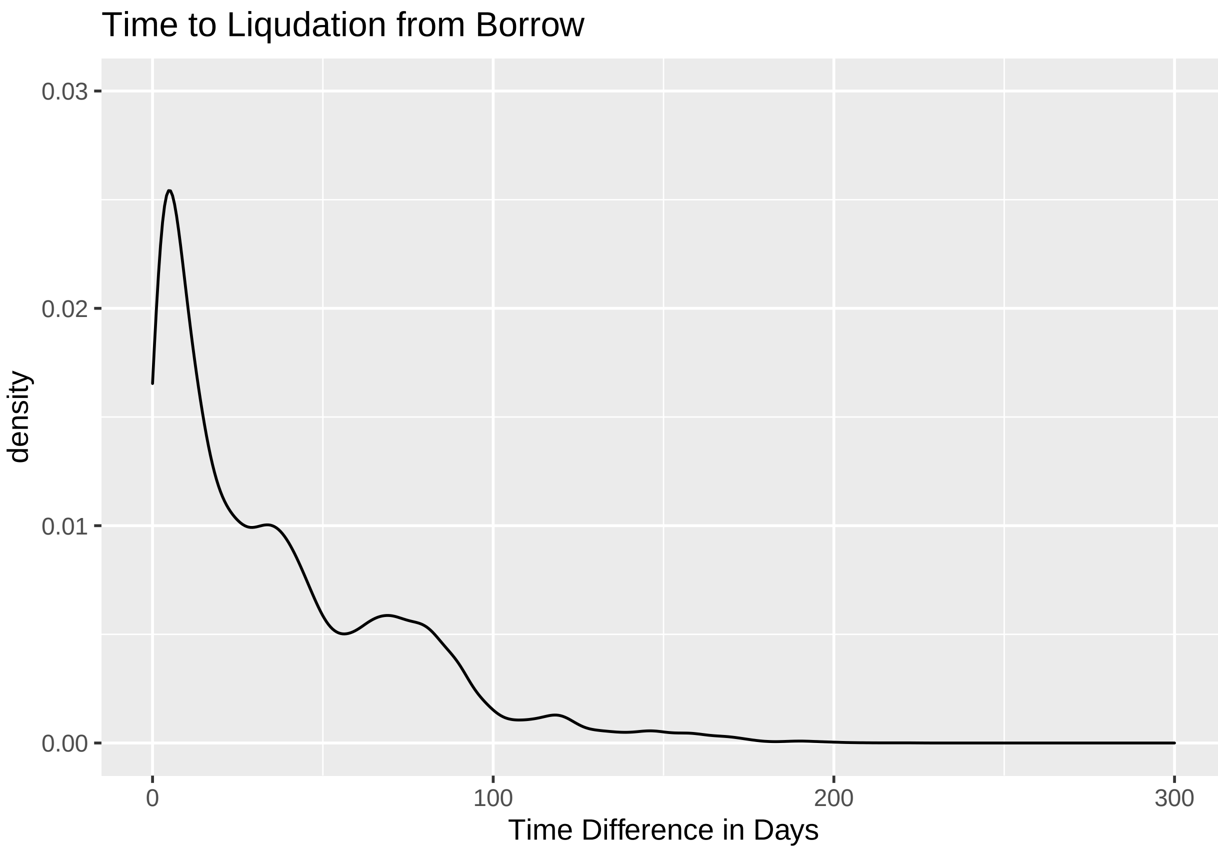
```
##           user  timestamp
## 1 8.502518e+47 1621340435
## 2 4.635974e+47 1622477822
## 3 3.735263e+47 1619775984
## 4 6.896232e+47 1615481632
## 5 1.089455e+48 1626914745
## 6 2.178337e+47 1620936688
```

```r
#join table
liqTable <- left_join(borrow.df,liquidation.df,by="user")%>%
  arrange(user)%>%
  rename(borrowTime=timestamp.x)%>%
  rename(liquidationTime = timestamp.y)%>%
  mutate(timeDiff = borrowTime-liquidationTime)%>%
  filter(timeDiff>0)

head(liqTable)
```

```
##           user borrowTime liquidationTime timeDiff
## 1 1.325103e+44 1623286150      1621468415  1817735
## 2 1.325103e+44 1623286150      1621398995  1887155
## 3 1.325103e+44 1623286150      1621426665  1859485
## 4 1.325103e+44 1625426662      1621468415  3958247
## 5 1.325103e+44 1625426662      1621398995  4027667
## 6 1.325103e+44 1625426662      1621426665  3999997
```

```r
# Basic density for liquidation
p <- ggplot(liqTable, aes(x=timeDiff/86400)) +
  xlim(0,300) +
  ylim(0,0.03)+
  xlab("Time Difference in Days")+
  geom_density()+
  ggtitle("Time to Liqudation from Borrow")
p
```



Weekly Work Summary

- RCS ID: mishrs4
- Project Name: Defi
- Summary of work since last week

https://docs.google.com/presentation/d/1YSJbGJxOD4ZigHVGgeTTbZ-rMI33HEqbT3aeKIGfgO4/edit#slide=id.p

Personal Contribution

I completed survival analysis of the various transaction types from borrowing from the Defi pool. I previously was not censoring the data and trying to do survival analysis through simple plotting of comparing the data by user overlap. I had several issues with installing the survival analysis package, but was finally able to get it to run and generate the following graphs.

Discussion of Primary Findings

The main question I was looking to answer is "How well does DeFi replicate traditional finance? How do the transaction types differ?

I found liquidation was most likely to occur the fastest, then deposits, redeemption, and then repayment of a borrow.

This differs from my previous week's results using my simple join of the users rather than the survival package.

Something I found of interest that I don't completely understand yet, though plan on figuring out is why deposits and redeems don't function as borrowing does. I attempted to plot the entire set of deposits and redeems, to reach a max limit of space. So instead I truncated the data set to only include 5000 rows and see if I could work with a smaller set but still look at a decent number of users. The graph is significantly different than those of borrowing, which may be inherent to the nature of Defi where a deposit and redeem functions much less like a loan and more a simple transaction which seems to be reflected in the lack of a downward curve in the analysis of the data.

Currently I have analyzed how the time it takes to repay an amount borrowed looks over time, the time it takes to redeem an amount deposited looks over time, the time it takes to deposit an amount deposited looks over time, and the time it takes to reach liquidation. In the graphs there is a lot more of a spread for redeeming than repaying or depositing.