Seasonality Analysis

Haya Naviwala

5/12/2022

R Markdown

```
library(tidyverse)
```

```
## Registered S3 methods overwritten by 'tibble':
##   method     from
##   format.tbl pillar
##   print.tbl  pillar
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.0 --
```

```
## v ggplot2 3.3.5     v purrr   0.3.3
## v tibble  2.1.3     v dplyr   1.0.7
## v tidyr   1.0.0     v stringr 1.4.0
## v readr   1.3.1     v forcats 0.4.0
```

```
## Warning: package 'ggplot2' was built under R version 3.6.2
```

```
## Warning: package 'dplyr' was built under R version 3.6.2
```

```
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
dt0 <- read_csv("telcoData.csv")
```

```
## Parsed with column specification:
## cols(
##   .default = col_character(),
##   SeniorCitizen = col_double(),
##   tenure = col_double(),
##   MonthlyCharges = col_double(),
##   TotalCharges = col_double()
## )
```

```
## See spec(...) for full column specifications.
```

```
glimpse(dt0)
```

```
## Observations: 7,043
## Variables: 21
## $ customerID       <chr> "7590-VHVEG", "5575-GNVDE", "3668-QPYBK", "77...
## $ gender           <chr> "Female", "Male", "Male", "Male", "Female", "...
## $ SeniorCitizen    <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ...
## $ Partner          <chr> "Yes", "No", "No", "No", "No", "No", "No", "N...
## $ Dependents       <chr> "No", "No", "No", "No", "No", "No", "Yes", "N...
## $ tenure           <dbl> 1, 34, 2, 45, 2, 8, 22, 10, 28, 62, 13, 16, 5...
## $ PhoneService     <chr> "No", "Yes", "Yes", "No", "Yes", "Yes", "Yes"...
## $ MultipleLines    <chr> "No phone service", "No", "No", "No phone ser...
## $ InternetService  <chr> "DSL", "DSL", "DSL", "DSL", "Fiber optic", "F...
## $ OnlineSecurity   <chr> "No", "Yes", "Yes", "Yes", "No", "No", "No", ...
```

```
## $ OnlineBackup     <chr> "Yes", "No", "Yes", "No", "No", "No", "Yes", ...
## $ DeviceProtection <chr> "No", "Yes", "No", "Yes", "No", "Yes", "No", ...
## $ TechSupport      <chr> "No", "No", "No", "Yes", "No", "No", "No", "N...
## $ StreamingTV      <chr> "No", "No", "No", "No", "No", "Yes", "Yes", "...
## $ StreamingMovies  <chr> "No", "No", "No", "No", "No", "Yes", "No", "N...
## $ Contract         <chr> "Month-to-month", "One year", "Month-to-month...
## $ PaperlessBilling <chr> "Yes", "No", "Yes", "No", "Yes", "Yes", "Yes"...
## $ PaymentMethod    <chr> "Electronic check", "Mailed check", "Mailed c...
## $ MonthlyCharges   <dbl> 29.85, 56.95, 53.85, 42.30, 70.70, 99.65, 89....
## $ TotalCharges     <dbl> 29.85, 1889.50, 108.15, 1840.75, 151.65, 820....
## $ Churn            <chr> "No", "No", "Yes", "No", "Yes", "Yes", "No", ...
```

```
names(dt0)
```

```
##  [1] "customerID"       "gender"           "SeniorCitizen"
##  [4] "Partner"          "Dependents"       "tenure"
##  [7] "PhoneService"     "MultipleLines"    "InternetService"
## [10] "OnlineSecurity"   "OnlineBackup"     "DeviceProtection"
## [13] "TechSupport"      "StreamingTV"      "StreamingMovies"
## [16] "Contract"         "PaperlessBilling" "PaymentMethod"
## [19] "MonthlyCharges"   "TotalCharges"     "Churn"
```

```
#replacing yes w churn and no w nochurn
dt1 <- dt0 %>%
  mutate(SeniorCitizen =
           ifelse(SeniorCitizen == 0,
                  "notSenior", "Senior"),
         Churn = ifelse(Churn == "Yes",
                        "Churn", "noChurch"))
dt1 %>% select(customerID, SeniorCitizen, Churn) %>%
  slice_tail(n=5)
```

```
## Warning: `...` is not empty.
##
## We detected these problematic arguments:
## * `needs_dots`
##
## These dots only exist to allow future extensions and should be empty.
## Did you misspecify an argument?
```

```
## # A tibble: 5 x 3
##   customerID SeniorCitizen Churn
##   <chr>      <chr>         <chr>
## 1 6840-RESVB notSenior     noChurch
## 2 2234-XADUH notSenior     noChurch
## 3 4801-JZAZL notSenior     noChurch
## 4 8361-LTMKD Senior        Churn
## 5 3186-AJIEK notSenior     noChurch
```

```
dt1 %>% is.na() %>% colSums()
```

```
##       customerID           gender    SeniorCitizen          Partner
##                0                0                0                0
##       Dependents           tenure     PhoneService    MultipleLines
##                0                0                0                0
##  InternetService   OnlineSecurity     OnlineBackup DeviceProtection
##                0                0                0                0
```

```
##      TechSupport      StreamingTV  StreamingMovies         Contract
##                0                0                0                0
## PaperlessBilling    PaymentMethod   MonthlyCharges     TotalCharges
##                0                0                0               11
##            Churn
##                0
```

```
dt1 %>%
  select(tenure, MonthlyCharges, TotalCharges) %>%
  summary()
```

```
##      tenure      MonthlyCharges    TotalCharges
##  Min.   : 0.00   Min.   : 18.25   Min.   :  18.8
##  1st Qu.: 9.00   1st Qu.: 35.50   1st Qu.: 401.4
##  Median :29.00   Median : 70.35   Median :1397.5
##  Mean   :32.37   Mean   : 64.76   Mean   :2283.3
##  3rd Qu.:55.00   3rd Qu.: 89.85   3rd Qu.:3794.7
##  Max.   :72.00   Max.   :118.75   Max.   :8684.8
##                                   NA's   :11
```

```
filter(dt1)
```

```
## Warning: `...` is not empty.
##
## We detected these problematic arguments:
## * `needs_dots`
##
## These dots only exist to allow future extensions and should be empty.
## Did you misspecify an argument?
```

```
## # A tibble: 7,043 x 21
##    customerID gender SeniorCitizen Partner Dependents tenure PhoneService
##    <chr>      <chr>  <chr>         <chr>   <chr>       <dbl> <chr>
##  1 7590-VHVEG Female notSenior     Yes     No              1 No
##  2 5575-GNVDE Male   notSenior     No      No             34 Yes
##  3 3668-QPYBK Male   notSenior     No      No              2 Yes
##  4 7795-CFOCW Male   notSenior     No      No             45 No
##  5 9237-HQITU Female notSenior     No      No              2 Yes
##  6 9305-CDSKC Female notSenior     No      No              8 Yes
##  7 1452-KIOVK Male   notSenior     No      Yes            22 Yes
##  8 6713-OKOMC Female notSenior     No      No             10 No
##  9 7892-POOKP Female notSenior     Yes     No             28 Yes
## 10 6388-TABGU Male   notSenior     No      Yes            62 Yes
## # ... with 7,033 more rows, and 14 more variables: MultipleLines <chr>,
## #   InternetService <chr>, OnlineSecurity <chr>, OnlineBackup <chr>,
## #   DeviceProtection <chr>, TechSupport <chr>, StreamingTV <chr>,
## #   StreamingMovies <chr>, Contract <chr>, PaperlessBilling <chr>,
## #   PaymentMethod <chr>, MonthlyCharges <dbl>, TotalCharges <dbl>,
## #   Churn <chr>
```

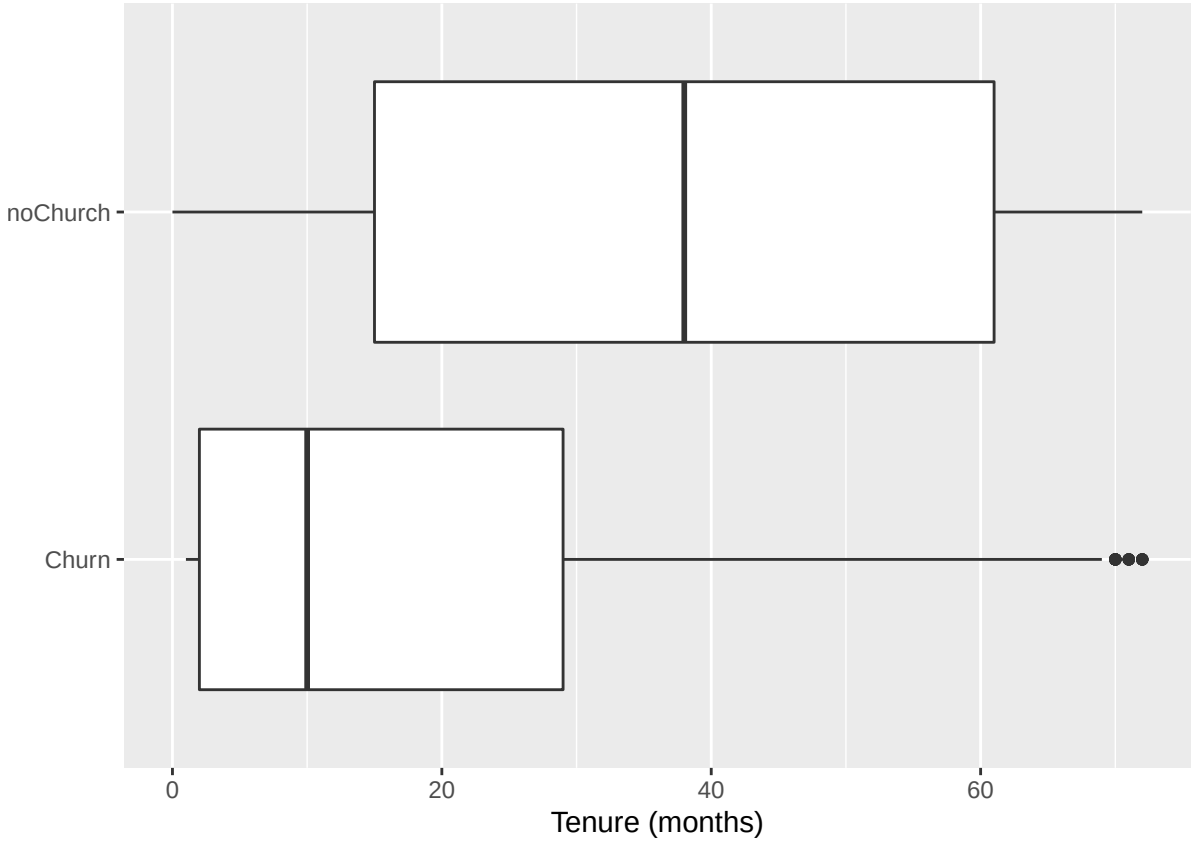
```
dt1 %>%
  group_by(Churn) %>%
  summarise(Q1Tenure = quantile(tenure,0.25),
            medTenure = median(tenure),
            Q3Tenure = quantile(tenure, 0.75))
```

```
## Warning: `...` is not empty.
```

```
##
## We detected these problematic arguments:
## * `needs_dots`
##
## These dots only exist to allow future extensions and should be empty.
## Did you misspecify an argument?
```

```
## # A tibble: 2 x 4
##   Churn    Q1Tenure medTenure Q3Tenure
##   <chr>       <dbl>     <dbl>    <dbl>
## 1 Churn           2        10       29
## 2 noChurch       15        38       61
```

```
churnBoxplot <- dt1 %>%
  ggplot(aes(x = Churn, y = tenure)) +
  geom_boxplot() +
  xlab("") + ylab("Tenure (months)") +
  coord_flip()
churnBoxplot
```




```
#see how many ppl left the company so far
dt1 %>%
  count(Churn)
```

```
## Warning: `...` is not empty.
##
## We detected these problematic arguments:
## * `needs_dots`
##
```

```
## These dots only exist to allow future extensions and should be empty.
## Did you misspecify an argument?
```

```
## # A tibble: 2 x 2
##   Churn        n
##   <chr>    <int>
## 1 Churn     1869
## 2 noChurch  5174
```

```
#coverting frequency
dt1 %>%
  count(Churn) %>%
  mutate(relFreq = n / sum(n))
```

```
## Warning: `...` is not empty.
##
## We detected these problematic arguments:
## * `needs_dots`
##
## These dots only exist to allow future extensions and should be empty.
## Did you misspecify an argument?
```

```
## # A tibble: 2 x 3
##   Churn        n relFreq
##   <chr>    <int>   <dbl>
## 1 Churn     1869   0.265
## 2 noChurch  5174   0.735
```

```
#generates frequency
dt1 %>% select(Churn) %>%
  table()
```

```
## .
##    Churn noChurch
##     1869     5174
```

```
#converting to relative frequency
dt1 %>% select(Churn) %>%
  table() %>% prop.table()
```

```
## .
##     Churn  noChurch
## 0.2653699 0.7346301
```

```
#generate percentage of senior citizens
dt1 %>% select(SeniorCitizen) %>%
  table() %>% prop.table
```

```
## .
## notSenior    Senior
## 0.8378532 0.1621468
```

```
#finding # of senior citizens who churn
dt1 %>%
  count(SeniorCitizen, Churn) %>%
  mutate(relFreq = n / sum(n))
```

```
## Warning: `...` is not empty.
##
```

```
## We detected these problematic arguments:
## * `needs_dots`
##
## These dots only exist to allow future extensions and should be empty.
## Did you misspecify an argument?
```

```
## # A tibble: 4 x 4
##   SeniorCitizen Churn        n relFreq
##   <chr>         <chr>    <int>   <dbl>
## 1 notSenior     Churn     1393  0.198
## 2 notSenior     noChurch  4508  0.640
## 3 Senior        Churn      476  0.0676
## 4 Senior        noChurch   666  0.0946
```

```
#removing frequency (n)
dt1 %>%count(SeniorCitizen, Churn) %>%
  mutate(relFreq = n / sum(n)) %>%
  select(-n)
```

```
## Warning: `...` is not empty.
##
## We detected these problematic arguments:
## * `needs_dots`
##
## These dots only exist to allow future extensions and should be empty.
## Did you misspecify an argument?
```

```
## # A tibble: 4 x 3
##   SeniorCitizen Churn    relFreq
##   <chr>         <chr>      <dbl>
## 1 notSenior     Churn     0.198
## 2 notSenior     noChurch  0.640
## 3 Senior        Churn     0.0676
## 4 Senior        noChurch  0.0946
```

```
#converting data to pivot table
dt1 %>%
  count(SeniorCitizen, Churn) %>%
  mutate(relFreq = n / sum(n)) %>%
  select(-n) %>%
  spread(SeniorCitizen, relFreq)
```

```
## Warning: `...` is not empty.
##
## We detected these problematic arguments:
## * `needs_dots`
##
## These dots only exist to allow future extensions and should be empty.
## Did you misspecify an argument?
```

```
## # A tibble: 2 x 3
##   Churn    notSenior Senior
##   <chr>        <dbl>  <dbl>
## 1 Churn        0.198 0.0676
## 2 noChurch     0.640 0.0946
```

```
#generate two way table
dt1 %>% select(Churn, SeniorCitizen) %>%
```

```
  table() %>% prop.table()
```

```
##           SeniorCitizen
## Churn       notSenior     Senior
##   Churn    0.19778503 0.06758484
##   noChurch 0.64006815 0.09456198
```

```
#conditional probabilities
dt1 %>%
  count(SeniorCitizen, Churn) %>%
  mutate(relFreq = n/ sum(n)) %>%
  group_by(SeniorCitizen) %>%
  mutate(condProbBySenior = relFreq/sum(relFreq))
```

```
## Warning: `...` is not empty.
##
## We detected these problematic arguments:
## * `needs_dots`
##
## These dots only exist to allow future extensions and should be empty.
## Did you misspecify an argument?
```

```
## # A tibble: 4 x 5
## # Groups:   SeniorCitizen [2]
##   SeniorCitizen Churn        n relFreq condProbBySenior
##   <chr>         <chr>    <int>   <dbl>            <dbl>
## 1 notSenior     Churn     1393  0.198             0.236
## 2 notSenior     noChurch  4508  0.640             0.764
## 3 Senior        Churn      476  0.0676            0.417
## 4 Senior        noChurch   666  0.0946            0.583
```

```
dt1 %>% select(Churn,SeniorCitizen) %>%
  table() %>% prop.table(2)
```

```
##           SeniorCitizen
## Churn       notSenior    Senior
##   Churn    0.2360617 0.4168126
##   noChurch 0.7639383 0.5831874
```

```
dt1 %>% select(Churn,SeniorCitizen) %>%
  table() %>% prop.table(2)*100
```

```
##           SeniorCitizen
## Churn       notSenior   Senior
##   Churn     23.60617 41.68126
##   noChurch  76.39383 58.31874
```

```
dt1 %>%
  group_by(MonthlyCharges>median(MonthlyCharges)) %>%
  select(Churn) %>%
  table() %>% prop.table()
```

```
## Adding missing grouping variables: `MonthlyCharges > median(MonthlyCharges)`
```

```
##                                         Churn
## MonthlyCharges > median(MonthlyCharges)      Churn   noChurch
##                                  FALSE 0.08973449 0.41118841
##                                  TRUE  0.17563538 0.32344172
```

```
dt1 %>% filter(SeniorCitizen=="notSenior") %>% select(TotalCharges) %>% summary()
```

```
##   TotalCharges
##  Min.   :  18.8
##  1st Qu.: 365.6
##  Median :1295.8
##  Mean   :2181.1
##  3rd Qu.:3566.4
##  Max.   :8684.8
##  NA's   :11
```

```
dt1 %>% group_by(....) %>%  summarize(...)
```

```
## Error: Must group by variables found in `.data`.
## * Column `....` is not found.
```

```
dt1 %>% filter(SeniorCitizen=="notSenior") %>% select(TotalCharges) %>% summary()
```

```
##   TotalCharges
##  Min.   :  18.8
##  1st Qu.: 365.6
##  Median :1295.8
##  Mean   :2181.1
##  3rd Qu.:3566.4
##  Max.   :8684.8
##  NA's   :11
```

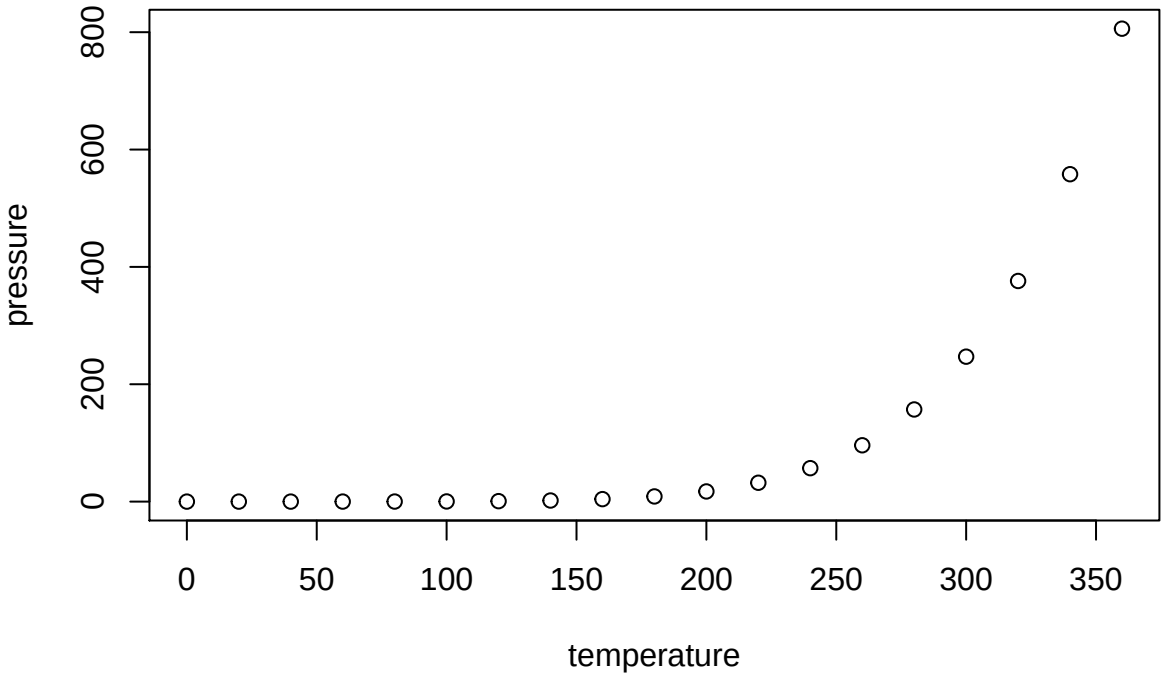
```
dt2 <- dt1 %>% mutate(

  contractLength=ifelse(Contract=="Month-to-month", "shortTerm", "longTerm"),

  autoPayment=ifelse(PaymentMethod=="Electronic check" |

                       PaymentMethod=="Mailed check", "manual", "automatic"))


dt1 %>%

  group_by(TotalCharges < mean(TotalCharges, na.rm = T)) %>%

  select(Churn) %>% table() %>% prop.table()
```

```
## Adding missing grouping variables: `TotalCharges < mean(TotalCharges, na.rm = T)`
```

```
##                                                Churn
## TotalCharges < mean(TotalCharges, na.rm = T)      Churn   noChurch
##                                        FALSE 0.06754835 0.30844710
##                                        TRUE  0.19823663 0.42576792
```

Including Plots

You can also embed plots, for example:



Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.