Intern_Assessment_Rawan_Hammad

Rawan Hammad

1/8/2022

This R-Markdown file will guide you through my method of recreating the summary statistics of the credit card types, current credit limits, and product types of higher risk, typical, and lower risk hypothetical credit card accounts.

In the first step of this analysis, I imported all the libraries I will likely need.

```
library(tidyr)
library(knitr)
library(readxl)
library(tidyverse)
library(dplyr)
library(gtsummary)
```

Next, I imported the three hypothetical datasets provided.

```
#importing the high risk cardholder file
cards_high_risk_2021 <- read.csv("C:/Users/rawan/OneDrive/Desktop/DePaul/Job
Assessments/Federal Reserve/cards-high-risk-2021.csv")

#importing the low risk cardholder file
cards_low_risk_2021 <- read.csv("C:/Users/rawan/OneDrive/Desktop/DePaul/Job
Assessments/Federal Reserve/cards-low-risk-2021.csv")

#importing the typical risk cardholder file
cards_typical_risk_2021 <-
read.csv("C:/Users/rawan/OneDrive/Desktop/DePaul/Job Assessments/Federal
Reserve/cards-typical-risk-2021.csv")
```

Now, it wouldn't be a bad idea to use the built-in summary function to view the data.

Let's first view the high risk dataset.

```
summary(cards_high_risk_2021)
```

```
##     loan_id       accountoriginationyear   activeflag    borrowerincome
##  Min.   :  1.00   Min.   :2003           Min.   :0.000   Min.   :     0
##  1st Qu.: 50.75   1st Qu.:2013           1st Qu.:0.000   1st Qu.: 21133
##  Median :100.50   Median :2016           Median :0.000   Median : 50785
##  Mean   :100.50   Mean   :2015           Mean   :0.025   Mean   : 81009
##  3rd Qu.:150.25   3rd Qu.:2018           3rd Qu.:0.000   3rd Qu.:108955
##  Max.   :200.00   Max.   :2019           Max.   :1.000   Max.   :395760
##  creditcardtype currentcreditlimit cycleendingbalance cycleendingretailapr
##  Min.   :1.00   Min.   :  220      Min.   :   20      Min.   : 0.37
```

```
##  1st Qu.:1.00   1st Qu.: 1332      1st Qu.:  695      1st Qu.:13.11
##  Median :1.00   Median : 2650      Median : 1510      Median :18.84
##  Mean   :1.17   Mean   : 3554      Mean   : 2240      Mean   :18.14
##  3rd Qu.:1.00   3rd Qu.: 4522      3rd Qu.: 2932      3rd Qu.:23.82
##  Max.   :2.00   Max.   :22080      Max.   :11050      Max.   :29.94
##   dayspastdue      monthendclosedrevokedflag originalcreditlimit
##  Min.   :  0.000   Min.   :0.000             Min.   :  100.0
##  1st Qu.:  0.000   1st Qu.:0.000             1st Qu.:  517.5
##  Median :  0.000   Median :0.000             Median : 1120.0
##  Mean   :  4.675   Mean   :0.025             Mean   : 2018.4
##  3rd Qu.:  0.000   3rd Qu.:0.000             3rd Qu.: 2422.5
##  Max.   :164.000   Max.   :1.000             Max.   :20920.0
##   producttype    refreshedcreditscoreprimaryborrower
##  Min.   :1.000   Min.   :450.0
##  1st Qu.:2.000   1st Qu.:602.0
##  Median :2.000   Median :649.0
##  Mean   :1.895   Mean   :635.2
##  3rd Qu.:2.000   3rd Qu.:677.0
##  Max.   :2.000   Max.   :699.0
```

The low risk dataset:

```
summary(cards_low_risk_2021)

##     loan_id       accountoriginationyear   activeflag borrowerincome
##  Min.   :  1.00   Min.   :2000           Min.   :0    Min.   :     0
##  1st Qu.: 50.75   1st Qu.:2011           1st Qu.:0    1st Qu.: 18078
##  Median :100.50   Median :2014           Median :0    Median : 47730
##  Mean   :100.50   Mean   :2013           Mean   :0    Mean   : 91745
##  3rd Qu.:150.25   3rd Qu.:2017           3rd Qu.:0    3rd Qu.:125095
##  Max.   :200.00   Max.   :2019           Max.   :0    Max.   :459850
##  creditcardtype currentcreditlimit cycleendingbalance cycleendingretailapr
##  Min.   :1.00   Min.   :  460      Min.   :   40      Min.   : 1.01
##  1st Qu.:1.00   1st Qu.: 4262      1st Qu.:  470      1st Qu.:11.93
##  Median :1.00   Median : 8110      Median : 1510      Median :15.97
##  Mean   :1.12   Mean   : 9124      Mean   : 2908      Mean   :15.96
##  3rd Qu.:1.00   3rd Qu.:12358      3rd Qu.: 3870      3rd Qu.:20.89
##  Max.   :2.00   Max.   :32460      Max.   :18690      Max.   :29.95
##   dayspastdue   monthendclosedrevokedflag originalcreditlimit  producttype
##  Min.   :0.00   Min.   :0                 Min.   :  140       Min.
:1.000
##  1st Qu.:0.00   1st Qu.:0                 1st Qu.: 2245       1st
Qu.:2.000
##  Median :0.00   Median :0                 Median : 4135       Median
:2.000
##  Mean   :0.07   Mean   :0                 Mean   : 6096       Mean
:1.835
##  3rd Qu.:0.00   3rd Qu.:0                 3rd Qu.: 7685       3rd
Qu.:2.000
##  Max.   :6.00   Max.   :0                 Max.   :29920       Max.
```

```
:2.000
##  refreshedcreditscoreprimaryborrower
##  Min.   :700.0
##  1st Qu.:729.0
##  Median :763.5
##  Mean   :765.4
##  3rd Qu.:798.0
##  Max.   :850.0
```

And the typical risk dataset:

```
summary(cards_typical_risk_2021)

##     loan_id       accountoriginationyear   activeflag borrowerincome
##  Min.   :  1.00   Min.   :2003           Min.   :0    Min.   :     0
##  1st Qu.: 50.75   1st Qu.:2011           1st Qu.:0    1st Qu.: 16168
##  Median :100.50   Median :2015           Median :0    Median : 55445
##  Mean   :100.50   Mean   :2014           Mean   :0    Mean   : 84045
##  3rd Qu.:150.25   3rd Qu.:2018           3rd Qu.:0    3rd Qu.:113578
##  Max.   :200.00   Max.   :2019           Max.   :0    Max.   :489080
##  creditcardtype  currentcreditlimit cycleendingbalance
cycleendingretailapr
##  Min.   :1.000   Min.   :  280      Min.   :   20      Min.   : 0.84
##  1st Qu.:1.000   1st Qu.: 2212      1st Qu.:  495      1st Qu.:10.86
##  Median :1.000   Median : 4780      Median : 1205      Median :14.46
##  Mean   :1.115   Mean   : 6463      Mean   : 2547      Mean   :15.10
##  3rd Qu.:1.000   3rd Qu.: 8802      3rd Qu.: 3055      3rd Qu.:20.26
##  Max.   :2.000   Max.   :28620      Max.   :20660      Max.   :29.81
##   dayspastdue     monthendclosedrevokedflag originalcreditlimit
producttype
##  Min.   :  0.00   Min.   :0                Min.   :   90.0     Min.
:1.00
##  1st Qu.:  0.00   1st Qu.:0                1st Qu.:  927.5     1st
Qu.:2.00
##  Median :  0.00   Median :0                Median : 2375.0     Median
:2.00
##  Mean   :  4.56   Mean   :0                Mean   : 4444.1     Mean
:1.85
##  3rd Qu.:  0.00   3rd Qu.:0                3rd Qu.: 5877.5     3rd
Qu.:2.00
##  Max.   :172.00   Max.   :0                Max.   :26930.0     Max.
:2.00
##  refreshedcreditscoreprimaryborrower
##  Min.   :450.0
##  1st Qu.:649.5
##  Median :709.0
##  Mean   :701.7
##  3rd Qu.:780.5
##  Max.   :848.0
```

We can see that all the variables match/overlap and there is no missing data in all three files. We can now create a summary of the three variables we're interested in: credit card types, current credit limit, and the product type. However, since the provided values for all these options are numerical, yet the output summary requires a categorical description, I've decided to bin a few values.

Now I'd like to take a minute to explain why I'm binning these values. I'm using a built-in function within the gtsummary library to create this summary. It is fast and efficient, and will provide us with all the data we need.

```
#creating a dataframe called highs and only including the credit card type, credit limit, and product type
highs <- cards_high_risk_2021 %>% select(creditcardtype, currentcreditlimit, producttype)

#the tbl summary is the built-in function within gtsummary that creates the summary below
highs %>% tbl_summary()

## Table printed with {flextable}, not {gt}. Learn why at
## http://www.danieldsjoberg.com/gtsummary/articles/rmarkdown.html
## To suppress this message, include `message = FALSE` in the code chunk header.
```

Characteristic	N = 200[1]
creditcardtype	
1	166 (83%)
2	34 (17%)
currentcreditlimit	2,650 (1,332, 4,522)
producttype	
1	21 (10%)
2	179 (90%)

[1]n (%); Median (IQR)

Based on the summary above, we're given percentages that match those in the example provided. However, we are not given any details on what the 1 and 2 values in product type and credit card type are. We also don't know what the current credit limit data is really telling us. This is why we need to bin the data.

Let's continue working with the higher risk file.

First, let's bin the credit limits:

```
#create a new data frame for the high risk data
high_risk <- cards_high_risk_2021

#create a new column for the binned credit limit
high_risk$"Current Credit Limit" <- high_risk$currentcreditlimit

#bin names
bin_names <- c("$1500 and less", "$1501-$7500", "over $7500")

#bin limits
bin_limits <- c(-Inf, 1501, 7500, Inf)

#implement the binning
high_risk$"Current Credit Limit" <- cut(high_risk$"Current Credit Limit",
breaks = bin_limits, labels = bin_names)
```

Next, we'll use the same method to rename 1's and 2's in credit card type and product type (by basically binning them):

```
#CREDIT CARD TYPE

#create a new column for the binned credit card type
high_risk$"Credit Card Type" <- high_risk$creditcardtype

#bin names
bin_names <- c("General Purpose", "Private Label")

#bin limits
bin_limits <- c(0,1,2)

#implement the binning
high_risk$"Credit Card Type" <- cut(high_risk$"Credit Card Type", breaks =
bin_limits, labels = bin_names)



#PRODUCT TYPE

#create a new column for the binned product type
high_risk$"Product Type" <- high_risk$producttype

#bin names
bin_names <- c("Co-brand", "Other")

#bin limits
bin_limits <- c(0,1,2)

#implement the binning
```

```
high_risk$"Product Type" <- cut(high_risk$"Product Type", breaks =
bin_limits, labels = bin_names)
```

We are now ready to view our final summary for the higher risk credit card accounts.

```
#create a new dataframe
highs <- high_risk %>% select("Credit Card Type", "Current Credit Limit",
"Product Type")


#now create the summary
highs <- highs %>% tbl_summary(digits = list(all_categorical() ~ c(0, 1)))
%>%
  modify_header(label ~ "**Variables**") %>%
  modify_caption("**Table 1. Summary Statistics of Hypothetical Higher-Risk
Credit Card Accounts**") %>%
  bold_labels()

highs
```

```
## Table printed with {flextable}, not {gt}. Learn why at
## http://www.danieldsjoberg.com/gtsummary/articles/rmarkdown.html
## To suppress this message, include `message = FALSE` in the code chunk
header.
```

Table 1. Summary Statistics of Hypothetical Higher-Risk Credit Card Accounts

Variables	N = 200[1]
Credit Card Type	
General Purpose	166 (83.0%)
Private Label	34 (17.0%)
Current Credit Limit	
$1500 and less	60 (30.0%)
$1501-$7500	121 (60.5%)
over $7500	19 (9.5%)
Product Type	
Co-brand	21 (10.5%)
Other	179 (89.5%)

[1]n (%)

We'll follow the same steps above to create a similar summary for the lower and typical risks.

Lower Risk credit cardholders:

```
#CURRENT CREDIT LIMIT

#create a new data frame for the high risk csv data
low_risk <- cards_low_risk_2021

#create a new column for the binned credit limit
low_risk$"Current Credit Limit" <- low_risk$currentcreditlimit

#bin names
bin_names <- c("$1500 and less", "$1501-$7500", "over $7500")

#bin limits
bin_limits <- c(-Inf, 1501, 7500, Inf)

#implement the binning
low_risk$"Current Credit Limit" <- cut(low_risk$"Current Credit Limit",
breaks = bin_limits, labels = bin_names)



#CREDIT CARD TYPE

#create a new column for the binned credit card type
low_risk$"Credit Card Type" <- low_risk$creditcardtype

#bin names
bin_names <- c("General Purpose", "Private Label")

#bin limits
bin_limits <- c(0,1,2)

#implement the binning
low_risk$"Credit Card Type" <- cut(low_risk$"Credit Card Type", breaks =
bin_limits, labels = bin_names)




#PRODUCT TYPE

#create a new column for the binned product type
low_risk$"Product Type" <- low_risk$producttype

#bin names
bin_names <- c("Co-brand", "Other")
```

```r
#bin limits
bin_limits <- c(0,1,2)

#implement the binning
low_risk$"Product Type" <- cut(low_risk$"Product Type", breaks = bin_limits,
labels = bin_names)

#create a new dataframe
lows <- low_risk %>% select("Credit Card Type", "Current Credit Limit",
"Product Type")


#now create the summary
lows <- lows %>% tbl_summary(digits = list(all_categorical() ~ c(0, 1))) %>%
  modify_header(label ~ "**Variables**") %>%
  modify_caption("**Table 2. Summary Statistics of Hypothetical Lower-Risk
Credit Card Accounts**") %>%
  bold_labels()

lows
```

```
## Table printed with {flextable}, not {gt}. Learn why at
## http://www.danieldsjoberg.com/gtsummary/articles/rmarkdown.html
## To suppress this message, include `message = FALSE` in the code chunk
header.
```

Table 2. Summary Statistics of Hypothetical Lower-Risk Credit Card Accounts

Variables	N = 200[1]
Credit Card Type	
General Purpose	176 (88.0%)
Private Label	24 (12.0%)
Current Credit Limit	
$1500 and less	16 (8.0%)
$1501-$7500	75 (37.5%)
over $7500	109 (54.5%)
Product Type	
Co-brand	33 (16.5%)
Other	167 (83.5%)

[1]n (%)

Finally, we'll run this for the typical risk credit card accounts:

```
#CURRENT CREDIT LIMIT

#create a new data frame for the high risk csv data
typical_risk <- cards_typical_risk_2021

#create a new column for the binned credit limit
typical_risk$"Current Credit Limit" <- typical_risk$currentcreditlimit

#bin names
bin_names <- c("$1500 and less", "$1501-$7500", "over $7500")

#bin limits
bin_limits <- c(-Inf, 1501, 7500, Inf)

#implement the binning
typical_risk$"Current Credit Limit" <- cut(typical_risk$"Current Credit
Limit", breaks = bin_limits, labels = bin_names)



#CREDIT CARD TYPE

#create a new column for the binned credit card type
typical_risk$"Credit Card Type" <- typical_risk$creditcardtype

#bin names
bin_names <- c("General Purpose", "Private Label")

#bin limits
bin_limits <- c(0,1,2)

#implement the binning
typical_risk$"Credit Card Type" <- cut(typical_risk$"Credit Card Type",
breaks = bin_limits, labels = bin_names)



#PRODUCT TYPE

#create a new column for the binned product type
typical_risk$"Product Type" <- typical_risk$producttype

#bin names
bin_names <- c("Co-brand", "Other")
```

```r
#bin limits
bin_limits <- c(0,1,2)

#implement the binning
typical_risk$"Product Type" <- cut(typical_risk$"Product Type", breaks =
bin_limits, labels = bin_names)

#create a new dataframe
typicals <- typical_risk %>% select("Credit Card Type", "Current Credit
Limit", "Product Type")


#now create the summary
typicals <- typicals %>% tbl_summary(digits = list(all_categorical() ~ c(0,
1))) %>%
  modify_header(label ~ "**Variables**") %>%
  modify_caption("**Table 3. Summary Statistics of Hypothetical Typical-Risk
Credit Card Accounts**") %>%
  bold_labels()

typicals
```

```
## Table printed with {flextable}, not {gt}. Learn why at
## http://www.danieldsjoberg.com/gtsummary/articles/rmarkdown.html
## To suppress this message, include `message = FALSE` in the code chunk
header.
```

Table 3. Summary Statistics of Hypothetical Typical-Risk Credit Card Accounts

Variables	N = 200[1]
Credit Card Type	
General Purpose	177 (88.5%)
Private Label	23 (11.5%)
Current Credit Limit	
$1500 and less	33 (16.5%)
$1501-$7500	105 (52.5%)
over $7500	62 (31.0%)
Product Type	
Co-brand	30 (15.0%)
Other	170 (85.0%)

[1]n (%)

Finally, we can show all the tables below for comparison. I tried combining the three tables into 1, but I couldn't get a visually appealing result. Therefore, I decided to maintain the user-friendly results we have below. If you happen to be familiar with a better way to combine these three summaries into one, I'd be happy to discuss it!

Thanks for your time!

FINAL RESULTS BELOW:

```
#plotting the summaries
highs
```

Table 1. Summary Statistics of Hypothetical Higher-Risk Credit Card Accounts

Variables	N = 200[1]
Credit Card Type	
General Purpose	166 (83.0%)
Private Label	34 (17.0%)
Current Credit Limit	
$1500 and less	60 (30.0%)
$1501-$7500	121 (60.5%)
over $7500	19 (9.5%)
Product Type	
Co-brand	21 (10.5%)
Other	179 (89.5%)

[1]n (%)

```
lows
```

Table 2. Summary Statistics of Hypothetical Lower-Risk Credit Card Accounts

Variables	N = 200[1]
Credit Card Type	
General Purpose	176 (88.0%)
Private Label	24 (12.0%)
Current Credit Limit	
$1500 and less	16 (8.0%)
$1501-$7500	75 (37.5%)

Variables	N = 200[1]
over $7500	109 (54.5%)
Product Type	
Co-brand	33 (16.5%)
Other	167 (83.5%)

[1]n (%)

```
typicals
```

Table 3. Summary Statistics of Hypothetical Typical-Risk Credit Card Accounts

Variables	N = 200[1]
Credit Card Type	
General Purpose	177 (88.5%)
Private Label	23 (11.5%)
Current Credit Limit	
$1500 and less	33 (16.5%)
$1501-$7500	105 (52.5%)
over $7500	62 (31.0%)
Product Type	
Co-brand	30 (15.0%)
Other	170 (85.0%)

[1]n (%)