Portfolio Analysis by Onyeka Okonkwo

Code

#BUSINESS INFO AND PROBLEM Data is for a loan company serving the under-served market. The company provides credit facilities to customers for uses ranging from Business, Education, Groceries and Personal effects.

#TASK 1. Descriptive analytics of the data 2. PAR Analysis showing evolution & recommendations, for this analysis please use PAR7, PAR15, PAR30 and PAR60

#IMPORT DATA

Import, Load libraries and View data

Hide

```
library(readr)
```

```
Warning: package ‘readr’ was built under R version 4.1.1
```

Hide

```
library(dplyr)
```

```
Warning: package ‘dplyr’ was built under R version 4.1.1

Attaching package: ‘dplyr’

The following objects are masked from ‘package:stats’:

    filter, lag

The following objects are masked from ‘package:base’:

    intersect, setdiff, setequal, union
```

Hide

```
library(lubridate)
```

```
Warning: package ‘lubridate’ was built under R version 4.1.1

Attaching package: ‘lubridate’

The following objects are masked from ‘package:base’:

    date, intersect, setdiff, union
```

Hide

```
df <- read_csv("C:/Users/onokonkwo/Desktop/Credit Case Study/data_case_study.csv")
```

```
Rows: 350844 Columns: 18
-- Column specification --------------------------------------------------------
-------
Delimiter: ","
chr (13): date, loan_id, loan_status, product_type, reason, approval_date, last_payment_date, de
fau...
dbl  (3): user_id, repaid, days_late

i Use `spec()` to retrieve the full column specification for this data.
i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

Hide

```
View(df)
```

Data structure

Hide

```
str(df)
```

```
spec_tbl_df [350,844 x 18] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
 $ date             : chr [1:350844] "15/03/2021" "15/03/2021" "15/03/2021" "15/03/2021" ...
 $ user_id          : num [1:350844] 1483192 1173117 1352358 1290014 1504236 ...
 $ loan_id          : chr [1:350844] "e1f35912-301a-4e7f-b769-4c0ae4482963" "d465a501-ed4b-432d-
b7b3-b8c39da25fbe" "3cc12530-9046-4b87-b327-1757cb95b361" "0c433515-d8e1-4286-9509-8ebfa974e9f9"
...
 $ principal        : num [1:350844] 14286 38095 11905 11905 11905 ...
 $ balance          : num [1:350844] 14286 38095 11905 11905 11905 ...
 $ repaid           : num [1:350844] 0 0 0 0 0 ...
 $ days_late        : num [1:350844] NA NA NA NA NA NA NA NA NA 20 ...
 $ loan_status      : chr [1:350844] "On Time" "On Time" "On Time" "On Time" ...
 $ product_type     : chr [1:350844] "SHORT" "SHORT" "SHORT" "SHORT" ...
 $ reason           : chr [1:350844] "Business" "Business" "Business" "Medical fees" ...
 $ approval_date    : chr [1:350844] "10/03/2021" "10/03/2021" "10/03/2021" "10/03/2021" ...
 $ last_payment_date: chr [1:350844] "25/03/2021" "25/03/2021" "09/04/2021" "09/04/2021" ...
 $ default_date     : chr [1:350844] "23/06/2021" "23/06/2021" "08/07/2021" "08/07/2021" ...
 $ bank             : chr [1:350844] "G" "G" "O" "C" ...
 $ gender           : chr [1:350844] "male" "male" "female" "male" ...
 $ state            : chr [1:350844] "Y25" "Y25" "BY" "D4" ...
 $ date_of_birth    : chr [1:350844] "23/07/1983" "25/11/1972" "06/05/1987" "19/01/1992" ...
 $ employment_status: chr [1:350844] "SELF-EMPLOYED" "SELF-EMPLOYED" "SELF-EMPLOYED" "SELF-EMPLO
YED" ...
 - attr(*, "spec")=
  .. cols(
  ..   date = col_character(),
  ..   user_id = col_double(),
  ..   loan_id = col_character(),
  ..   principal = col_number(),
  ..   balance = col_number(),
  ..   repaid = col_double(),
  ..   days_late = col_double(),
  ..   loan_status = col_character(),
  ..   product_type = col_character(),
  ..   reason = col_character(),
  ..   approval_date = col_character(),
  ..   last_payment_date = col_character(),
  ..   default_date = col_character(),
  ..   bank = col_character(),
  ..   gender = col_character(),
  ..   state = col_character(),
  ..   date_of_birth = col_character(),
  ..   employment_status = col_character()
  .. )
 - attr(*, "problems")=<externalptr>
```

#DATA CLEANING AND PREP

###Change class of variables

rounding numeric values to 2 decimal

Hide

```
df[,4:6] <- round(df[,4:6], digits = 2)
```

```
Warning: One or more parsing issues, see `problems()` for details
```

Characters to Factors

Hide

```
df[,8:10] <- lapply(df[,8:10], as.factor)
df$gender <- as.factor(df$gender)
df$employment_status <- as.factor(df$employment_status)
```

Character to Dates

Hide

```
df$date_of_birth <- dmy(df$date_of_birth)
df$approval_date <- dmy(df$approval_date)
```

Hide

```
df$date <- dmy(df$date)
df$last_payment_date <- dmy(df$last_payment_date)
df$default_date <- dmy(df$default_date)
```

Arrange Date column in chronological order

Hide

```
df %>% arrange(date)
```

date <date>	user_id <dbl>	loan_id <chr>	principal <dbl>	balance <dbl>	repai <dbl>
2021-03-15	1483192	e1f35912-301a-4e7f-b769-4c0ae4482963	14285.7	14285.7	0.0
2021-03-15	1173117	d465a501-ed4b-432d-b7b3-b8c39da25fbe	38095.2	38095.2	0.0
2021-03-15	1352358	3cc12530-9046-4b87-b327-1757cb95b361	11904.8	11904.8	0.0
2021-03-15	1290014	0c433515-d8e1-4286-9509-8ebfa974e9f9	11904.8	11904.8	0.0
2021-03-15	1504236	4a498d09-e54c-4d4b-95f5-ba3ed007d1aa	11904.8	11904.8	0.0
2021-03-15	1099554	f6426a8c-201d-424d-ac08-ea06c49d8f51	14285.7	14285.7	0.0
2021-03-15	1488311	5b485d9c-930f-4b82-9c38-1f96787984aa	7142.9	7142.9	0.0
2021-03-15	348033	15832e96-6733-47c8-baf9-1ab364a43466	52381.0	52381.0	0.0
2021-03-15	409808	476ff167-7a7a-4816-920e-e707c9c81b08	71428.6	71428.6	0.0
2021-03-15	1243070	2f844e03-b60c-466b-802d-dd95683cc68c	64285.7	64285.7	11904.7

1-10 of 350,844 rows | 1-7 of 18 columns Previous **1** 2 3 4 5 6 … 100 Next

Search for missing values, identifies how many and what columns they're in

Hide

```
sum(is.na(df))
```

```
[1] 217878
```

Hide

```
names(df[, !complete.cases(t(df))])
```

```
[1] "balance"       "days_late"     "date_of_birth"
```

Summary of Data

Hide

```
summary(df)
```

```
      date               user_id          loan_id            principal         balance
 Min.   :2021-03-15   Min.   :     26   Length:350844      Min.   :  7143   Min.   :     1.2
 1st Qu.:2021-03-29   1st Qu.: 517404   Class :character   1st Qu.: 11905   1st Qu.: 11904.8
 Median :2021-04-12   Median :1187943   Mode  :character   Median : 23810   Median : 23809.5
 Mean   :2021-04-08   Mean   : 992014                      Mean   : 33211   Mean   : 32530.2
 3rd Qu.:2021-04-26   3rd Qu.:1433756                      3rd Qu.: 52381   3rd Qu.: 52381.0
 Max.   :2021-05-03   Max.   :1627539                      Max.   :273810   Max.   :273809.5
                                                                            NA's   :144
     repaid         days_late       loan_status      product_type                       reason
 Min.   :     0   Min.   : 1.0     Late   :133118   LONG   :  2596   Business             :15061
9
 1st Qu.:     0   1st Qu.:14.0     On Time:217726   SCALING: 67164   Household Goods      : 3932
0
 Median :     0   Median :33.0                      SHORT  :281084   Medical fees         : 3439
7
 Mean   :  1591   Mean   :37.2                                       Personal/Confidential: 3438
2
 3rd Qu.:     0   3rd Qu.:59.0                                       Emergency            : 2928
4
 Max.   :250000   Max.   :90.0                                       Education            : 1816
2
                  NA's   :217726                                     (Other)              : 4468
0
 approval_date        last_payment_date     default_date            bank              gender
 Min.   :2020-10-30   Min.   :2020-12-16   Min.   :2021-03-16   Length:350844      female:110833
 1st Qu.:2021-02-22   1st Qu.:2021-03-18   1st Qu.:2021-06-16   Class :character   male  :240011
 Median :2021-03-15   Median :2021-04-09   Median :2021-07-08   Mode  :character
 Mean   :2021-03-09   Mean   :2021-04-03   Mean   :2021-07-02
 3rd Qu.:2021-04-02   3rd Qu.:2021-04-29   3rd Qu.:2021-07-28
 Max.   :2021-05-03   Max.   :2021-12-18   Max.   :2022-03-18

    state           date_of_birth            employment_status
 Length:350844      Min.   :1970-01-01   EMPLOYED     :162737
 Class :character   1st Qu.:1981-03-03   SELF-EMPLOYED:188107
 Mode  :character   Median :1987-02-14
                    Mean   :1986-07-04
                    3rd Qu.:1992-05-12
                    Max.   :2003-04-14
                    NA's   :8
```

###NOTES

1. Data length is 350844 rows and 18 columns (created additional for Age)
2. Date ranges from 15/3/2021 to 3/05/2021
3. Highest amount disbursed was N273,810 and lowest was N7,143
4. 37.9% of customers were late on payment
5. Company has 3 product category - Short, Scaling and Long. Majority of customers (80%) took Short product. This increases liquidity and availability of working capital for the company if payment is timely.
6. 42.9% required loan for business, while between 5% and 11% needed a loan for Household goods, Personal, Emergency, Education and Medical Fees.
7. Gender of customers was 68% male and 31% female.

8. 53.6% were self-employed and 46.3% were employed.
9. Customers age range from 18 to 51.

Check unique values in loan id, user id, state and bank

Hide

```
length(unique(df$user_id))
```

```
[1] 59988
```

Hide

```
length(unique(df$loan_id))
```

```
[1] 108752
```

Hide

```
length(unique(df$state))
```

```
[1] 37
```

Hide

```
length(unique(df$bank))
```

```
[1] 19
```

###NOTES:

1. 59998 customers over the period
2. 108752 unique loan IDs i.e loans disbursed during the period
3. Customers take multiple loans, often without paying up the prior loan. See examples in users 1483192 and 1488311
4. Customers are resident in 37 states
5. Loan was disbursed into 19 different banks

#TREATMENT OF MISSING VALUES (NA)

Date of Birth (8) impute with median value

Hide

```
df$date_of_birth[is.na(df$date_of_birth)] <- median(df$date_of_birth, na.rm = TRUE)
```

User ID 439447 (1258b7a2-6a23-4146-aed4-e0117ea29234) has no Date of Birth recorded.

Days Late (217726) Calculate for values i.e. if last payment date is less than date, then calculate days late, otherwise return value as 0.

Hide

```
df$days_late <- ifelse((df$last_payment_date < df$date), (df$date - df$last_payment_date), 0)
```

Balance (144) - calculate field and fill. We'll run multiple test conditions for this.

###Test 1 (144 NA) Replace with corresponding principal value when repayment is 0

Hide

```
df$balance <- ifelse(df$repaid==0 & is.na(df$balance), df$principal, df$balance)
```

###Test 2 (135 NA) Group by User ID and Loan ID, then fill NA in balance with Principal if the repaid value is same as above

Hide

```
df2 <- df%>%
  group_by(user_id, loan_id) %>%
  mutate(balance = ifelse(is.na(balance) & repaid == lag(repaid, n=1), principal, balance))
```

###(contd. Test 2 - 76 NA) Group by User ID and Loan ID, then fill NA in balance with Principal if the repaid value is same as below

Hide

```
df3 <- df2%>%
  group_by(user_id, loan_id) %>%
  mutate(balance = ifelse(is.na(balance) & repaid == lead(repaid, n=1), principal, balance))
```

###Test 3 (52 NA) Group by User ID and Loan ID, then fill NA in Balance with 0 if customer paid more than existing loan balance, otherwise leave value as NA.

Hide

```
df4 <- df3 %>%
  group_by(user_id, loan_id) %>%
  mutate(balance = ifelse(is.na(balance) & (repaid - lag(repaid, n=1))>lag(balance, n=1), 0, bal
ance))
```

###Test 4 (15 NA) Group by user Id and Loan ID, if the repaid value is greater than existing loan balance, then fill with 0, otherwise leave as NA

Hide

```
df5 <- df4%>%
  group_by(user_id, loan_id) %>%
  mutate(balance = ifelse(is.na(balance) & repaid > lag(balance, n=1), 0, balance))
```

This code seems similar to the previous chunk so no NAs were affected. May be okay to omit, but safe to keep it.

###Test 5 (15 NA) Return 0 if repaid value is greater than loan principal

Hide

```
df6 <- df5%>%
  group_by(user_id, loan_id) %>%
  mutate(balance = ifelse(is.na(balance) & repaid > principal, 0, balance))
```

###check if there's any NA left

Hide

```
sum(is.na(df6))
```

```
[1] 0
```

Hide

```
names(df6[, !complete.cases(t(df6))])
```

```
character(0)
```

All cleared!! Now we can move forward.

Create New column for Age of Customers using Approval Date and DOB

Hide

```
library(dplyr)
calc_age <- function(birthDate, refDate = Sys.Date(), unit = "year") {

  require(lubridate)

  if(grepl(x = unit, pattern = "year")) {
    as.period(interval(birthDate, refDate), unit = 'year')$year
  } else if(grepl(x = unit, pattern = "month")) {
    as.period(interval(birthDate, refDate), unit = 'month')$month
  } else if(grepl(x = unit, pattern = "week")) {
    floor(as.period(interval(birthDate, refDate), unit = 'day')$day / 7)
  } else if(grepl(x = unit, pattern = "day")) {
    as.period(interval(birthDate, refDate), unit = 'day')$day
  } else {
    print("Argument 'unit' must be one of 'year', 'month', 'week', or 'day'")
    NA
  }

}
df6$Age = calc_age(birthDate = df6$date_of_birth, refDate = df6$approval_date, unit = "year")
```

rename df6

Hide

```
Cleandf <- df6
```

VIEW DATA

Hide

```
View(Cleandf)
```

#DATA VISUALISATION AND EDA

Hide

```
library(ggplot2)
```

```
Warning: package ‘ggplot2’ was built under R version 4.1.1
```

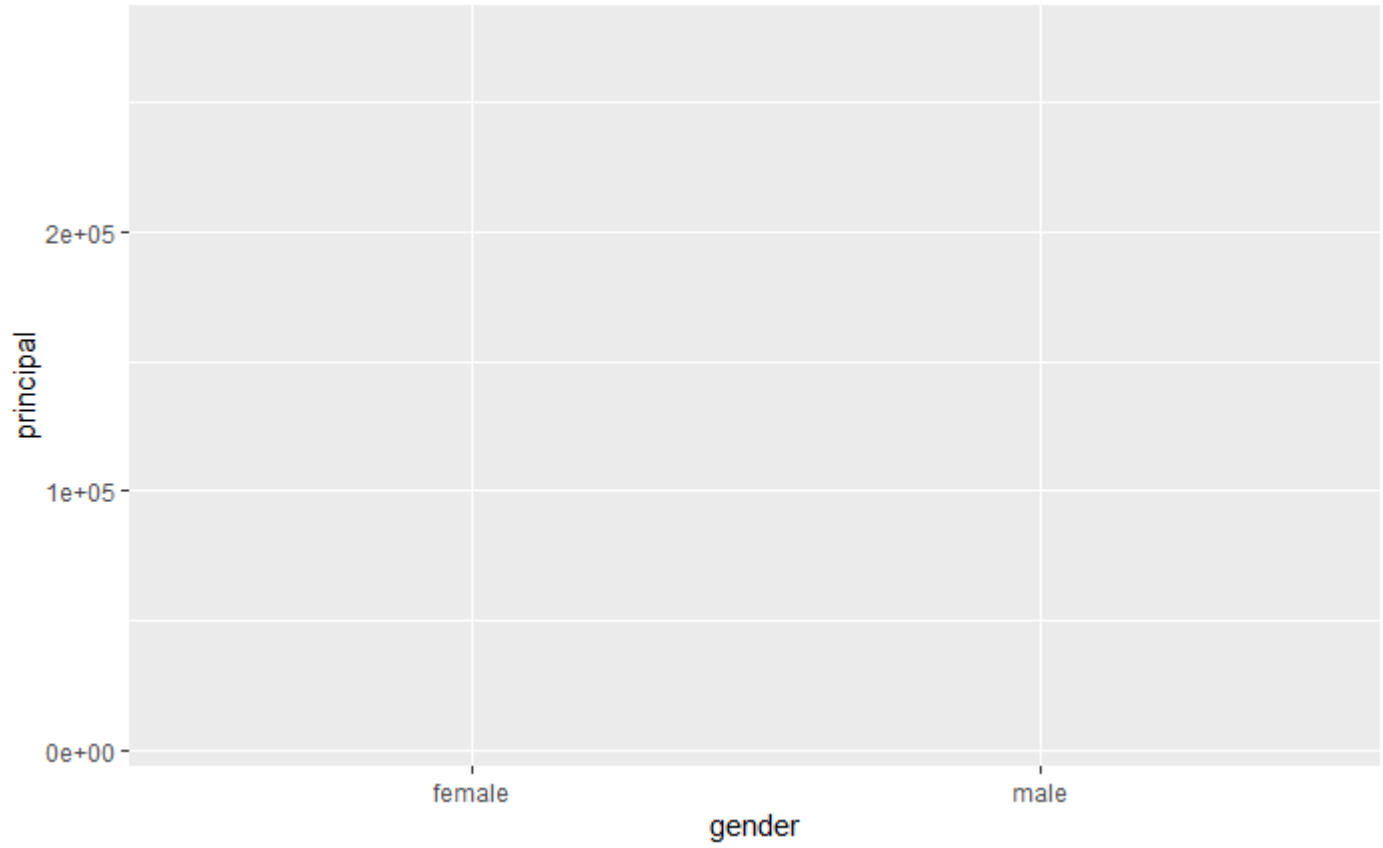
Hide

```
library(DataExplorer)
```

```
Warning: package ‘DataExplorer’ was built under R version 4.1.1
Registered S3 method overwritten by 'data.table':
  method           from
  print.data.table
Registered S3 method overwritten by 'htmlwidgets':
  method           from
  print.htmlwidget tools:rstudio
```

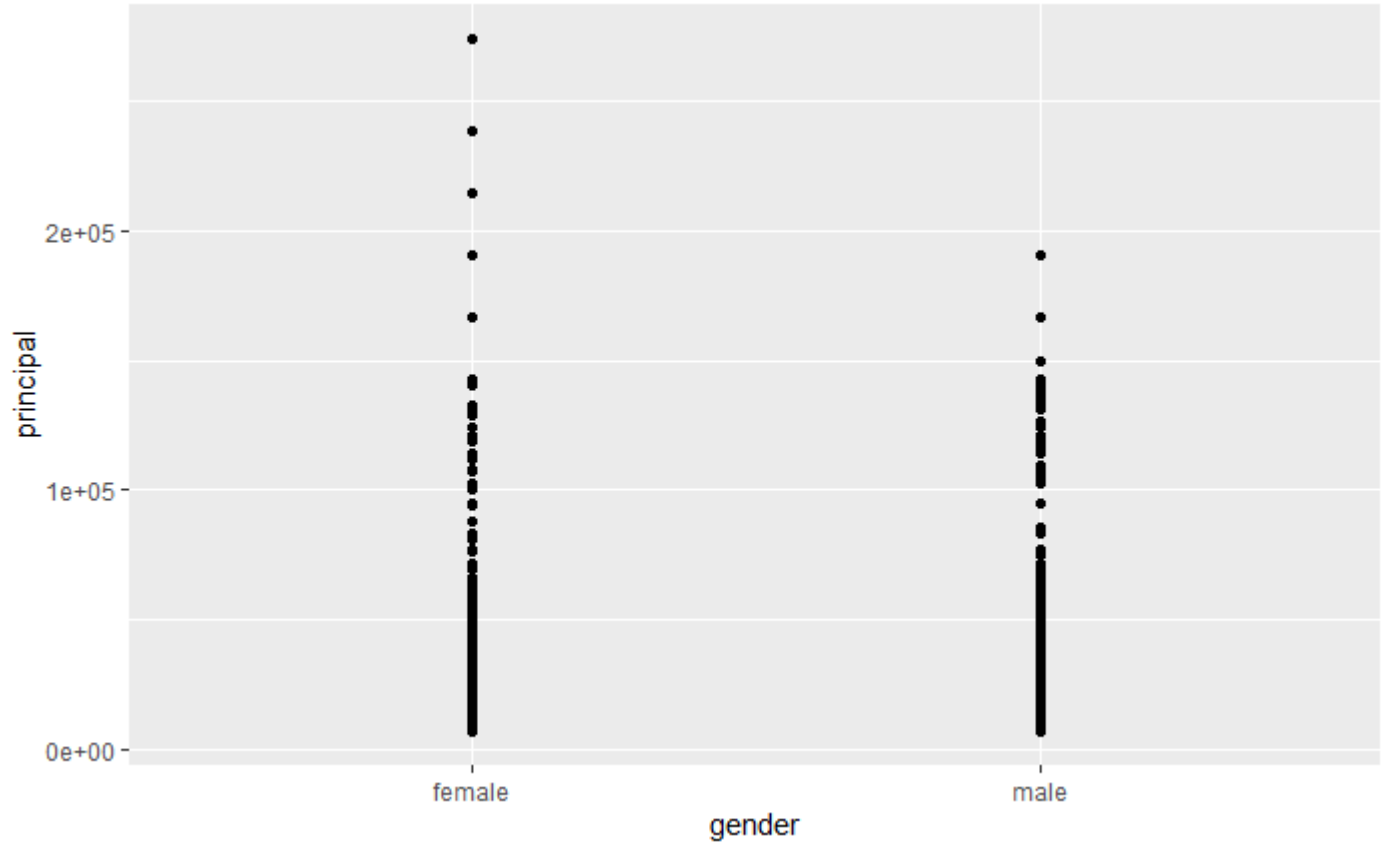
Hide

```
p <- ggplot(data = Cleandf, aes(x = gender , y = principal))
print(p)
```



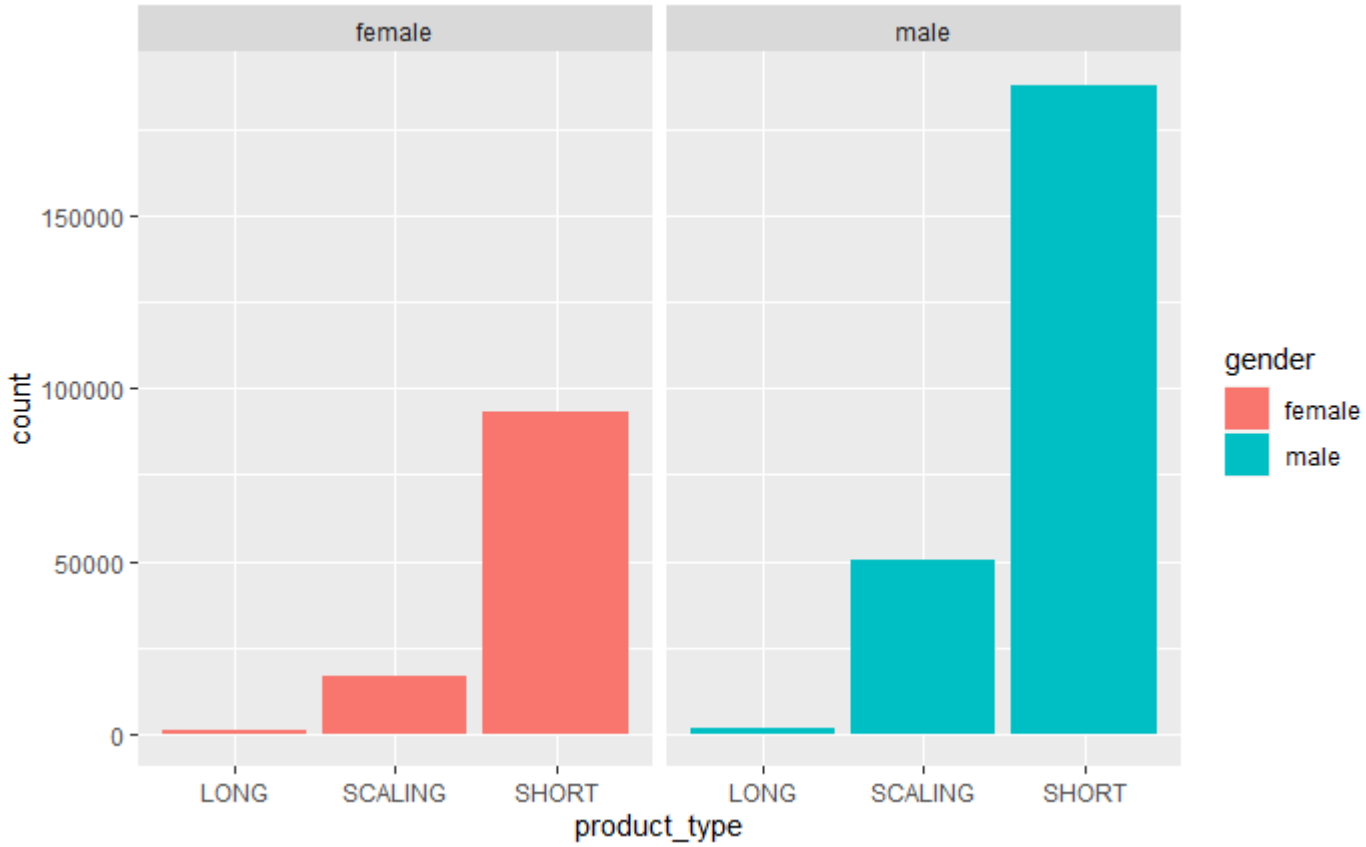
Hide

```
print(p + geom_point())
```



Hide

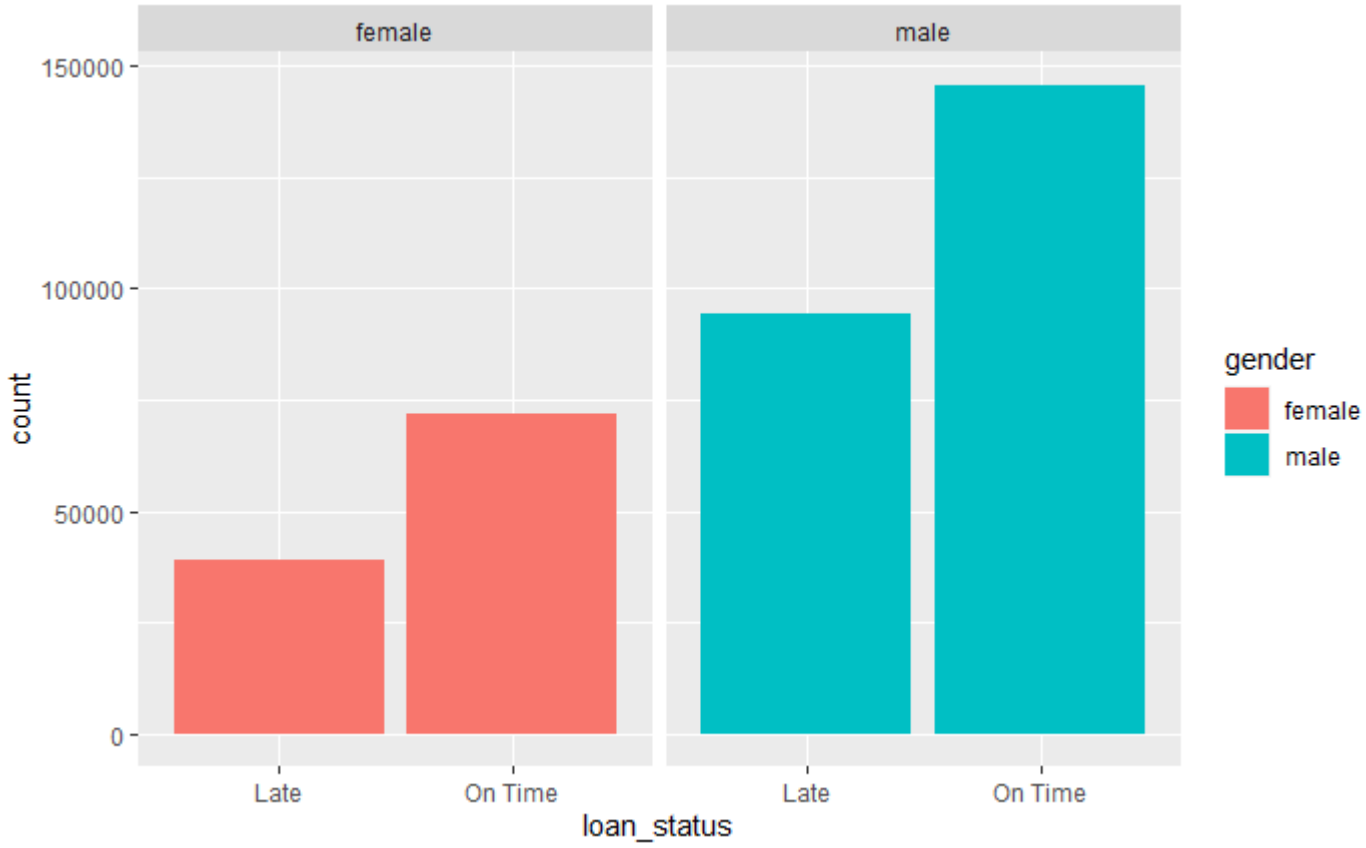
```
ggplot(group_by(Cleandf, loan_id), aes(x = product_type, fill = gender))+
  geom_bar() +
  facet_wrap(~ gender)
```



Between the 3 product categories for both genders, more customers took Short loans, and men took more loans in general than women.

Hide

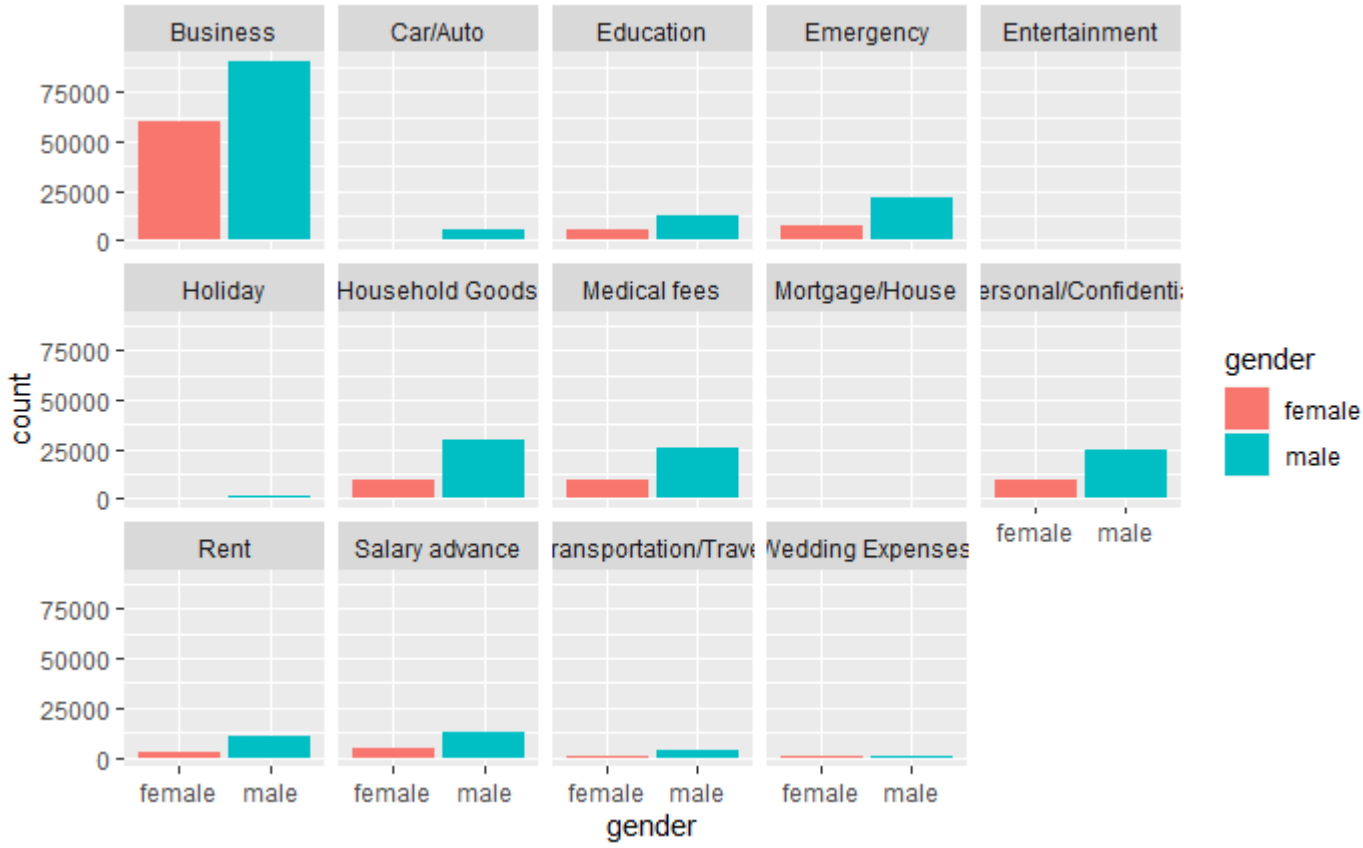
```
ggplot(data = Cleandf, aes(x = loan_status, fill = gender))+
  geom_bar() +
  facet_wrap(~ gender)
```



About 62% women pay on time, compared to about 60% of men. The difference in ability to pay on time does not appear to be significant.

Hide

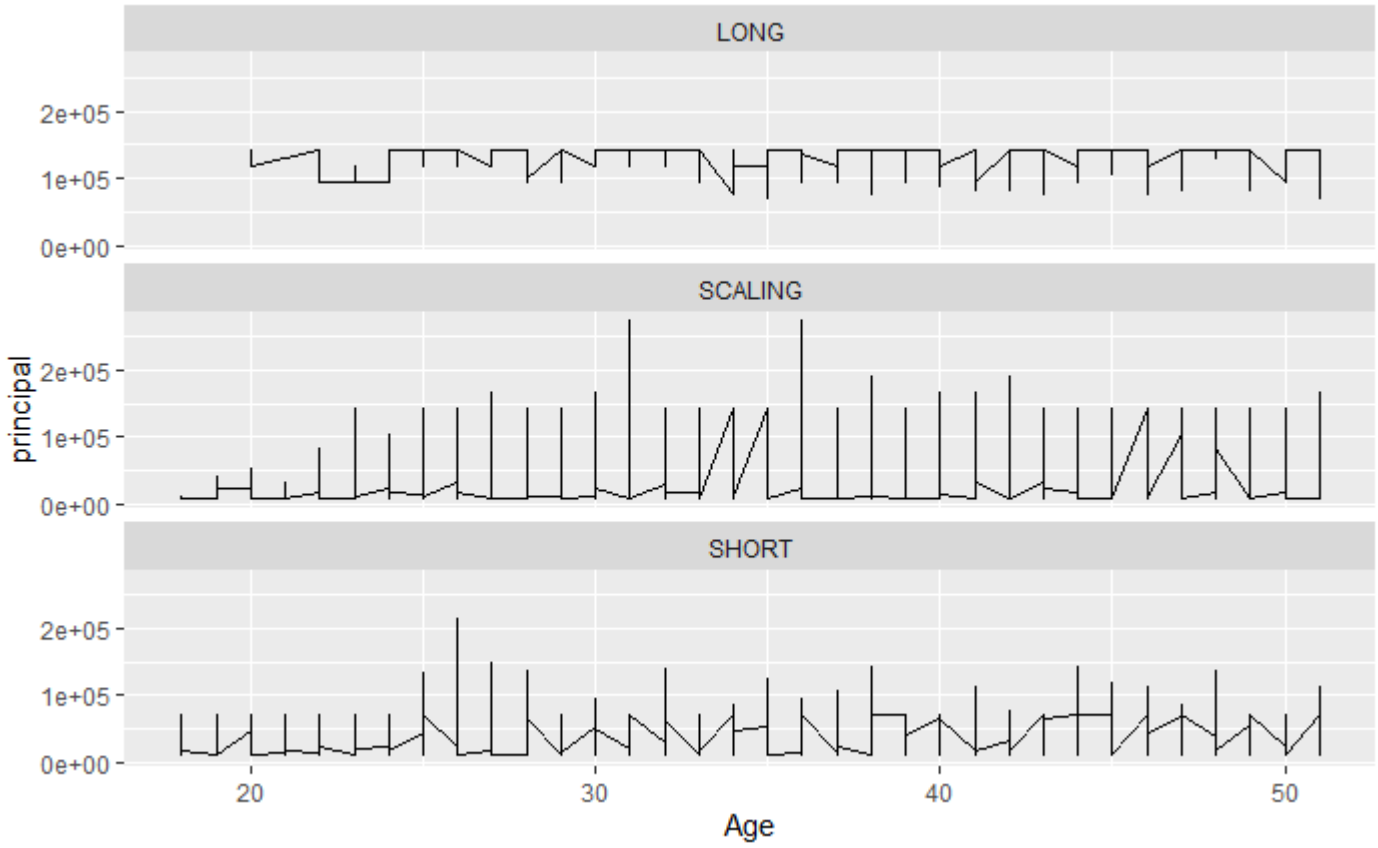
```
ggplot(group_by(Cleandf, loan_id), aes(x = gender, fill = gender))+
  geom_bar() +
  facet_wrap(~ reason, nrow = 3)
```



Both gender took more business loans, and others in these category topped reasons for loans collected - emergency, household goods, medical fees and person/confidential reasons.

Hide

```
ggplot(group_by(Cleandf, loan_id), aes(x = Age, y = principal))+
  geom_line()+
  facet_wrap(~ product_type, nrow = 3)
```



In scaling category, customers between ages 30 and 40 took the most loans. For short loans ages 25 - 30 took more loans. Collection of long loans are evenly distributed across the age groups.

Hide

```
ggplot(group_by(Cleandf, loan_id), aes(employment_status, fill = gender))+
  geom_bar() +
  facet_wrap(~ gender, nrow = 2)
```

Among the customers, there are more self-employed women and more employed men.

Hide

```
create_report(Cleandf)
```

```
processing file: report.rmd
```

```
  |
  |
|   0%
  |
  |..
|   2%
   inline R code fragments


  |
  |....
|   5%
label: global_options (with options)
List of 1
 $ include: logi FALSE


  |
  |.......
|   7%
  ordinary text without R code


  |
  |.........
|  10%
label: introduce

  |
  |...........
|  12%
  ordinary text without R code


  |
  |.............
|  14%
label: plot_intro

  |
  |...............
|  17%
  ordinary text without R code


  |
  |.................
|  19%
label: data_structure

  |
  |...................
```

```
|  21%
  ordinary text without R code


  |
  |.....................
|  24%
label: missing_profile

  |
  |.......................
|  26%
  ordinary text without R code


  |
  |.........................
|  29%
label: univariate_distribution_header

  |
  |.............................
|  31%
  ordinary text without R code


  |
  |...............................
|  33%
label: plot_histogram

  |
  |.................................
|  36%
  ordinary text without R code


  |
  |...................................
|  38%
label: plot_density

  |
  |.....................................
|  40%
  ordinary text without R code


  |
  |.......................................
|  43%
label: plot_frequency_bar

  |
```

```
  |.....................................
|  45%
  ordinary text without R code


  |
  |.......................................
|  48%
label: plot_response_bar

  |
  |.........................................
|  50%
  ordinary text without R code


  |
  |.........................................
|  52%
label: plot_with_bar

  |
  |.................................................
|  55%
  ordinary text without R code


  |
  |.....................................................
|  57%
label: plot_normal_qq

  |
  |.......................................................
|  60%
  ordinary text without R code


  |
  |.........................................................
|  62%
label: plot_response_qq

  |
  |...........................................................
|  64%
  ordinary text without R code


  |
  |.............................................................
|  67%
label: plot_by_qq
```

```
  |
  |.....................................................................
|  69%
  ordinary text without R code


  |
  |.....................................................................
|  71%
label: correlation_analysis

  |
  |.....................................................................
|  74%
  ordinary text without R code


  |
  |.....................................................................
|  76%
label: principal_component_analysis

  |
  |.....................................................................
|  79%
  ordinary text without R code


  |
  |.....................................................................
|  81%
label: bivariate_distribution_header

  |
  |.....................................................................
|  83%
  ordinary text without R code


  |
  |.....................................................................
|  86%
label: plot_response_boxplot

  |
  |.....................................................................
|  88%
  ordinary text without R code


  |
  |.....................................................................
|  90%
label: plot_by_boxplot
```

```
  |
  |.....................................................................
|  93%
  ordinary text without R code


  |
  |.....................................................................
|  95%
label: plot_response_scatterplot

  |
  |.....................................................................
|  98%
  ordinary text without R code


  |
  |.....................................................................
| 100%
label: plot_by_scatterplot
```

```
output file: C:/Users/onokonkwo/Desktop/R code/report.knit.md
```

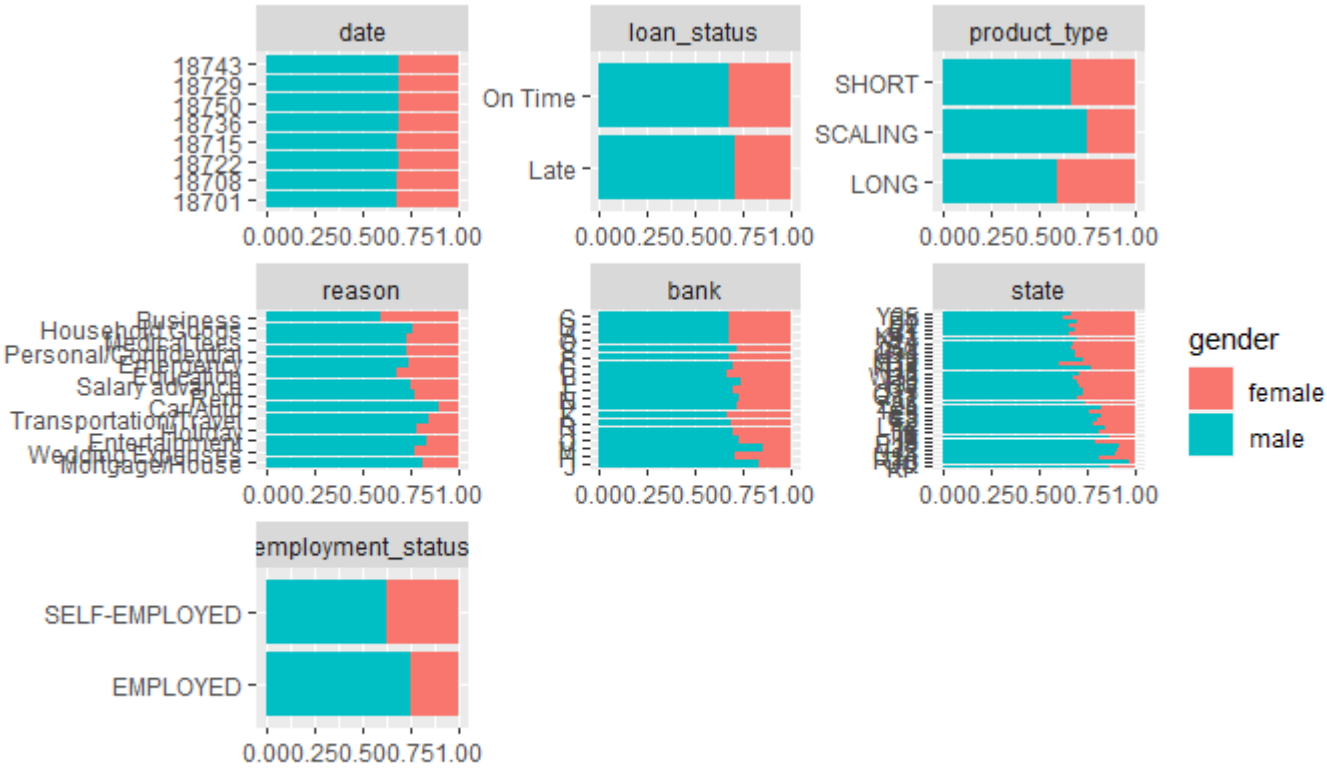
```
"C:/Program Files/RStudio/bin/pandoc/pandoc" +RTS -K512m -RTS "C:/Users/onokonkwo/Desktop/R cod
e/report.knit.md" --to html4 --from markdown+autolink_bare_uris+tex_math_single_backslash --outp
ut pandoc4680972805.html --lua-filter "C:\Users\onokonkwo\Documents\R\R-4.1.0\library\rmarkdown
\rmarkdown\lua\pagebreak.lua" --lua-filter "C:\Users\onokonkwo\Documents\R\R-4.1.0\library\rmark
down\rmarkdown\lua\latex-div.lua" --self-contained --variable bs3=TRUE --standalone --section-di
vs --table-of-contents --toc-depth 6 --template "C:\Users\onokonkwo\Documents\R\R-4.1.0\library
\rmarkdown\rmd\h\default.html" --no-highlight --variable highlightjs=1 --variable theme=yeti --i
nclude-in-header "C:\Users\ONOKON~1\AppData\Local\Temp\RtmpSitxZj\rmarkdown-str46804ba6456.html"
--mathjax --variable "mathjax-url:https://mathjax.rstudio.com/latest/MathJax.js?config=TeX-AMS-M
ML_HTMLorMML"
```

```
Output created: report.html
```

Hide

```
plot_bar(Cleandf, by = "gender")
```

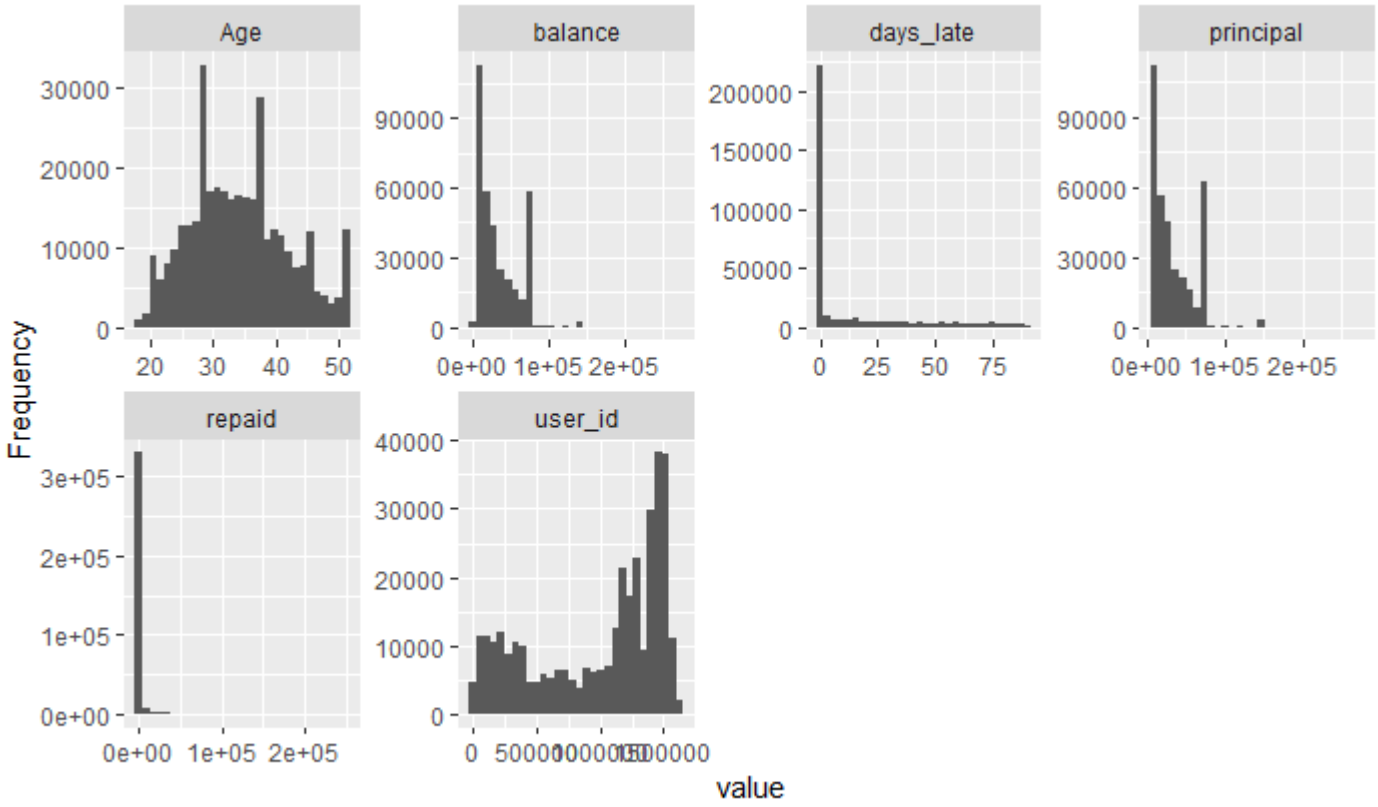
```
5 columns ignored with more than 50 categories.
loan_id: 108752 categories
approval_date: 173 categories
last_payment_date: 229 categories
default_date: 229 categories
date_of_birth: 11177 categories
```



Loan Status - more women paid on time than men. Product Type - men took more scaling loan product, while women took more of Long product. Reason - women took more loans for business and education, and men for car/auto and transport/travel. Employment Status - More self-employed women and employed men.

Hide

```
plot_histogram(Cleandf)
```

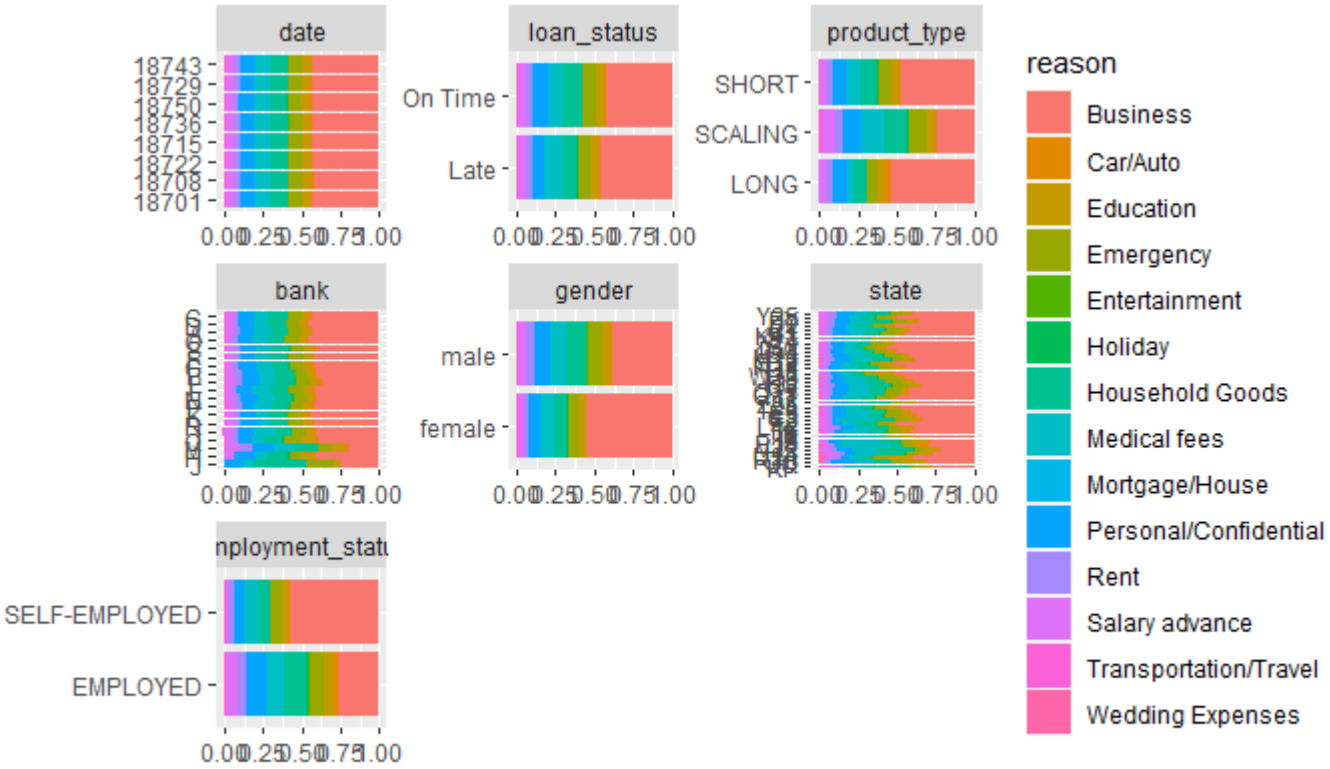


Age - most customers between 28 and 38

Hide

```
plot_bar(Cleandf, by = "reason")
```

```
5 columns ignored with more than 50 categories.
loan_id: 108752 categories
approval_date: 173 categories
last_payment_date: 229 categories
default_date: 229 categories
date_of_birth: 11177 categories
```



Product type - More long product taken for business reason Employment status - Self-employed customers took more business loans

Hide

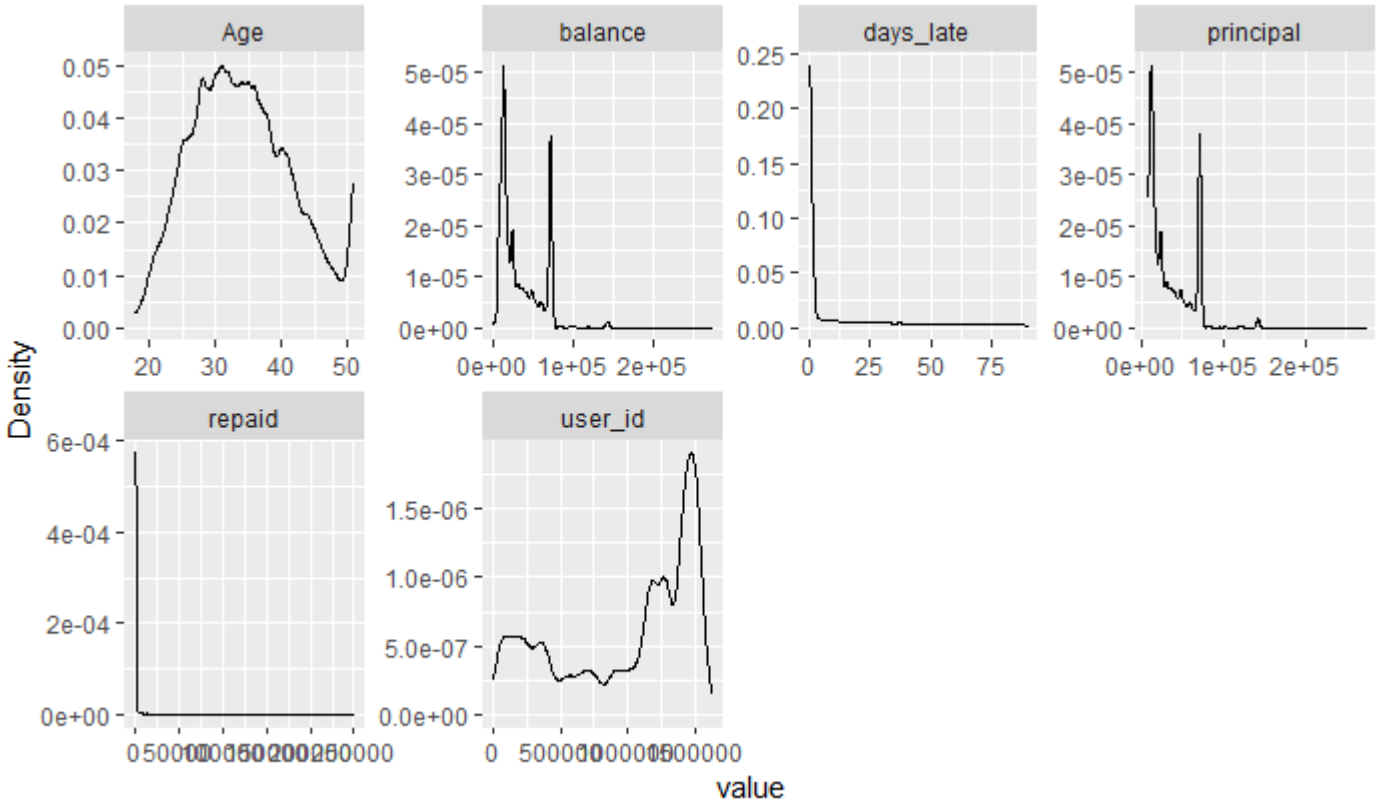
```
plot_bar(Cleandf)
```

```
5 columns ignored with more than 50 categories.
loan_id: 108752 categories
approval_date: 173 categories
last_payment_date: 229 categories
default_date: 229 categories
date_of_birth: 11177 categories
```

Loan status - more loans are paid on time. Gender - more than 50% more men than female customers. expand customer geographic by sex State - High concentration of customers in state Y25. expand customer geographics by location Bank - Majority of customers use bank C. consider partnerships with other banks

Hide

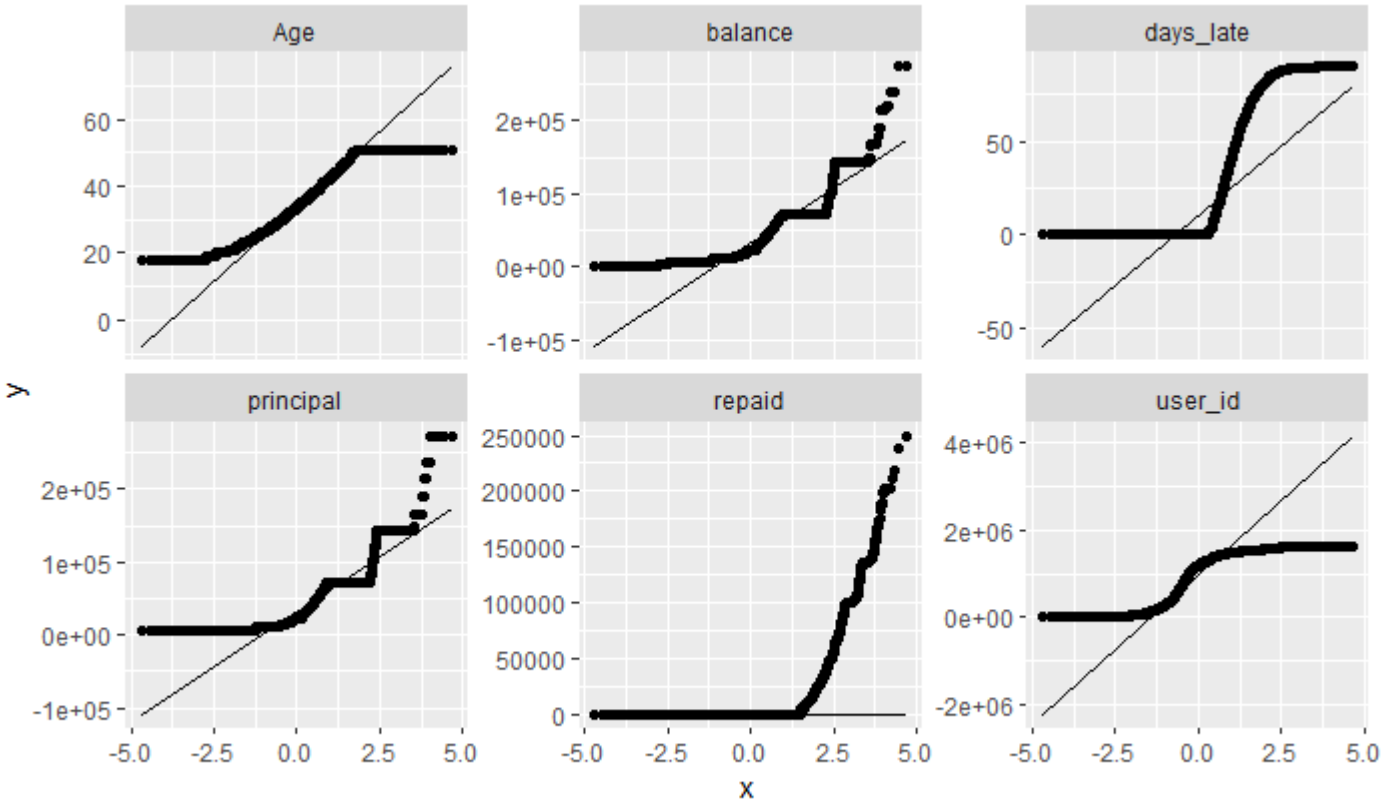
```
plot_density(Cleandf)
```



Age - customers between 25 and 40 Balance - balance owed is between 0 and 600,000. days late - between 0 and 7 days principal - loan amount collected is between 7,000 and 600,000 repaid - most repayment amount is less than 10,000

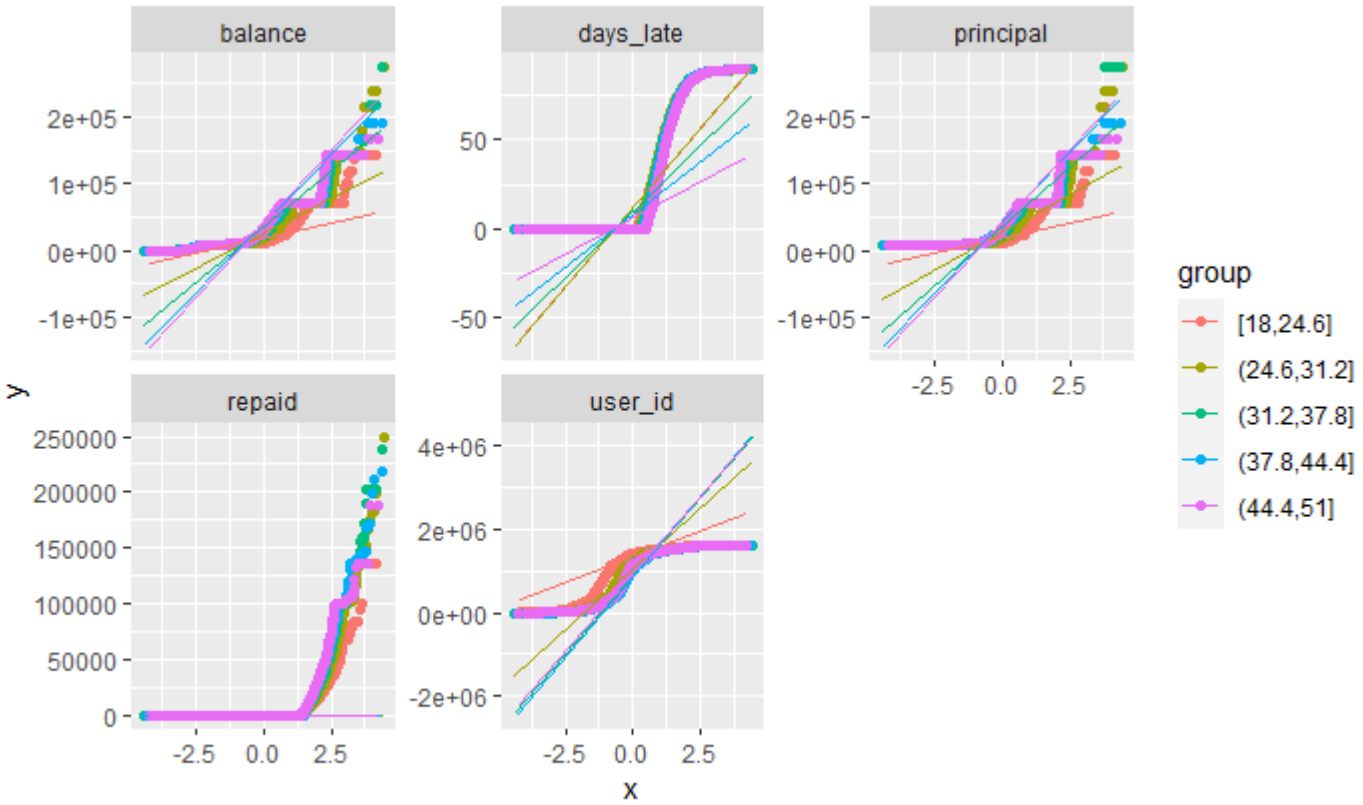
Hide

```
plot_qq(Cleandf)
```



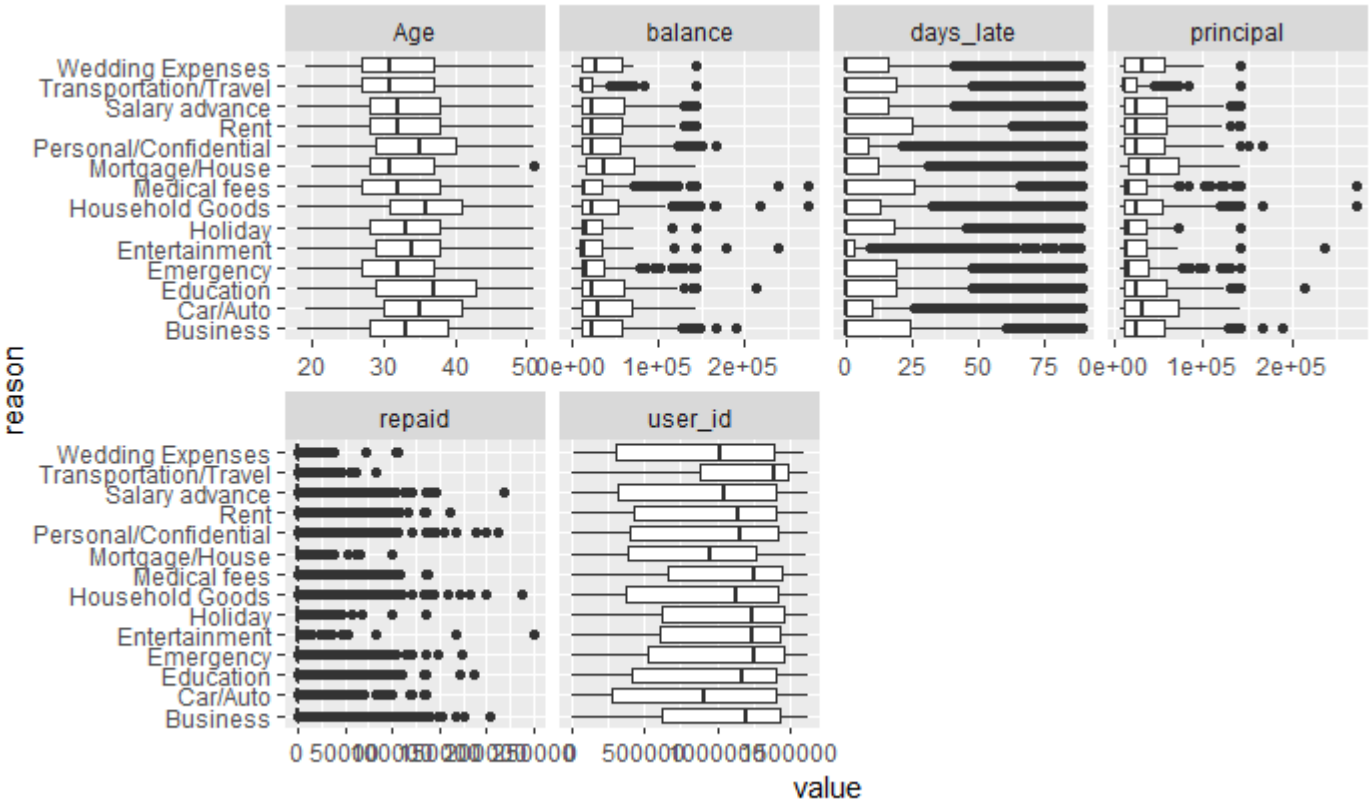
Hide

```
plot_qq(Cleandf, by = "Age")
```



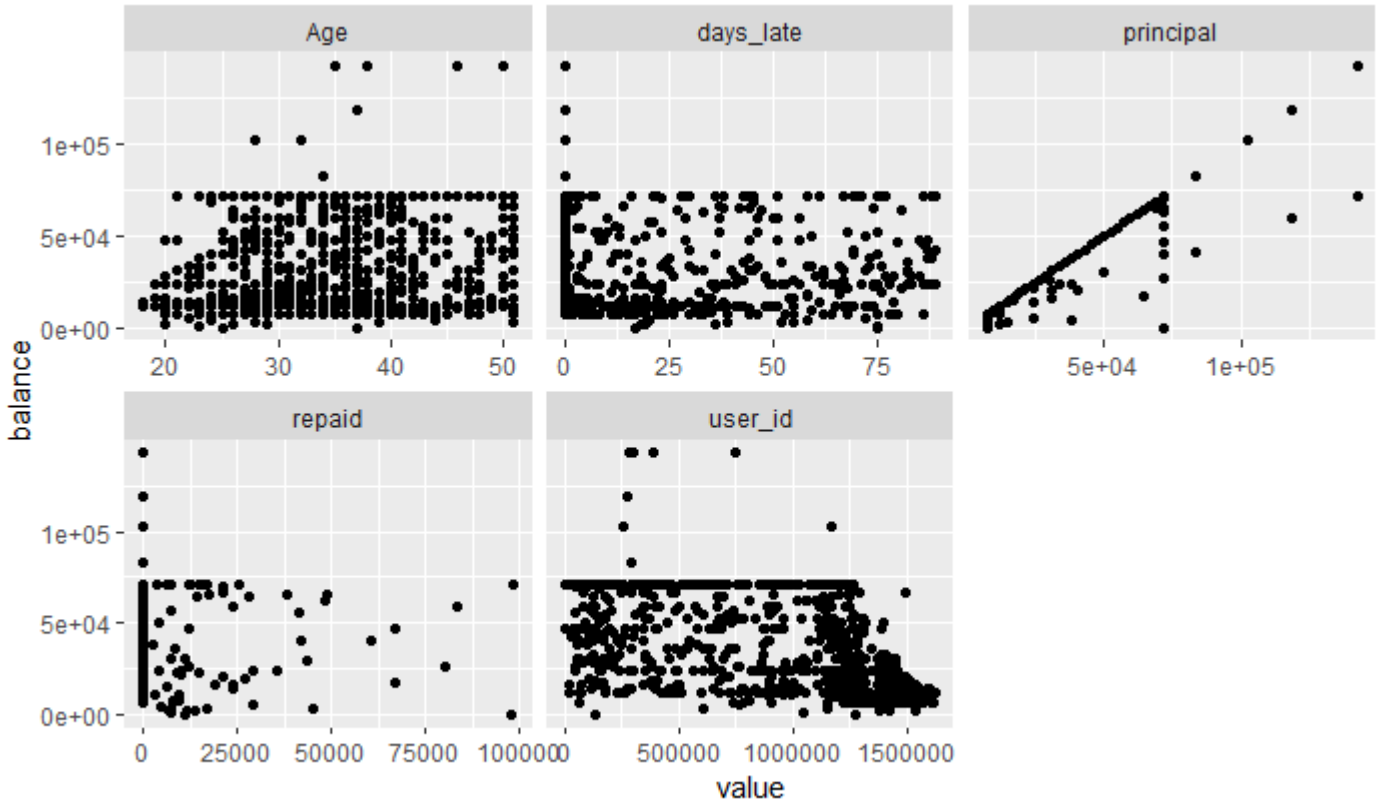
Hide

```
plot_boxplot(Cleandf, by = "reason")
```



Hide

```
plot_scatterplot(split_columns(Cleandf)$continuous, by = "balance", sampled_rows = 1000L)
```



Save as csv file

Hide

```
write.csv(Cleandf, "CleanCS.csv")
```

#GROSS LOAN PORTFOLIO Needs to be grouped by user ID, loan ID. Then sum unique values in Balance

Hide

```
df7 <- Cleandf %>%
group_by(date, loan_id)
```

Hide

```
length(unique(df7$balance))
```

```
[1] 3597
```

Hide

```
sum(unique(df7$balance))
```

```
[1] 103659135
```

NOTES

1. Total GLP =

2. Appears there's a disconnect between the loan amount repaid and the balance left for some customers.

#Portfolio At Risk (PAR) ANALYSIS and EVOLUTION (7, 15, 30, 60)

Portfolio-at-risk (PAR) ratio : Portfolio at risk (X days) / Gross loan portfolio Evolution (Per week?) 15/3, 22/3, 29/3, 5/4, 12/4, 19/4, 26/4, 3/5 (8 weeks in total)

Task - sum by week Group by Date (week) and loan ID

###Wk 1 - 15/3

Hide

```
Wk1 <- Cleandf %>%
  filter(date == "2021/03/15") %>%
  select(date, user_id, loan_id, principal, balance, days_late)%>%
  group_by(date, user_id, loan_id)
```

Hide

```
View(Wk1)
```

GLP of Week 1 - 1,375,404,576

Hide

```
sum(Wk1$balance)
```

```
[1] 1375404576
```

Hide

```
length(unique(Cleandf$user_id))
```

```
[1] 59988
```

EVOLUTION FOR WEEK 1

sum of balance when days late is 7 and above - 27.39%

Hide

```
Par7wk1 <- Wk1 %>%
  filter(days_late > 7)
```

Hide

```
sum(Par7wk1$balance)/1375404576
```

```
[1] 0.273875
```

sum of balance when days late is 15 and above - 23.28%

Hide

```
Par15wk1 <- Wk1 %>%
  filter(days_late > 15)
```

Hide

```
sum(Par15wk1$balance)/1375404576
```

```
[1] 0.2328037
```

sum of balance when days late is 30 and above 17.44%

Hide

```
Par30wk1 <- Wk1 %>%
  filter(days_late > 30)
```

Hide

```
sum(Par30wk1$balance)/1375404576
```

```
[1] 0.1744588
```

sum of balance when days late is 60 and above - 7.86%

Hide

```
Par60wk1 <- Wk1 %>%
  filter(days_late > 60)
```

Hide

```
sum(Par60wk1$balance)/1375404576
```

```
[1] 0.07856392
```

###Wk 2 - 22/3

Hide

```
Wk2 <- Cleandf %>%
  filter(date == "2021/03/22") %>%
  select(date, user_id, loan_id, principal, balance, days_late)%>%
  group_by(date, user_id, loan_id)
```

GLP Wk2 - 1,405,305,146

Hide

```
View(Wk2)
sum(Wk2$balance)
```

```
[1] 1405305146
```

EVOLUTION FOR WEEK 2

sum of balance when days late is 7 and above - 26.26%

Hide

```
Par7wk2 <- Wk2 %>%
  filter(days_late > 7)
```

Hide

```
sum(Par7wk2$balance)/1405305146
```

```
[1] 0.2625645
```

sum of balance when days late is 15 and above - 23.16%

Hide

```
Par15wk2 <- Wk2 %>%
  filter(days_late > 15)
```

Hide

```
sum(Par15wk2$balance)/1405305146
```

```
[1] 0.2316474
```

sum of balance when days late is 30 and above - 17.13%

Hide

```
Par30wk2 <- Wk2 %>%
  filter(days_late > 30)
```

Hide

```
sum(Par30wk2$balance)/1405305146
```

```
[1] 0.1712799
```

sum of balance when days late is 60 and above - 8.19%

Hide

```
Par60wk2 <- Wk2 %>%
  filter(days_late > 60)
```

Hide

```
sum(Par60wk2$balance)/1405305146
```

```
[1] 0.08185723
```

###Wk 3 - 29/3

Hide

```
Wk3 <- Cleandf %>%
  filter(date == "2021/03/29") %>%
  select(date, user_id, loan_id, principal, balance, days_late)%>%
  group_by(date, user_id, loan_id)
```

GLP wk 3 = 1,411,598,292

Hide

```
View(Wk3)
sum(Wk3$balance)
```

```
[1] 1411598292
```

WEEK 3 EVOLUTION

sum of balance when days late is 7 and above - 26.51%

Hide

```
Par7wk3 <- Wk3 %>%
  filter(days_late > 7)
```

Hide

```
sum(Par7wk3$balance)/1411598292
```

```
[1] 0.2650545
```

sum of balance when days late is 15 and above - 23.08%

Hide

```
Par15wk3 <- Wk3 %>%
  filter(days_late > 15)
```

Hide

```
sum(Par15wk3$balance)/1411598292
```

```
[1] 0.2307892
```

sum of balance when days late is 30 and above - 17.52%

Hide

```
Par30wk3 <- Wk3 %>%
  filter(days_late > 30)
```

Hide

```
sum(Par30wk3$balance)/1411598292
```

```
[1] 0.1751651
```

sum of balance when days late is 60 and above - 8.46%

Hide

```
Par60wk3 <- Wk3 %>%
  filter(days_late > 60)
```

Hide

```
sum(Par60wk3$balance)/1411598292
```

```
[1] 0.08464391
```

###Wk 4 - 5/4

Hide

```
Wk4 <- Cleandf %>%
  filter(date == "2021/04/05") %>%
  select(date, user_id, loan_id, principal, balance, days_late)%>%
  group_by(date, user_id, loan_id)
```

GLP = 1,414,476,977

Hide

```
View(Wk4)
sum(Wk4$balance)
```

```
[1] 1414476977
```

WEEK 4 EVOLUTION

sum of balance when days late is 7 and above - 26.40%

Hide

```
Par7wk4 <- Wk4 %>%
  filter(days_late > 7)
```

Hide

```
sum(Par7wk4$balance)/1414476977
```

```
[1] 0.2639984
```

sum of balance when days late is 15 and above - 22.92%

Hide

```
Par15wk4 <- Wk4 %>%
  filter(days_late > 15)
```

Hide

```
sum(Par15wk4$balance)/1414476977
```

```
[1] 0.2292069
```

sum of balance when days late is 30 and above - 17.87%

Hide

```
Par30wk4 <- Wk4 %>%
  filter(days_late > 30)
```

Hide

```
sum(Par30wk4$balance)/1414476977
```

```
[1] 0.1787441
```

sum of balance when days late is 60 and above - 8.18%

Hide

```
Par60wk4 <- Wk4 %>%
  filter(days_late > 60)
```

Hide

```
sum(Par60wk4$balance)/1414476977
```

```
[1] 0.08176224
```

###Wk 5 - 12/4

Hide

```
Wk5 <- Cleandf %>%
  filter(date == "2021/04/12") %>%
  select(date, user_id, loan_id, principal, balance, days_late)%>%
  group_by(date, user_id, loan_id)
```

GLP wk 5 = 1,458,110,474

Hide

```
View(Wk5)
sum(Wk5$balance)
```

```
[1] 1458110474
```

WEEK 5 EVOLUTION

sum of balance when days late is 7 and above - 27.42%

Hide

```
Par7wk5 <- Wk5 %>%
  filter(days_late > 7)
```

Hide

```
sum(Par7wk5$balance)/1458110474
```

```
[1] 0.2742133
```

sum of balance when days late is 15 and above - 22.75%

Hide

```
Par15wk5 <- Wk5 %>%
  filter(days_late > 15)
```

Hide

```
sum(Par15wk5$balance)/1458110474
```

```
[1] 0.2275141
```

sum of balance when days late is 30 and above - 17.63%

Hide

```
Par30wk5 <- Wk5 %>%
  filter(days_late > 30)
```

Hide

```
sum(Par30wk5$balance)/1458110474
```

```
[1] 0.1762583
```

sum of balance when days late is 60 and above - 8.09%

Hide

```
Par60wk5 <- Wk5 %>%
  filter(days_late > 60)
```

Hide

```
sum(Par60wk5$balance)/1458110474
```

```
[1] 0.08093415
```

###Wk 6 - 19/4

Hide

```
Wk6 <- Cleandf %>%
  filter(date == "2021/04/19") %>%
  select(date, user_id, loan_id, principal, balance, days_late)%>%
  group_by(date, user_id, loan_id)
```

GLP wk 6 - 1,452,277,510

Hide

```
View(Wk6)
sum(Wk6$balance)
```

```
[1] 1452277510
```

WEEK 6 EVOLUTION

sum of balance when days late is 7 and above - 27.53%

Hide

```
Par7wk6 <- Wk6 %>%
  filter(days_late > 7)
```

Hide

```
sum(Par7wk6$balance)/1452277510
```

```
[1] 0.2752921
```

sum of balance when days late is 15 and above - 23.56%

Hide

```
Par15wk6 <- Wk6 %>%
  filter(days_late > 15)
```

Hide

```
sum(Par15wk6$balance)/1452277510
```

```
[1] 0.2357595
```

sum of balance when days late is 30 and above - 17.05%

Hide

```
Par30wk6 <- Wk6 %>%
  filter(days_late > 30)
```

Hide

```
sum(Par30wk6$balance)/1452277510
```

```
[1] 0.170562
```

sum of balance when days late is 60 and above - 7.61%

Hide

```
Par60wk6 <- Wk6 %>%
  filter(days_late > 60)
```

Hide

```
sum(Par60wk6$balance)/1452277510
```

```
[1] 0.07613098
```

###Wk 7 - 26/4

Hide

```
Wk7 <- Cleandf %>%
  filter(date == "2021/04/26") %>%
  select(date, user_id, loan_id, principal, balance, days_late)%>%
  group_by(date, user_id, loan_id)
```

GLP wk 7 - 1,461,959,963

Hide

```
View(Wk7)
sum(Wk7$balance)
```

```
[1] 1461959963
```

WEEK 7 EVOLUTION

sum of balance when days late is 7 and above - 27.65%

Hide

```
Par7wk7 <- Wk7 %>%
  filter(days_late > 7)
```

Hide

```
sum(Par7wk7$balance)/1461959963
```

```
[1] 0.2765478
```

sum of balance when days late is 15 and above - 23.78%

Hide

```
Par15wk7 <- Wk7 %>%
  filter(days_late > 15)
```

Hide

```
sum(Par15wk7$balance)/1461959963
```

```
[1] 0.2377876
```

sum of balance when days late is 30 and above - 16.82%

Hide

```
Par30wk7 <- Wk7 %>%
  filter(days_late > 30)
```

Hide

```
sum(Par30wk7$balance)/1461959963
```

```
[1] 0.1682005
```

sum of balance when days late is 60 and above - 7.45%

Hide

```
Par60wk7 <- Wk7 %>%
  filter(days_late > 60)
```

Hide

```
sum(Par60wk7$balance)/1461959963
```

```
[1] 0.074472
```

###Wk 8 - 3/5

Hide

```
Wk8 <- Cleandf %>%
  filter(date == "2021/05/03") %>%
  select(date, user_id, loan_id, principal, balance, days_late)%>%
  group_by(date, user_id, loan_id)
```

GLP Wk 8 = 1,432,066,814

Hide

```
View(Wk8)
sum(Wk8$balance)
```

```
[1] 1432066814
```

WEEK 8 EVOLUTION

sum of balance when days late is 7 and above - 28.37%

Hide

```
Par7wk8 <- Wk8 %>%
  filter(days_late > 7)
```

Hide

```
sum(Par7wk8$balance)/1432066814
```

```
[1] 0.2836927
```

sum of balance when days late is 15 and above - 24.52%

Hide

```
Par15wk8 <- Wk8 %>%
  filter(days_late > 15)
```

Hide

```
sum(Par15wk8$balance)/1432066814
```

```
[1] 0.2452345
```

sum of balance when days late is 30 and above - 18.26%

Hide

```
Par30wk8 <- Wk8 %>%
  filter(days_late > 30)
```

Hide

```
sum(Par30wk8$balance)/1432066814
```

```
[1] 0.1826491
```

sum of balance when days late is 60 and above - 8.29%

Hide

```
Par60wk8 <- Wk8 %>%
  filter(days_late > 60)
```

Hide

```
sum(Par60wk8$balance)/1432066814
```

```
[1] 0.08294334
```

```
```