

Exercise solution of lesson 5

Hao Wang

3/8/2022

0. Load libraries

```
knitr::opts_chunk$set(echo = TRUE)

library(tidyverse)
library(rlang)

library(readxl)
```

4. Exercise

4.1 Exercise 1.

- Read the sheet **PBI New** and **SAS New** of Excel file: **Test PBI and SAS new and endorsement 20220214.xlsx** into R.

Solution:

```
PBI_new <- readxl::read_excel(path = "L:/BIRA/R Training/Test PBI and SAS new and endorsement 20220214.xlsx",
                             sheet = "PBI New") %>%
  as_tibble()
```

```
## New names:
## * ' ' -> ...14
## * ' ' -> ...15
## * ' ' -> ...16
```

```
str(PBI_new)
```

```
## tibble [344 x 17] (S3: tbl_df/tbl/data.frame)
## $ Policy Number           : chr [1:344] "PSG-05-078888E-000" "PSG-05-078889E-000" "PSG-05-078890E-000" ...
## $ Transaction Number      : num [1:344] 66331 66331 66331 66331 55119 ...
## $ Policy Issued Date       : chr [1:344] "Friday, November 12, 2021" "Friday, November 12, 2021" "Friday, November 12, 2021" ...
## $ Request Received Date    : chr [1:344] "Wednesday, November 10, 2021" "Wednesday, November 10, 2021" "Wednesday, November 10, 2021" ...
## $ MBC Originating Officer ID : num [1:344] 29 29 29 29 29 ...
## $ MBC Originating Officer First Name: chr [1:344] "Jeff" "Jeff" "Jeff" "Jeff" ...
```

```
## $ MBC Originating Officer Last Name : chr [1:344] "McQuaid" "McQuaid" "McQuaid" "McQuaid" ...
## $ Reporting Month : chr [1:344] "Tuesday, November 30, 2021" "Tuesday, November 30, 2021" ...
## $ Claims Expiry Date : POSIXct[1:344], format: "2026-08-30" "2022-09-26" ...
## $ Policy Created Date : POSIXct[1:344], format: "2021-08-25 11:16:20" "2021-08-25 11:16:20" ...
## $ DateDiff : num [1:344] 1 1 1 1 34 34 14 14 14 5 ...
## $ Quarter Year : chr [1:344] "Q4 2021" "Q4 2021" "Q4 2021" "Q4 2021" ...
## $ Policy MOA Date : chr [1:344] "Wednesday, November 10, 2021" "Wednesday, November 10, 2021" ...
## $ ...14 : logi [1:344] NA NA NA NA NA NA ...
## $ ...15 : logi [1:344] NA NA NA NA NA NA ...
## $ ...16 : logi [1:344] NA NA NA NA NA NA ...
## $ product : chr [1:344] "PSG" "PSG" "PSG" "PSG" ...
```

```
SAS_new <- readxl::read_excel(path = "L:/BIRA/R Training/Test PBI and SAS new and endorsement 20220214.1",
                             sheet = "SAS New ") %>%
  as_tibble()
str(SAS_new)
```

```
## tibble [355 x 45] (S3: tbl_df/tbl/data.frame)
## $ Month : num [1:355] 202112 202112 202112 202112 202112 ...
## $ CIB_Workcell : chr [1:355] NA NA NA NA ...
## $ Count_toward_SLA : chr [1:355] "Y" "Y" "Y" "Y" ...
## $ Cal_Days_New_Opportunity : num [1:355] 3 5 1 2 NA 3 1 1 1 1 ...
## $ Bus_Days_New_Opportunity : num [1:355] 3 3 1 2 NA 3 1 1 1 1 ...
## $ New_Opportunity_SLA : logi [1:355] NA NA NA NA NA NA ...
## $ New_Opportunity_SLA_Met : logi [1:355] NA NA NA NA NA NA ...
## $ Count_of_New_Opportunity_SLAs : num [1:355] 1 1 1 1 0 1 1 1 1 1 ...
## $ Cal_Days_In_Process : num [1:355] 1 1 2 1 1 1 1 1 1 5 ...
## $ Bus_Days_In_Process : num [1:355] 1 1 2 1 1 1 1 1 1 3 ...
## $ In_Process_SLA : num [1:355] 2 2 2 2 2 2 2 2 2 2 ...
## $ In_Process_SLA_Met : num [1:355] 1 1 1 1 1 1 1 1 1 0 ...
## $ Count_of_In_Process_SLAs : num [1:355] 1 1 1 1 1 1 1 1 1 1 ...
## $ Cal_Days_End_to_End : num [1:355] 3 5 2 2 NA 3 1 1 1 5 ...
## $ Bus_Days_End_to_End : num [1:355] 3 3 2 2 NA 3 1 1 1 3 ...
## $ End_to_End_Bus_Days_SLA : num [1:355] 3 3 3 3 3 3 3 3 3 3 ...
## $ End_to_End_SLA_Met : num [1:355] 1 1 1 1 NA 1 1 1 1 1 ...
## $ Count_of_End_to_End_SLAs : num [1:355] 1 1 1 1 0 1 1 1 1 1 ...
## $ MBC_RGN_DSC : chr [1:355] "UNKNOWN" "Ontario Region" "UNKNOWN" "Western Region" ...
## $ MBC_TRX_TYPE_DSC : chr [1:355] "Master Agreement - Account" "Master Agreement - Account" ...
## $ MBC_TRAN_NUM : chr [1:355] "90427" "56270" "90426" "92180" ...
## $ PolicyNo_Rev : chr [1:355] "PSG-05-079718-000" "PSG-05-079676-000" "PSG-05-079676-000" ...
## $ POL_CREATED_DATE : POSIXct[1:355], format: "2021-12-16" "2021-12-13" ...
## $ MOA_DATE : POSIXct[1:355], format: "2021-12-16" "2021-12-13" ...
## $ POL_PAYMENT_RECEIVED_DATE : POSIXct[1:355], format: NA "2021-12-15" ...
## $ POL_APPLICATION_DATE : logi [1:355] NA NA NA NA NA NA ...
## $ POL_MTIP_POLICY_ISSUANCE_DATE : POSIXct[1:355], format: NA "2021-12-16" ...
## $ PLCY_EXP_DT : POSIXct[1:355], format: "2024-06-05" "2022-01-25" ...
## $ CLM_EXP_DT : POSIXct[1:355], format: "2024-09-05" "2022-04-25" ...
## $ RQT_RCVD_DT : POSIXct[1:355], format: "2021-12-14" "2021-12-09" ...
## $ PLCY_STRT_DT : POSIXct[1:355], format: "2021-12-16" "2021-12-13" ...
## $ MBC_CNTRPRTY_TYPE_DSC : chr [1:355] "Coindemnitor" NA NA NA ...
## $ Exporter_CI_NUMBER : chr [1:355] "698793991" "214001902" "343361396" "710308741" ...
## $ Exporter_LEGAL_NAME : chr [1:355] "Azure Power Global Limited" "Norvic Maritime Holdings Ltd" ...
## $ Exporter_CUSTOMER_TYPE : chr [1:355] "IN" "Commercial Segment" "Strategic Segment" "Commercial Segment" ...
```

```
## $ Exporter_Region_Long      : chr [1:355] "xOther" "xOther" "xOther" "xOther" ...
## $ Exporter_COMPANY_SIZE    : chr [1:355] "L" "Large" "Large" "Large" ...
## $ Exporter_BUS_ORG_TEAM_CD : chr [1:355] "04" "03" "04" "05" ...
## $ Exporter_BUS_ORG_TEAM_NM : chr [1:355] "Infrastructure & Environment" "Light Manufacturing" ...
## $ Exporter_BUS_ORG_TEAM_NM_SH : chr [1:355] "INFR" "LTM" "INFR" "ICT" ...
## $ Exporter_Market_Sector_Alternate: chr [1:355] "INF" "LTM" "INF" "ICT" ...
## $ Exporter_Region_Short    : chr [1:355] NA NA NA NA ...
## $ CALENDAR_DATE            : POSIXct[1:355], format: "2022-01-31" "2022-01-31" ...
## $ CALENDAR_YEAR            : num [1:355] 2022 2022 2022 2022 2022 ...
## $ REPORTING_PERIOD         : num [1:355] 4 4 4 4 4 4 4 4 4 ...
```

- How many records in each tab?

```
nrow(PBI_new)
```

```
## [1] 344
```

```
nrow(SAS_new)
```

```
## [1] 355
```

- what's the min, max, and mean of column “DateDiff” in sheet **PBI New** and column “Bus_Days_End_to_End” in sheet **SAS New**.

```
PBI_new %>% summarise(min_DateDiff = min(DateDiff),
                      max_DateDiff = max(DateDiff),
                      mean_DateDiff = mean(DateDiff),
                      )
```

```
## # A tibble: 1 x 3
##   min_DateDiff max_DateDiff mean_DateDiff
##         <dbl>         <dbl>         <dbl>
## 1             1             34             2.92
```

```
SAS_new %>% summarise(min_Bus_Days_End_to_End = min(Bus_Days_End_to_End),
                      max_Bus_Days_End_to_End = max(Bus_Days_End_to_End),
                      mean_Bus_Days_End_to_End = mean(Bus_Days_End_to_End),
                      )
```

```
## # A tibble: 1 x 3
##   min_Bus_Days_End_to_End max_Bus_Days_End_to_End mean_Bus_Days_End_to_End
##               <dbl>               <dbl>               <dbl>
## 1                   NA                   NA                   NA
```

4.2 Exercise 2.

- Load the Denodo view bi_crm.Transfile_Company_Table_2022 in R.

```
library(odbc)

# setup the connection
con <- dbConnect(odbc::odbc(), "DenodoODBC", timeout = 30)

# query table using SQL
transfile_company_table <-
  dbGetQuery(con, 'SELECT * FROM bi_crm."Transfile_Company_Table_2022"') %>%
  as_tibble()
```

- sum of “Small Business Financial Customer Served Partnership Count”, and sum of “Medium Financial Customer Served Count”

```
transfile_company_table %>%
  summarise(`Total Small Business Financial Customer Served Partnership Count` =
    sum(`Small Business Financial Customer Served Partnership Count`, na.rm = T),
    `Total Medium Financial Customer Served Count` = sum(`Medium Financial Customer Served Count`, na.rm = T))
```

```
## # A tibble: 1 x 2
##   `Total Small Business Financial Customer Ser~` `Total Medium Financial Custome~`
##                                     <int64>                                     <int64>
## 1                                     3210                                     3024
```