Technical Test

Lauw Dastin Falentino Laurianto

2023-08-24

```
#Library Setup
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.2     v readr     2.1.4
## v forcats   1.0.0     v stringr   1.5.0
## v ggplot2   3.4.3     v tibble    3.2.1
## v lubridate 1.9.2     v tidyr     1.3.0
## v purrr     1.0.1
## -- Conflicts ---------------------------------------- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the ]8;;http://conflicted.r-lib.org/conflicted package]8;; to force all conflicts to become errors
```

```
library(rmarkdown)
```

```
#Read CSV
dataset_1 <- read.csv("Stockbit_Bibit_PST_dataset1.csv", header = TRUE)
dataset_2 <- read.csv("Stockbit_Bibit_PST_dataset2.csv", header = TRUE)
```

```
#Question 1
#Write a query that finds the top 3 users with most active (frequency) on buying for each group 17 - 22 and 23 -
30

##Group 17 - 22 Years Old
#Only take customer between 17 - 22 Years Old
group_17_22 <- subset(dataset_1, between(user_age, 17, 22))

#Getting the transaction data for 17 - 22 Years Old group
buy_freq_group_17_22 <- left_join(select(group_17_22, c("user_id")), dataset_2,
                                  by = "user_id")

#Checking if there is a transaction from this group of user
#Buying if it's a first transaction for the user or there is an increase on investment_amount
#Selling if there is a decrease on investment amount
#No transaction if there is no changes on the investment amount or it's NA
#Value 1 means buying
#Value 0 means no transaction or selling
i = 1
for (x in buy_freq_group_17_22$user_id) {
  if (i == 1) {
    if (is.na(buy_freq_group_17_22$Saham_invested_amount) == FALSE) {
      buy_freq_group_17_22$flag_Saham[i] <- 1
    } else {
      buy_freq_group_17_22$flag_Saham[i] <- 0
    }

    if (is.na(buy_freq_group_17_22$Pasar_Uang_invested_amount) == FALSE) {
      buy_freq_group_17_22$flag_PU[i] <- 1
    } else {
      buy_freq_group_17_22$flag_PU[i] <- 0
    }

    if (is.na(buy_freq_group_17_22$Pendapatan_Tetap_invested_amount) == FALSE) {
      buy_freq_group_17_22$flag_PT[i] <- 1
    } else {
      buy_freq_group_17_22$flag_PT[i] <- 0
    }

    if (is.na(buy_freq_group_17_22$Campuran_invested_amount) == FALSE) {
      buy_freq_group_17_22$flag_Campuran[i] <- 1
    } else {
      buy_freq_group_17_22$flag_Campuran[i] <- 0
    }
  } else if ((buy_freq_group_17_22$user_id[i] ==
              buy_freq_group_17_22$user_id[i-1]) == FALSE){
    if (is.na(buy_freq_group_17_22$Saham_invested_amount) == FALSE) {
      buy_freq_group_17_22$flag_Saham[i] <- 1
```

```
    } else {
      buy_freq_group_17_22$flag_Saham[i] <- 0
    }

    if (is.na(buy_freq_group_17_22$Pasar_Uang_invested_amount) == FALSE) {
      buy_freq_group_17_22$flag_PU[i] <- 1
    } else {
      buy_freq_group_17_22$flag_PU[i] <- 0
    }

    if (is.na(buy_freq_group_17_22$Pendapatan_Tetap_invested_amount) == FALSE) {
      buy_freq_group_17_22$flag_PT[i] <- 1
    } else {
      buy_freq_group_17_22$flag_PT[i] <- 0
    }

    if (is.na(buy_freq_group_17_22$Campuran_invested_amount) == FALSE) {
      buy_freq_group_17_22$flag_Campuran[i] <- 1
    } else {
      buy_freq_group_17_22$flag_Campuran[i] <- 0
    }
  } else {
    transaction_Saham <- buy_freq_group_17_22$Saham_invested_amount[i] - buy_freq_group_17_22$Saham_invested_amou
nt[i-1]

    transaction_PU <- buy_freq_group_17_22$Pasar_Uang_invested_amount[i] - buy_freq_group_17_22$Pasar_Uang_invest
ed_amount[i-1]

    transaction_PT <- buy_freq_group_17_22$Pendapatan_Tetap_invested_amount[i] - buy_freq_group_17_22$Pendapatan_
Tetap_invested_amount[i-1]

    transaction_Campuran <- buy_freq_group_17_22$Campuran_invested_amount[i] - buy_freq_group_17_22$Campuran_inve
sted_amount[i-1]

    if (is.na(transaction_Saham) == TRUE) {
      buy_freq_group_17_22$flag_Saham[i] <- 0
    } else if (transaction_Saham < 0) {
      buy_freq_group_17_22$flag_Saham[i] <- 0
    } else if (transaction_Saham > 0){
      buy_freq_group_17_22$flag_Saham[i] <- 1
    } else {
      buy_freq_group_17_22$flag_Saham[i] <- 0
    }

    if (is.na(transaction_PU) == TRUE) {
      buy_freq_group_17_22$flag_PU[i] <- 0
    } else if (transaction_PU < 0) {
      buy_freq_group_17_22$flag_PU[i] <- 0
    } else if (transaction_PU > 0){
      buy_freq_group_17_22$flag_PU[i] <- 1
    } else {
      buy_freq_group_17_22$flag_PU[i] <- 0
    }

    if (is.na(transaction_PT) == TRUE) {
      buy_freq_group_17_22$flag_PT[i] <- 0
    } else if (transaction_PT < 0) {
      buy_freq_group_17_22$flag_PT[i] <- 0
    } else if (transaction_PT > 0){
      buy_freq_group_17_22$flag_PT[i] <- 1
    } else {
      buy_freq_group_17_22$flag_PT[i] <- 0
    }

    if (is.na(transaction_Campuran) == TRUE) {
      buy_freq_group_17_22$flag_Campuran[i] <- 0
    } else if (transaction_Campuran < 0) {
      buy_freq_group_17_22$flag_Campuran[i] <- 0
    } else if (transaction_Campuran > 0){
      buy_freq_group_17_22$flag_Campuran[i] <- 1
    } else {
      buy_freq_group_17_22$flag_Campuran[i] <- 0
    }
  }

  buy_freq_group_17_22$buy_per_date[i] <- buy_freq_group_17_22$flag_Saham[i] + buy_freq_group_17_22$flag_PU[i] +
buy_freq_group_17_22$flag_PT[i] + buy_freq_group_17_22$flag_Campuran[i]

  i <- i + 1
```

```
}

#Finding the total buying frequency for group 17 - 22 Years Old
result_17_22 <- aggregate(buy_freq_group_17_22$buy_per_date,
                          list(buy_freq_group_17_22$user_id), FUN = sum)
colnames(result_17_22)[colnames(result_17_22) == "Group.1"] = "user_id"
colnames(result_17_22)[colnames(result_17_22) == "x"] = "buying_freq"
result_17_22 <- result_17_22[order(result_17_22$buying_freq,
                                   decreasing = TRUE), ]

#Showing the top 3 user from group 17 - 22 Years Old
head(left_join(result_17_22, dataset_1, by = "user_id"), 3)
```

1
2
3

3 rows | 1-1 of 10 columns

```
##Group 23 - 30 Years Old
#Only take customer between 23 - 30 Years Old
group_23_30 <- subset(dataset_1, between(user_age, 23, 30))

#Getting the transaction data for 23 - 30 Years Old group
buy_freq_group_23_30 <- left_join(select(group_23_30, c("user_id")), dataset_2,
                                  by = "user_id")

#Checking if there is a transaction from this group of user
#Buying if it's a first transaction for the user or there is an increase on investment_amount
#Selling if there is a decrease on investment amount
#No transaction if there is no changes on the investment amount or it's NA
#Value 1 means buying
#Value 0 means no transaction or selling
i = 1
for (x in buy_freq_group_23_30$user_id) {
  if (i == 1) {
    if (is.na(buy_freq_group_23_30$Saham_invested_amount) == FALSE) {
      buy_freq_group_23_30$flag_Saham[i] <- 1
    } else {
      buy_freq_group_23_30$flag_Saham[i] <- 0
    }

    if (is.na(buy_freq_group_23_30$Pasar_Uang_invested_amount) == FALSE) {
      buy_freq_group_23_30$flag_PU[i] <- 1
    } else {
      buy_freq_group_23_30$flag_PU[i] <- 0
    }

    if (is.na(buy_freq_group_23_30$Pendapatan_Tetap_invested_amount) == FALSE) {
      buy_freq_group_23_30$flag_PT[i] <- 1
    } else {
      buy_freq_group_23_30$flag_PT[i] <- 0
    }

    if (is.na(buy_freq_group_23_30$Campuran_invested_amount) == FALSE) {
      buy_freq_group_23_30$flag_Campuran[i] <- 1
    } else {
      buy_freq_group_23_30$flag_Campuran[i] <- 0
    }
  } else if ((buy_freq_group_23_30$user_id[i] ==
              buy_freq_group_23_30$user_id[i-1]) == FALSE){
    if (is.na(buy_freq_group_23_30$Saham_invested_amount) == FALSE) {
      buy_freq_group_23_30$flag_Saham[i] <- 1
    } else {
      buy_freq_group_23_30$flag_Saham[i] <- 0
    }

    if (is.na(buy_freq_group_23_30$Pasar_Uang_invested_amount) == FALSE) {
      buy_freq_group_23_30$flag_PU[i] <- 1
    } else {
      buy_freq_group_23_30$flag_PU[i] <- 0
    }

    if (is.na(buy_freq_group_23_30$Pendapatan_Tetap_invested_amount) == FALSE) {
```

```
      buy_freq_group_23_30$flag_PT[i] <- 1
    } else {
      buy_freq_group_23_30$flag_PT[i] <- 0
    }

    if (is.na(buy_freq_group_23_30$Campuran_invested_amount) == FALSE) {
      buy_freq_group_23_30$flag_Campuran[i] <- 1
    } else {
      buy_freq_group_23_30$flag_Campuran[i] <- 0
    }
  } else {
    transaction_Saham <- buy_freq_group_23_30$Saham_invested_amount[i] - buy_freq_group_23_30$Saham_invested_amou
nt[i-1]

    transaction_PU <- buy_freq_group_23_30$Pasar_Uang_invested_amount[i] - buy_freq_group_23_30$Pasar_Uang_invest
ed_amount[i-1]

    transaction_PT <- buy_freq_group_23_30$Pendapatan_Tetap_invested_amount[i] - buy_freq_group_23_30$Pendapatan_
Tetap_invested_amount[i-1]

    transaction_Campuran <- buy_freq_group_23_30$Campuran_invested_amount[i] - buy_freq_group_23_30$Campuran_inve
sted_amount[i-1]

    if (is.na(transaction_Saham) == TRUE) {
      buy_freq_group_23_30$flag_Saham[i] <- 0
    } else if (transaction_Saham < 0) {
      buy_freq_group_23_30$flag_Saham[i] <- 0
    } else if (transaction_Saham > 0){
      buy_freq_group_23_30$flag_Saham[i] <- 1
    } else {
      buy_freq_group_23_30$flag_Saham[i] <- 0
    }

    if (is.na(transaction_PU) == TRUE) {
      buy_freq_group_23_30$flag_PU[i] <- 0
    } else if (transaction_PU < 0) {
      buy_freq_group_23_30$flag_PU[i] <- 0
    } else if (transaction_PU > 0){
      buy_freq_group_23_30$flag_PU[i] <- 1
    } else {
      buy_freq_group_23_30$flag_PU[i] <- 0
    }

    if (is.na(transaction_PT) == TRUE) {
      buy_freq_group_23_30$flag_PT[i] <- 0
    } else if (transaction_PT < 0) {
      buy_freq_group_23_30$flag_PT[i] <- 0
    } else if (transaction_PT > 0){
      buy_freq_group_23_30$flag_PT[i] <- 1
    } else {
      buy_freq_group_23_30$flag_PT[i] <- 0
    }

    if (is.na(transaction_Campuran) == TRUE) {
      buy_freq_group_23_30$flag_Campuran[i] <- 0
    } else if (transaction_Campuran < 0) {
      buy_freq_group_23_30$flag_Campuran[i] <- 0
    } else if (transaction_Campuran > 0){
      buy_freq_group_23_30$flag_Campuran[i] <- 1
    } else {
      buy_freq_group_23_30$flag_Campuran[i] <- 0
    }
  }

  buy_freq_group_23_30$buy_per_date[i] <- buy_freq_group_23_30$flag_Saham[i] + buy_freq_group_23_30$flag_PU[i] +
buy_freq_group_23_30$flag_PT[i] + buy_freq_group_23_30$flag_Campuran[i]

  i <- i + 1
}

#Finding the total buying frequency for group 23 - 30 Years Old
result_23_30 <- aggregate(buy_freq_group_23_30$buy_per_date,
                          list(buy_freq_group_23_30$user_id), FUN = sum)
colnames(result_23_30)[colnames(result_23_30) == "Group.1"] = "user_id"
colnames(result_23_30)[colnames(result_23_30) == "x"] = "buying_freq"
result_23_30 <- result_23_30[order(result_23_30$buying_freq,
                                   decreasing = TRUE), ]

#Showing the top 3 user from group 23 - 30 Years Old
```

```
head(left_join(result_23_30, dataset_1, by = "user_id"), 3)
```

1
2
3

3 rows | 1-1 of 10 columns

```
#Question 2
#Write a query that finds the top 3 users with most active (frequency) on selling (Reksadana Saham Portfolio Only
) who are female and income source not from "Keuntungan Bisnis"

#Subset to only get the customer that meet the criteria
g_female_not_kb <- subset(dataset_1, user_gender == "Female")
g_female_not_kb <- subset(g_female_not_kb,
                          user_income_source != "Keuntungan Bisnis")

#Getting the transaction data of user from the requirement group
transaction_data_group <- left_join(select(g_female_not_kb, c("user_id")),
                                dataset_2, by = "user_id")

#Getting the only last transaction from user portofolio to know what investment product they ever have till the e
nd of the year.
dataset_2  <- dataset_2[order(dataset_2$date, decreasing = TRUE), ]
unique_dataset_2 <- select(dataset_2, -c("date"))
unique_dataset_2 <- distinct(unique_dataset_2, user_id, .keep_all = TRUE)

#Getting user portofolio data that have Reksadana Saham Only for an entire year
i = 1
for (x in unique_dataset_2$user_id) {
  if (is.na(unique_dataset_2$Saham_invested_amount[i]) == FALSE &&
        is.na(unique_dataset_2$Pasar_Uang_invested_amount[i]) == TRUE &&
        is.na(unique_dataset_2$Pendapatan_Tetap_invested_amount[i]) == TRUE &&
            is.na(unique_dataset_2$Campuran_invested_amount[i]) == TRUE) {

      unique_dataset_2$flag[i] <- "Reksadana Saham Portofolio Only"

  } else {
      unique_dataset_2$flag[i] <- "Not Reksadana Saham Portofolio Only"
  }

  i <- i + 1
}

user_saham_only <- subset(unique_dataset_2,
                          flag == "Reksadana Saham Portofolio Only",
                          select = c("user_id"))

#Combine all the requirement into one data frame
result <- left_join(user_saham_only,
                      select(transaction_data_group,
                             c("user_id", "date",
                               "Saham_AUM", "Saham_invested_amount")),
                    by = "user_id")

result <- drop_na(result)

#Checking if there is a sell transaction from this group of user
result  <- result[order(result$user_id, decreasing = FALSE), ]

i = 1
for (x in result$user_id) {
  if (i == 1) {
    result$flag[i] <- "Buying"
  } else if ((result$user_id[i] == result$user_id[i-1]) == FALSE){
    result$flag[i] <- "Buying"
  } else {
    transaction <- result$Saham_invested_amount[i] - result$Saham_invested_amount[i-1]
    if (transaction > 0) {
      result$flag[i] <- "Buying"
    } else if (transaction < 0) {
      result$flag[i] <- "Selling"
    } else {
      result$flag[i] <- "No Transaction"
```

```
    }
  }
  i <- i + 1
}

#Colleting all the Selling Transaction Data
result <- subset(result, flag == "Selling")
user_selling_freq <- count(result, user_id)
user_selling_freq  <- user_selling_freq[order(user_selling_freq$n,
                                             decreasing = TRUE), ]
colnames(user_selling_freq)[colnames(user_selling_freq) == "n"] = "selling_freq"

#Showing the top 3 Result
head(left_join(user_selling_freq, dataset_1, by = "user_id"), 3)
```

1
2
3

3 rows | 1-1 of 10 columns