# Question 1.

In ‘Results’ tab: Calculate Match/ No-Match Signals for each data source, for each Term defined in ‘Match Rules’ tab.  You can use 1 = Match and 0 = No-Match.

a. Calculate the Match Rate for each data source, for each Term.

## Solution: 1

As per the given match rule in the question –

Case 1:

Match (FirstName And LastName)

match.fullname <- function(row){

row %>%

content.match(c('FirstName', 'LastName')) %>%

all() %>%

return()

}

Case 2:

Match ((Address1 OR (HouseNumber And StreetName)) And (PostalCode Or City))

match.address <- function(row){

first <- row %>%

content.match(c('Address1')) %>%

all()

second <- row %>%

content.match(c('HouseNumber', 'StreetName')) %>%

all()

third <- row %>%

content.match(c('PostalCode', 'City')) %>%

any()

return((first || second) && third)

}

Case 3:

Match (DayOfBirth And MonthOfBirth And YearOfBirth)

match.DOB <- function(row){

row %>%

content.match(c('DayOfBirth', 'MonthOfBirth', 'YearOfBirth')) %>%

all() %>%

return()

}

Case 4:

Match (FullName And Address)

|  |
| --- |
| Match (FullName And Date Of Birth) |
| Match (FullName And (Address Or DateOfBirth)) |
| Match (FullName And Address And DateOfBirth) |

terms <- results %>%

merge(transactions) %>%

by\_row(..f = match.fullname, .to = 'FullName', .collate='rows') %>%

by\_row(..f = match.address, .to = 'Address', .collate='rows') %>%

by\_row(..f = match.DOB, .to = 'DateOfBirth', .collate='rows') %>%

mutate(`Name And Address` = FullName & Address,

`Name And DateOfBirth` = FullName & DateOfBirth,

`Name And (Address Or DateOfBirth)` = FullName & (Address | DateOfBirth),

`Name And Address And DateOfBirth` = FullName & Address & DateOfBirth

) %>%

select(c(15, 2, 18:24))

terms %>%

datatable()



temp <- terms %>%

select(-TransactionID) %>%

group\_by(DatasourceName) %>%

summarise\_all(mean) %>%

mutate\_if(is.numeric, ~round(., 2))

temp %>%

datatable()

# Question 2:

In ‘Overall’ tab: Calculate Match/NoMatch Signals for each term for each transaction.  A transaction is a “Match” for a given Term if 1 or more data sources returned a positive Match on that Term.  You can use 1 = Match and 0 = NoMatch.

a. Calculate the Match Rate over all transactions, for each Term.

## Solution:

As per the given condition, the required query will be –

terms %>%

select(-DatasourceName) %>%

group\_by(TransactionID) %>%

summarize\_all(any) %>%

datatable()

1. The query for the match rate over all the transactions is given below -

terms %>%

select(-DatasourceName) %>%

group\_by(TransactionID) %>%

summarise\_all(mean) %>%

mutate\_if(is.numeric, ~round(., 2)) %>%

datatable()

# Question 3

Rank the 4 data sources for optimization based on Maximizing Verification Rate and Minimizing Cost.

a. Verification Rule for optimization is defined as: Match on FullName And (Address Or DateOfBirth)

Costs:

1. Consumer = $0.64

2. Credit Agency = $0.88

3. Credit Agency 2 = $0.52

4. Credit Agency 3 = $0.52

4. Additional insights, findings and/or recommendations

## Solution:

terms %>%

select(DatasourceName, `Name And (Address Or DateOfBirth)`) %>%

group\_by(DatasourceName) %>%

summarize(matches = sum(`Name And (Address Or DateOfBirth)`)) %>%

mutate(cost.per.request = c(0.64, 0.88, 0.52, 0.52),

requests = rep(nrow(transactions), 4),

total.cost = cost.per.request \* requests,

cost.per.match = total.cost / matches) %>%

mutate\_if(is.numeric, ~round(., 2)) %>%

arrange(cost.per.match) %>%

datatable()

As I found the last case, we are changing the data then the result will get change so it will perform best if the data source will be better and much informative.