Vendor Prospects Testing

# Libraries

library(readxl)

## Warning: package 'readxl' was built under R version 4.1.2

library(dplyr)

##   
## 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

library(tidyr)  
library(writexl)

# Load the data

NACAC <- read\_excel("NACAC College Fair Lists Prospect Records.xlsx")  
College\_Board <- read\_excel("College Board Search Name Buy Prospect Records.xlsx")  
PCU <- read\_excel("PCU Prospect Records.xlsx")  
Niche <- read\_excel("Niche Lead Delivery Prospect Records.xlsx")  
Active\_Match <- read\_excel("Active Match + Prospect Records.xlsx")  
  
Vendor\_Prospects <- bind\_rows(NACAC,College\_Board,PCU,Niche,Active\_Match)

# Create Function to Split the Recd into date and time and sort the data. Then get rid of duplicates

TimeSplit <- function(DataFrame) {  
 Result <- DataFrame %>%  
 mutate(TimeRecd = substr(DateRecd,12,nchar(DateRecd))) %>%  
 mutate(DateRecd = as.Date(substr(DateRecd,1,10), "%m/%d/%Y")) %>%  
 arrange(Ref,DateRecd)  
 Result <- Result[!duplicated(Result$Ref), ]  
  
 return(Result)  
}

# Fixed the data

NACAC\_Clean <- TimeSplit(NACAC)  
College\_Board\_Clean <- TimeSplit(College\_Board)  
PCU\_Clean <- TimeSplit(PCU)  
Niche\_Clean <- TimeSplit(Niche)  
Active\_Match\_Clean <- TimeSplit(Active\_Match)   
  
Vendor\_Prospects\_Clean <- bind\_rows(NACAC\_Clean, College\_Board\_Clean, PCU\_Clean, Niche\_Clean, Active\_Match\_Clean) %>%  
 arrange(Ref,DateRecd)

# Exporting data

#write\_xlsx(Vendor\_Prospects\_Clean,"C:\\Users\\foilbn\\Documents\\Vendor\_Prospects.xlsx")

# Plan of action: Create five columns for five sources. Source 1 is the first vendor who reached out to the prospect

V <- Vendor\_Prospects\_Clean %>%  
 select(Ref,SourceFormat,DateRecd,TimeRecd)  
  
reshape(V, timevar = "SourceFormat",idvar = "Ref", direction = "wide")

## # A tibble: 235,299 x 3  
## Ref `DateRecd.c("College Board Searc~ `TimeRecd.c("College Board Searc~  
## <chr> <date> <chr>   
## 1 000001653 NA <NA>   
## 2 000005725 NA <NA>   
## 3 000006796 NA <NA>   
## 4 000008698 NA <NA>   
## 5 000008725 NA <NA>   
## 6 000011732 NA <NA>   
## 7 000018368 NA <NA>   
## 8 000020110 NA <NA>   
## 9 000024864 NA <NA>   
## 10 000027091 NA <NA>   
## # ... with 235,289 more rows

# Initial conditions  
  
A = 0  
I = 50  
# Create the source columns  
Vendor\_Prospects\_Clean <- Vendor\_Prospects\_Clean %>%  
 mutate(Source\_1 = NA, Source\_1\_Date = NA, Source\_2 = NA, Source\_2\_Date = NA, Source\_3 = NA, Source\_3\_Date = NA,Source\_4 = NA, Source\_4\_Date = NA,Source\_5 = NA, Source\_5\_Date = NA)  
Vendor\_final <- Vendor\_Prospects\_Clean  
Vendor\_final$SourceFormat <- NULL  
Vendor\_final$DateRecd <- NULL  
  
  
for (i in 1:nrow(Vendor\_Prospects\_Clean)) {  
 ID = Vendor\_Prospects\_Clean[i,1]  
 if (Vendor\_Prospects\_Clean[I,1] == ID) {  
 A = A + 1  
 if (A == 2) {  
 Vendor\_final[I,18] = Vendor\_Prospects\_Clean[i,3]  
 Vendor\_final[I,19] = Vendor\_Prospects\_Clean[i,4]  
 }else {  
 if (A ==3) {  
 Vendor\_final[I,20] = Vendor\_Prospects\_Clean[i,3]  
 Vendor\_final[I,21] = Vendor\_Prospects\_Clean[i,4]  
 }else {  
 if (A==4) {  
 Vendor\_final[I,22] = Vendor\_Prospects\_Clean[i,3]  
 Vendor\_final[I,23] = Vendor\_Prospects\_Clean[i,4]  
 }else {  
 Vendor\_final[I,24] = Vendor\_Prospects\_Clean[i,3]  
 Vendor\_final[I,25] = Vendor\_Prospects\_Clean[i,4]  
 }  
 }  
 }  
 } else {  
 ID = Vendor\_Prospects\_Clean[i,1]  
 A = 1  
 I = i # In order to remember the first instance of the Row ID  
 Vendor\_final[I,16] = Vendor\_Prospects\_Clean[I,3]  
 Vendor\_final[I,17] = Vendor\_Prospects\_Clean[I,4]  
 }  
   
# if (i%%5000 == 0) {  
# print(i)  
# }  
   
   
 }

# Fix the data so we only have the final rows

Vendor\_final <- Vendor\_final %>%  
 drop\_na(Source\_1)

# Exporting data

#write\_xlsx(Vendor\_final,"C:\\Users\\foilbn\\Documents\\Vendor\_final.xlsx")

# How many people have reoccuring Refs?

T <- Vendor\_Prospects\_Clean %>%  
 # filter(SourceFormat == "Niche Lead Delivery") %>%  
 group\_by(Ref) %>%  
 summarise(num\_vendors = n())  
 # arrange(desc(num\_vendors)) %>%  
# group\_by(num\_vendors) %>%  
 # summarise(Num\_occ = n())  
nrow(T)

## [1] 235299

# Individual search

Ind <- Vendor\_Prospects\_Clean %>%  
 filter(Ref == 037500275)   
   
Ind

## # A tibble: 0 x 27  
## # ... with 27 variables: Ref <chr>, RefreshDate <chr>, SourceFormat <chr>,  
## # DateRecd <date>, Entry Term <chr>, Prospect Date <chr>, Inquiry Date <chr>,  
## # App Submit Date <chr>, Sex <chr>, Active Region <chr>,  
## # Active US 5-digit ZIP Code <chr>, Active Geomarket <chr>,  
## # Intended Major <chr>, Major <chr>, Application Source <chr>,  
## # Distance from 01845 <chr>, TimeRecd <chr>, Source\_1 <lgl>,  
## # Source\_1\_Date <lgl>, Source\_2 <lgl>, Source\_2\_Date <lgl>, ...

Vendor\_Prospects\_Clean %>%  
 group\_by(Ref) %>%  
 summarise(frequancy = n()) %>%  
 arrange(desc(frequancy))

## # A tibble: 235,299 x 2  
## Ref frequancy  
## <chr> <int>  
## 1 037500275 4  
## 2 056690035 4  
## 3 056903263 4  
## 4 088854781 4  
## 5 105569666 4  
## 6 117046753 4  
## 7 123838629 4  
## 8 136661466 4  
## 9 143020417 4  
## 10 151316110 4  
## # ... with 235,289 more rows