Vendor Analysis

# 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

OP <- read\_excel("Vendor\_Prospects.xlsx")  
TP <- read\_excel("Transposed\_VP.xlsx")  
VP <- read\_excel("Full\_VP.xlsx")  
CD <- read\_excel("Noah - Connecting Data 20211020-094022.xlsx")  
Combined\_data <- left\_join(by = "Ref",VP,CD)

# Change specific columns into factors

Combined\_data <- Combined\_data %>%  
 mutate(Admit = recode(Admit,  
 "1" = 1,  
 "0" = 0)) %>%  
 mutate(Enrolled = recode(Enrolled,  
 "1" = 1,  
 "0" = 0))

# Create a binary column for if they applied and deposite date

Combined\_data<- Combined\_data %>%  
 mutate(Applied = ifelse(is.na(`App Submit Date`) == TRUE, 0,1)) %>%  
 mutate(Deposited = ifelse(is.na(`Deposit Date`) == TRUE, 0,1))

# Transposed update with enrollement, admit, applied

merging <- Combined\_data %>%  
 select(Ref,Enrolled,Admit,Applied,Deposited)  
TP\_full <- left\_join(TP,merging,by = c("Ref" = "Ref"))

# Display the combined data

head(Combined\_data)

## # A tibble: 6 x 28  
## Ref RefreshDate SourceFormat DateRecd `Entry Term`  
## <chr> <chr> <chr> <dttm> <chr>   
## 1 000001653 10/15/21 12:37:56 PM College Board~ 2020-01-24 00:00:00 Fall 2021   
## 2 000005725 10/15/21 12:37:56 PM College Board~ 2020-01-24 00:00:00 Fall 2021   
## 3 000006796 10/15/21 12:37:56 PM College Board~ 2020-01-24 00:00:00 Fall 2021   
## 4 000008698 10/15/21 12:37:56 PM College Board~ 2020-08-18 00:00:00 Fall 2021   
## 5 000008725 10/15/21 12:37:56 PM College Board~ 2020-08-18 00:00:00 Fall 2021   
## 6 000011732 10/15/21 12:37:56 PM College Board~ 2020-01-24 00:00:00 Fall 2021   
## # ... with 23 more variables: Prospect Date <chr>, Inquiry Date <chr>,  
## # App Submit Date <chr>, Sex <chr>, Active Region <chr>,  
## # Active US 5-digit ZIP Code.x <chr>, Active Geomarket <chr>,  
## # Intended Major <chr>, Major <chr>, Application Source <chr>,  
## # Distance from 01845 <chr>, TimeRecd <chr>, Deposit Date <chr>, Admit <dbl>,  
## # Enrolled <dbl>, Recalculated GPA <chr>, Filed FAFSA <chr>,  
## # Parent Adjusted Gross Income (PAGI) <lgl>, Ping - Total Count <dbl>, ...

# How many people didnt get admitted but enrolled?

Combined\_data %>%  
 mutate(no\_ap\_bt\_enrl = ifelse(Admit == 0 & Enrolled == 1, 1,0) ) %>%  
 group\_by(no\_ap\_bt\_enrl) %>%  
 summarise(frequancy = n())

## # A tibble: 3 x 2  
## no\_ap\_bt\_enrl frequancy  
## <dbl> <int>  
## 1 0 241980  
## 2 1 3  
## 3 NA 151

# There is three people who werent admitted but still enrolled?  
# 151 na's  
# I am going to filter out the outliers for my vendor analysis

# Summary

summary(Combined\_data)

## Ref RefreshDate SourceFormat   
## Length:242134 Length:242134 Length:242134   
## Class :character Class :character Class :character   
## Mode :character Mode :character Mode :character   
##   
##   
##   
##   
## DateRecd Entry Term Prospect Date   
## Min. :2017-09-26 00:00:00 Length:242134 Length:242134   
## 1st Qu.:2020-01-24 00:00:00 Class :character Class :character   
## Median :2020-08-18 00:00:00 Mode :character Mode :character   
## Mean :2020-05-30 00:35:17   
## 3rd Qu.:2020-08-18 00:00:00   
## Max. :2021-10-11 00:00:00   
##   
## Inquiry Date App Submit Date Sex Active Region   
## Length:242134 Length:242134 Length:242134 Length:242134   
## Class :character Class :character Class :character Class :character   
## Mode :character Mode :character Mode :character Mode :character   
##   
##   
##   
##   
## Active US 5-digit ZIP Code.x Active Geomarket Intended Major   
## Length:242134 Length:242134 Length:242134   
## Class :character Class :character Class :character   
## Mode :character Mode :character Mode :character   
##   
##   
##   
##   
## Major Application Source Distance from 01845 TimeRecd   
## Length:242134 Length:242134 Length:242134 Length:242134   
## Class :character Class :character Class :character Class :character   
## Mode :character Mode :character Mode :character Mode :character   
##   
##   
##   
##   
## Deposit Date Admit Enrolled Recalculated GPA   
## Length:242134 Min. :0.00000 Min. :0.00000 Length:242134   
## Class :character 1st Qu.:0.00000 1st Qu.:0.00000 Class :character   
## Mode :character Median :0.00000 Median :0.00000 Mode :character   
## Mean :0.03072 Mean :0.00386   
## 3rd Qu.:0.00000 3rd Qu.:0.00000   
## Max. :1.00000 Max. :1.00000   
## NA's :151 NA's :151   
## Filed FAFSA Parent Adjusted Gross Income (PAGI) Ping - Total Count  
## Length:242134 Mode:logical Min. : 0.000   
## Class :character NA's:242134 1st Qu.: 0.000   
## Mode :character Median : 0.000   
## Mean : 2.841   
## 3rd Qu.: 0.000   
## Max. :1383.000   
## NA's :151   
## Active US 5-digit ZIP Code.y On Campus Visit Applied   
## Length:242134 Length:242134 Min. :0.00000   
## Class :character Class :character 1st Qu.:0.00000   
## Mode :character Mode :character Median :0.00000   
## Mean :0.03563   
## 3rd Qu.:0.00000   
## Max. :1.00000   
##   
## Deposited   
## Min. :0.000000   
## 1st Qu.:0.000000   
## Median :0.000000   
## Mean :0.004113   
## 3rd Qu.:0.000000   
## Max. :1.000000   
##

# What is the yield per vendor? (Percent of students that apply that get enrolled)

# What is the melt per vendor? (Percent of students that deposited but didnt enroll)

Vendor\_properies <- Combined\_data %>%  
 filter(!(Admit == 0 & Enrolled == 1)) %>%  
 group\_by(SourceFormat) %>%  
 summarise(Yield = sum(Enrolled)/sum(Admit)\*100, Melt = 100 - sum(Enrolled)/sum(Deposited)\*100, prospects\_2\_applied = sum(Applied)/n()\*100)  
  
Vendor\_properies

## # A tibble: 5 x 4  
## SourceFormat Yield Melt prospects\_2\_applied  
## <chr> <dbl> <dbl> <dbl>  
## 1 Active Match + 12.8 6.51 67.6   
## 2 College Board Search Name Buy 11.0 5.72 1.83  
## 3 NACAC College Fair Lists 18.5 0 9.25  
## 4 Niche Lead Delivery 13.1 7.51 16.6   
## 5 PCU 17.3 5.92 12.0

# We need the # of times each source occurs

Total <- Combined\_data %>%  
 group\_by(SourceFormat) %>%  
 summarise(Total\_Contact = n())  
Total

## # A tibble: 5 x 2  
## SourceFormat Total\_Contact  
## <chr> <int>  
## 1 Active Match + 3258  
## 2 College Board Search Name Buy 222145  
## 3 NACAC College Fair Lists 336  
## 4 Niche Lead Delivery 8551  
## 5 PCU 7844

# First Contact?

TP1<- TP %>%   
 group\_by(Source\_1) %>%  
 summarise(First\_Contact = n())  
  
TP2 <- TP %>%  
 drop\_na(Source\_2) %>%  
 group\_by(Source\_2) %>%  
 summarise(Second\_Contact = n())  
  
TP3 <- TP %>%  
 drop\_na(Source\_3) %>%  
 group\_by(Source\_3) %>%  
 summarise(Third\_Contact = n())  
  
TP4 <- TP %>%  
 drop\_na(Source\_4) %>%  
 group\_by(Source\_4) %>%  
 summarise(Fourth\_Contact = n())  
  
TP5 <- TP %>%  
 drop\_na(Source\_5) %>%  
 group\_by(Source\_5) %>%  
 summarise(Fifth\_Contact = n())  
  
TP\_Final <- left\_join(by = c("Source\_1" = "Source\_2"),TP1,TP2)  
TP\_Final <- left\_join(by = c("Source\_1" = "Source\_3"),TP\_Final,TP3)  
TP\_Final <- left\_join(by = c("Source\_1" = "Source\_4"),TP\_Final,TP4)  
TP\_Final <- left\_join(by = c("Source\_1" = "SourceFormat"),TP\_Final,Total)  
#TP\_Final <- left\_join(by = c("Source\_1" = "Source\_5"),TP\_Final,TP5)  
  
TP\_Final[is.na(TP\_Final)] <- 0  
names(TP\_Final)[1] <- "Sources"  
  
TP\_Final

## # A tibble: 5 x 6  
## Sources First\_Contact Second\_Contact Third\_Contact Fourth\_Contact  
## <chr> <int> <int> <int> <int>  
## 1 Active Match + 1932 1230 93 3  
## 2 College Board Searc~ 221236 896 13 0  
## 3 NACAC College Fair ~ 330 6 0 0  
## 4 Niche Lead Delivery 4838 3400 297 16  
## 5 PCU 6963 690 161 30  
## # ... with 1 more variable: Total\_Contact <int>

# Download TP\_Final

# write\_xlsx(TP\_Final,"C:\\Users\\foilbn\\Documents\\Transposed\_Contact.xlsx")  
# write\_xlsx(Vendor\_properies,"C:\\Users\\foilbn\\Documents\\Vendor\_Yield.xlsx")

# How effective is each vendor.

IP <- TP\_Final %>%  
 group\_by(Sources) %>%  
 summarise(First\_Contact\_Percent = round(First\_Contact/Total\_Contact \*100,2),  
 Second\_Contact\_Percent = round(Second\_Contact/Total\_Contact\*100,2),  
 Third\_Contact\_Percent = round(Third\_Contact/Total\_Contact\*100,2),  
 Fourth\_Contact\_Percent = round(Fourth\_Contact/Total\_Contact\*100,2))  
IP

## # A tibble: 5 x 5  
## Sources First\_Contact\_Pe~ Second\_Contact\_~ Third\_Contact\_P~ Fourth\_Contact\_~  
## <chr> <dbl> <dbl> <dbl> <dbl>  
## 1 Active M~ 59.3 37.8 2.85 0.09  
## 2 College ~ 99.6 0.4 0.01 0   
## 3 NACAC Co~ 98.2 1.79 0 0   
## 4 Niche Le~ 56.6 39.8 3.47 0.19  
## 5 PCU 88.8 8.8 2.05 0.38

# How about only percents per contact?

ILP <- TP\_Final %>%  
 group\_by(Sources) %>%  
 summarise(Percent\_1C = round(First\_Contact/nrow(TP)\*100,2),  
 Percent\_2C = round(Second\_Contact/sum(TP\_Final$Second\_Contact),4)\*100,  
 Percent\_3C = round(Third\_Contact/sum(TP\_Final$Third\_Contact),4)\*100,  
 Percent\_4C = round(Fourth\_Contact/sum(TP\_Final$Fourth\_Contact),4)\*100,  
 Percent\_Total = round(Total\_Contact/sum(TP\_Final$Total\_Contact),4)\*100)  
ILP

## # A tibble: 5 x 6  
## Sources Percent\_1C Percent\_2C Percent\_3C Percent\_4C Percent\_Total  
## <chr> <dbl> <dbl> <dbl> <dbl> <dbl>  
## 1 Active Match + 0.82 19.8 16.5 6.12 1.35  
## 2 College Board Searc~ 94.0 14.4 2.3 0 91.7   
## 3 NACAC College Fair ~ 0.14 0.1 0 0 0.14  
## 4 Niche Lead Delivery 2.06 54.6 52.7 32.6 3.53  
## 5 PCU 2.96 11.1 28.6 61.2 3.24

# For tommorw look for duplicates? EX. Niche occurs multiple times in the vendor prospects. Look at that and see how many times a vendor sends us the same person. We paythem per vendor sometimes

# Duplicates?

Duplicates <- OP %>%  
 group\_by(Ref,SourceFormat) %>%  
 summarise(Frequancy = n()) %>%  
 arrange(desc(Frequancy)) %>%  
 filter(Frequancy != 1) %>%  
 mutate(Frequancy = Frequancy - 1) %>%  
 arrange(desc(Frequancy)) %>%  
 group\_by(SourceFormat) %>%  
 summarise(Total\_Duplicates = sum(Frequancy))

## `summarise()` has grouped output by 'Ref'. You can override using the `.groups` argument.

# Find a good way to diplay such data

T<- OP %>%  
 group\_by(SourceFormat) %>%  
 summarise(Total\_Occurences = n())

# Final

DP <- left\_join(by = c("SourceFormat" = "SourceFormat"),T,Duplicates)

# Percents?

DP <- DP %>%   
 mutate(Percent\_Duplicates = round(Total\_Duplicates/Total\_Occurences\*100,2))  
DP

## # A tibble: 5 x 4  
## SourceFormat Total\_Occurences Total\_Duplicates Percent\_Duplica~  
## <chr> <int> <dbl> <dbl>  
## 1 Active Match + 3278 20 0.61  
## 2 College Board Search Name Buy 223215 1070 0.48  
## 3 NACAC College Fair Lists 371 35 9.43  
## 4 Niche Lead Delivery 9223 672 7.29  
## 5 PCU 8007 163 2.04

# Look at ref 024297179

# Joining data

Duplicate\_Final <- left\_join(by = c("SourceFormat" = "Sources"),DP,IP)

# Download Duplicate\_Final

write\_xlsx(Duplicate\_Final,"C:\\Users\\foilbn\\Documents\\Duplicate\_Final.xlsx")  
write\_xlsx(IP,"C:\\Users\\foilbn\\Documents\\IP.xlsx")  
write\_xlsx(ILP,"C:\\Users\\foilbn\\Documents\\ILP.xlsx")

#Relationship between the amount of sources find a reference vs their enrollment rate. Ex: ppl who are found 4 times are more likely to enroll vs 3 vs 2 vs 1

Contacted\_Once <- TP\_full[is.na(TP\_full$Source\_2), ]  
C1 <- Contacted\_Once %>%  
 filter(!(Admit == 0 & Enrolled == 1)) %>%  
 drop\_na(Source\_1) %>%  
 summarise(Contact = "Contacted Once", Yield = sum(Enrolled)/sum(Admit)\*100, Melt = 100 - sum(Enrolled)/sum(Deposited)\*100, prospects\_2\_applied = sum(Applied)/n()\*100)  
  
Contacted\_Twice <- TP\_full[is.na(TP\_full$Source\_3), ]  
C2 <- Contacted\_Twice %>%  
 filter(!(Admit == 0 & Enrolled == 1)) %>%  
 drop\_na(Source\_2) %>%  
 summarise(Contact = "Contacted Twice", Yield = sum(Enrolled)/sum(Admit)\*100, Melt = 100 - sum(Enrolled)/sum(Deposited)\*100, prospects\_2\_applied = sum(Applied)/n()\*100)  
  
Contacted\_Thrice <- TP\_full[is.na(TP\_full$Source\_4), ]  
C3 <- Contacted\_Thrice %>%  
 filter(!(Admit == 0 & Enrolled == 1)) %>%  
 drop\_na(Source\_3) %>%  
 summarise(Contact = "Contacted Thrice", Yield = sum(Enrolled)/sum(Admit)\*100, Melt = 100 - sum(Enrolled)/sum(Deposited)\*100, prospects\_2\_applied = sum(Applied)/n()\*100)  
  
C4 <- TP\_full %>%  
 filter(!(Admit == 0 & Enrolled == 1)) %>%  
 drop\_na(Source\_4) %>%  
 summarise(Contact = "Contacted Frice", Yield = sum(Enrolled)/sum(Admit)\*100, Melt = 100 - sum(Enrolled)/sum(Deposited)\*100, prospects\_2\_applied = sum(Applied)/n()\*100)  
  
CT <- rbind(C1,C2,C3,C4)  
CT

## # A tibble: 4 x 4  
## Contact Yield Melt prospects\_2\_applied  
## <chr> <dbl> <dbl> <dbl>  
## 1 Contacted Once 11.2 4.28 1.77  
## 2 Contacted Twice 12.9 6.47 29.3   
## 3 Contacted Thrice 13.8 10 69.3   
## 4 Contacted Frice 25 9.09 93.9

# Download CT

#write\_xlsx(CT,"C:\\Users\\foilbn\\Documents\\Contact\_Efficency .xlsx")