#importing data

Attendee <- read.csv('C:/A3866/R 103/AttendeeData.csv')

Universe <- read.csv('C:/A3866/R 103/UniverseData.csv')

#filtering only physicians only from datasets

Attended\_Physicians <- subset(Attendee, SP\_flag==1 & At\_Class=="Physician")

Universe\_Physicans <- subset(Universe, At\_Class=="Physician")

#Physicians in control who have not attended Speaker Program

library(data.table)

Universe\_Control<-subset(Universe\_Physicans, !Universe\_Physicans$At\_ID %in% Attended\_Physicians$At\_ID)

#ID which has 3 months sales and pre-sales

library(dplyr)

id\_sales<-Attended\_Physicians%>%filter(Month>3 & Month<22)%>%select(At\_ID)

Attendee\_new<-Attended\_Physicians[Attended\_Physicians$At\_ID %in% id\_sales$At\_ID,]

Attendee\_new<-Attendee\_new%>%arrange(At\_ID,Month)

#Attendee Phyisicans First Occurance

Attendee\_test <- Attendee\_new[match(unique(Attendee\_new$At\_ID), Attendee\_new$At\_ID),]

write.csv(Attendee\_test, 'C:/A3866/R 103/Attendee\_test.csv')

write.csv(Universe\_Control, 'C:/A3866/R 103/Universe\_Control.csv')

#uniquely specifying Columns due to error that may come while merging

colnames(Universe\_Control)=paste("Universe",colnames(Universe\_Control),sep="\_")

master=merge(Attendee\_test,Universe\_Control,by.x=c("At\_State","Month"),by.y = c("Universe\_At\_State","Universe\_Month"))

master=master%>%

select(-c(Universe\_At\_Class,At\_Class,SP\_flag))%>%arrange(At\_ID,Universe\_At\_ID,Month)

#replicating weights

master$weights=rep(c(4,2,1))

#Euclidean Distance

master$Euclidean\_Distance=0

master=master%>%

mutate(Euclidean\_Distance=(Sales-Universe\_Sales)^2\*weights)

matching\_phy\_id=master%>%group\_by(At\_ID,Universe\_At\_ID)%>%summarise(Distance=sqrt(sum(Euclidean\_Distance)),pre\_test\_sales=sum(Sales),pre\_Universe\_sales=sum(Universe\_Sales))

matching\_phy\_id=matching\_phy\_id%>%select(At\_ID,Universe\_At\_ID,Distance,pre\_Universe\_sales,pre\_test\_sales)%>%slice\_min(Distance)%>%slice\_max(pre\_Universe\_sales)

matched\_pair\_id = distinct(matching\_phy\_id,matching\_phy\_id$At\_ID,.keep\_all = TRUE)

matched\_pair\_id=matched\_pair\_id[,-6]

write.csv(matched\_pair\_id, 'C:/A3866/R 103/pairs.csv')

##Merging to find out the sales after promotion

master\_post=merge(Attendee\_test,Universe\_Control,by.x=c("At\_State","Month"),by.y = c("Universe\_At\_State","Universe\_Month"))

master\_post=master\_post%>%

select(-c(Universe\_At\_Class,At\_Class,SP\_flag,At\_State))%>%arrange(At\_ID,Universe\_At\_ID,Month)

IDs\_new=master\_post[master\_post$Universe\_At\_ID %in% matched\_pair\_id$Universe\_At\_ID,]

##pos-sales

Sum\_sales=IDs\_new%>%group\_by(Universe\_At\_ID,At\_ID)%>%summarise(post\_universe\_sales=sum(Universe\_Sales))

Matched\_Pairs=inner\_join(matched\_pair\_id,Sum\_sales,by=c("At\_ID","Universe\_At\_ID"), copy=TRUE)

post\_tes\_sales=Attendee\_test%>%group\_by(At\_ID)%>%summarise(post\_tes\_sales=sum(Sales))

Matched\_Pairs$post\_test\_sale=post\_tes\_sales$post\_tes\_sales

write.csv(Matched\_Pairs,'C:/A3866/R 103/Matched\_Pairs\_Vaibhav L\_A3866.csv')

#ROI Calculation

ROI=(((((sum(Matched\_Pairs$post\_test\_sale))-(sum(Matched\_Pairs$post\_universe\_sales))-(sum(Matched\_Pairs$pre\_test\_sales))-(sum(Matched\_Pairs$pre\_Universe\_sales)))\*5000)/500000)-1)\*100

paste("The ROI is is",ROI,"%")