Analisando o Resultado de Uma Campanha de Marketing - Versão 3

Data Science Academy

03 de Agosto, 2018

I. Problema de Negócio

A Agência de Viagens Borboleta Feliz possui os dados de todos os clientes que receberam campanhas de Marketing da empresa, mas que não comparam um pacote de férias.

Problema de negócio: Devemos continuar investindo nesses clientes (retarget)?

```
abandoned_data$Test  <- NA
abandoned_data$Test[abandoned_data$Test_Control == "test"] <- 1
abandoned_data$Test[abandoned_data$Test_Control == "control"] <- 0

summary(abandoned_data$Test)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.0000  0.0000  1.0000  0.5053  1.0000  1.0000
```

Standard Deviation = 0.5000012
q5 = 0
q95 = 1

```
abandoned_data$Has_State <- 0
abandoned_data$Has_State[abandoned_data$Address != ""] <- 1

summary(abandoned_data$Test[abandoned_data$Has_State == 1])
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.0000  0.0000  1.0000  0.5134  1.0000  1.0000
```

Standard Deviation = 0.4998865
q5 = 0
q95 = 1

II. Data Matching

```
abandoned_data[abandoned_data == ""] <- NA
reservation_data[reservation_data == ""] <- NA

# Email Matches
Email_Matches_Abandoned <- ifelse(!is.na(abandoned_data$Email),abandoned_data$Email %in% reservation_da

# Contact Phone Matches
ContactPhone_Matches_Abandoned <- ifelse(!is.na(abandoned_data$Contact_Phone),abandoned_data$Contact_Ph

# Last Name, Incoming Phone Matches
LastName_Incoming_Matches_Abandoned <- ifelse(!is.na(abandoned_data$Last_Name) & !is.na(abandoned_data$

# First Name, Last Name, Zip Matches
Names_Zip_Matches_Abandoned <- ifelse((!is.na(abandoned_data$First_Name) & !is.na(abandoned_data$Last_Na

# Combine all Matches
```

```
All_Matches_Abandoned = Email_Matches_Abandoned | ContactPhone_Matches_Abandoned | LastName_Incoming_Ma
abandoned_data_matches <- abandoned_data[All_Matches_Abandoned,]

# Remove Duplicates based on the keys
abandoned_data_matches <- abandoned_data_matches[!duplicated(abandoned_data_matches[,c("Email")],incompa
abandoned_data_matches <- abandoned_data_matches[!duplicated(abandoned_data_matches[,c("Contact_Phone")]
abandoned_incoming_dup <- duplicated(abandoned_data_matches[,c("Incoming_Phone")],incomparables = NA)
abandoned_lastname_dup <- duplicated(abandoned_data_matches[,c("Last_Name")],incomparables = NA)
abandoned_firstname_dup <- duplicated(abandoned_data_matches[,c("First_Name")],incomparables = NA)
abandoned_zipcode_dup <- duplicated(abandoned_data_matches[,c("Zipcode")],incomparables = NA)
abandoned_data_matches <- abandoned_data_matches[!(abandoned_incoming_dup & abandoned_lastname_dup),]
abandoned_data_matches <- abandoned_data_matches[!(abandoned_firstname_dup & abandoned_lastname_dup & al

# Store Outcome in original dataset
abandoned_data$Outcome <- 0
abandoned_data$Outcome[as.numeric(row.names(abandoned_data_matches))] <- 1
```

```
library(knitr)
treatments <- nrow(abandoned_data[abandoned_data$Test == 1,])
controls <- nrow(abandoned_data[abandoned_data$Test == 0,])
treatment_buy <- length(abandoned_data$Outcome[abandoned_data$Outcome == 1 & abandoned_data$Test == 1])
treatment_nobuy <- length(abandoned_data$Outcome[abandoned_data$Outcome == 0 & abandoned_data$Test == 1]
control_buy <- length(abandoned_data$Outcome[abandoned_data$Outcome == 1 & abandoned_data$Test == 0])
control_nobuy <- length(abandoned_data$Outcome[abandoned_data$Outcome == 0 & abandoned_data$Test == 0])

conv_rate_treatment <- treatment_buy/treatments*100
conv_rate_control <- control_buy/controls*100
cross_tab <- data.frame(treatment_buy,treatment_nobuy,control_buy,control_nobuy)
kable(cross_tab)
```

treatment_buy	treatment_nobuy	control_buy	control_nobuy
181	4085	42	4134

Conversion Rate for Treatment Group is 4.2428504 %.
Conversion Rate for Control Group is 1.0057471 %.

State: New York

```
#NY
treatment_buy_NY <- length(abandoned_data$Outcome[abandoned_data$Outcome == 1 & abandoned_data$Test ==
treatment_nobuy_NY <- length(abandoned_data$Outcome[abandoned_data$Outcome == 0 & abandoned_data$Test =
control_buy_NY <- length(abandoned_data$Outcome[abandoned_data$Outcome == 1 & abandoned_data$Test == 0 &
control_nobuy_NY <- length(abandoned_data$Outcome[abandoned_data$Outcome == 0 & abandoned_data$Test ==
cross_tab_NY <- data.frame(treatment_buy_NY,treatment_nobuy_NY,control_buy_NY,control_nobuy_NY)
kable(cross_tab_NY)
```

treatment_buy_NY	treatment_nobuy_NY	control_buy_NY	control_nobuy_NY
64	2285	16	2341

State: Ohio

```
#OH
treatment_buy_OH <- length(abandoned_data$Outcome[abandoned_data$Outcome == 1 & abandoned_data$Test ==
treatment_nobuy_OH <- length(abandoned_data$Outcome[abandoned_data$Outcome == 0 & abandoned_data$Test ==
control_buy_OH <- length(abandoned_data$Outcome[abandoned_data$Outcome == 1 & abandoned_data$Test == 0 &
control_nobuy_OH <- length(abandoned_data$Outcome[abandoned_data$Outcome == 0 & abandoned_data$Test == 0
cross_tab_OH <- data.frame(treatment_buy_OH,treatment_nobuy_OH,control_buy_OH,control_nobuy_OH)
kable(cross_tab_OH)
```

treatment_buy_OH	treatment_nobuy_OH	control_buy_OH	control_nobuy_OH
64	2295	15	2345

State: Arizona

```
#AZ
treatment_buy_AZ <- length(abandoned_data$Outcome[abandoned_data$Outcome == 1 & abandoned_data$Test ==
treatment_nobuy_AZ <- length(abandoned_data$Outcome[abandoned_data$Outcome == 0 & abandoned_data$Test ==
control_buy_AZ <- length(abandoned_data$Outcome[abandoned_data$Outcome == 1 & abandoned_data$Test == 0 &
control_nobuy_AZ <- length(abandoned_data$Outcome[abandoned_data$Outcome == 0 & abandoned_data$Test == 0
cross_tab_AZ <- data.frame(treatment_buy_AZ,treatment_nobuy_AZ,control_buy_AZ,control_nobuy_AZ)
kable(cross_tab_AZ)
```

treatment_buy_AZ	treatment_nobuy_AZ	control_buy_AZ	control_nobuy_AZ
63	2300	16	2349

State: Illinois

```
#IL
treatment_buy_IL <- length(abandoned_data$Outcome[abandoned_data$Outcome == 1 & abandoned_data$Test ==
treatment_nobuy_IL <- length(abandoned_data$Outcome[abandoned_data$Outcome == 0 & abandoned_data$Test ==
control_buy_IL <- length(abandoned_data$Outcome[abandoned_data$Outcome == 1 & abandoned_data$Test == 0 &
control_nobuy_IL <- length(abandoned_data$Outcome[abandoned_data$Outcome == 0 & abandoned_data$Test == 0
cross_tab_IL <- data.frame(treatment_buy_IL,treatment_nobuy_IL,control_buy_IL,control_nobuy_IL)
kable(cross_tab_IL)
```

treatment_buy_IL	treatment_nobuy_IL	control_buy_IL	control_nobuy_IL
62	2284	15	2353

State: California

```
#CA
treatment_buy_CA <- length(abandoned_data$Outcome[abandoned_data$Outcome == 1 & abandoned_data$Test ==
treatment_nobuy_CA <- length(abandoned_data$Outcome[abandoned_data$Outcome == 0 & abandoned_data$Test ==
control_buy_CA <- length(abandoned_data$Outcome[abandoned_data$Outcome == 1 & abandoned_data$Test == 0 &
control_nobuy_CA <- length(abandoned_data$Outcome[abandoned_data$Outcome == 0 & abandoned_data$Test == 0
cross_tab_CA <- data.frame(treatment_buy_CA,treatment_nobuy_CA,control_buy_CA,control_nobuy_CA)
kable(cross_tab_CA)
```

treatment_buy_CA	treatment_nobuy_CA	control_buy_CA	control_nobuy_CA
64	2293	15	2343

III. Data Cleaning

```
reservation_email_matches <- match(abandoned_data_matches$Email, reservation_data$Email, nomatch = 0, in
reservation_phone_matches <- match(abandoned_data_matches$Contact_Phone, reservation_data$Contact_Phone
reservation_name_incoming_matches <- ifelse(!is.na(abandoned_data_matches$Last_Name) & !is.na(abandoned_
reservation_name_zip_matches <- ifelse(!is.na(abandoned_data_matches$First_Name) & !is.na(abandoned_data
reservation_all_matches <- reservation_email_matches
reservation_all_matches <- ifelse(reservation_all_matches == 0,reservation_all_matches+reservation_phone
reservation_all_matches <- ifelse(reservation_all_matches == 0,reservation_all_matches+reservation_name_
reservation_all_matches <- ifelse(reservation_all_matches == 0,reservation_all_matches+reservation_name_
abandoned_all_matches <- as.numeric(row.names(abandoned_data_matches))

abandoned_data$Days_in_between <- 200
abandoned_data$Days_in_between[abandoned_data$Outcome == 1] <- as.numeric(as.Date(reservation_data$Sessi

cleaned_abandoned_data <- data.frame(c(1:nrow(abandoned_data)),abandoned_data$Test,abandoned_data$Outcom
colnames(cleaned_abandoned_data) <- c("Customer_ID","Test_Variable","Outcome","Days_in_Between","State")

write.csv(cleaned_abandoned_data,file = "cleaned_abandoned_data.csv")
```

IV. Statistical Analysis

Model-1: Outcome = alpha + beta * Test Variable + error*

```
lmodel1 <- lm(cleaned_abandoned_data$Outcome ~ cleaned_abandoned_data$Test_Variable)
kable(summary(lmodel1)$coef, digits=3)
```

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	0.010	0.002	4.073	0
cleaned_abandoned_data$Test_Variable	0.032	0.003	9.319	0

Outcome = 0.01 + 0.032 * Test_Variable + 0.002
Adjusted R-squared = 0.0100679

```
abandoned_data$Has_Email <- 0
abandoned_data$Has_Email[!is.na(abandoned_data$Email)] <- 1
```

Model-2: Outcome = alpha + beta1 * Test Variable + beta2 * Has Email * beta3 * Has_State + error*

```
lmodel2 <- lm(cleaned_abandoned_data$Outcome ~ cleaned_abandoned_data$Test_Variable + abandoned_data$Has
kable(summary(lmodel2)$coef, digits = 3)
```

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-0.002	0.003	-0.635	0.525
cleaned_abandoned_data$Test_Variable	0.031	0.003	9.034	0.000
abandoned_data$Has_Email	0.049	0.005	8.997	0.000
abandoned_data$Has_State	0.015	0.004	4.092	0.000

Outcome = -0.001 + 0.031 * Test Variable + 0.048 * Has Email * 0.014 * Has_State + 0.002
Adjusted R-squared = 0.0237249

The adjusted R-squared has increased to 0.023 after using the dummy variables - Has Email and Has State. Hence compared to the output from the first model, this fits better.

Model-3: Outcome = alpha + beta1 * Test Variable * Has Email * beta2 * Has State + error*

```
lmodel3 <- lm(cleaned_abandoned_data$Outcome ~ cleaned_abandoned_data$Test_Variable*abandoned_data$Has_
kable(summary(lmodel3)$coef)
```

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	0.0021784	0.0029693	0.7336306	0.4631943
cleaned_abandoned_data$Test_Variable	0.0231258	0.0036729	6.2963398	0.0000000
abandoned_data$Has_Email	0.0126783	0.0078555	1.6139451	0.1065767
abandoned_data$Has_State	0.0145458	0.0035473	4.1004965	0.0000416
cleaned_abandoned_data$Test_Variable$: $abandoned_data$Has_Email	0.0665159	0.0105552	6.3017373	0.0000000

v. Statistical Analysis: Response Times

Model-4: Outcome = alpha + beta * Days in Between * Test Variable + error

```
lmodel4 <- lm(cleaned_abandoned_data$Outcome ~ cleaned_abandoned_data$Days_in_Between*cleaned_abandoned_
kable(summary(lmodel4)$coef)
```

	Estimate	Std. Error	t value
(Intercept)	1.2790587	0.0032580	392.593735
cleaned_abandoned_data$Days_in_Between	-0.0063948	0.0000164	-390.696164
cleaned_abandoned_data$Test_Variable	0.0316088	0.0036339	8.698344
cleaned_abandoned_data$Days_in_Between$: $cleaned_abandoned_data$Test_Variable	-0.0001563	0.0000183	-8.537275

** This is the best linear regression model for this dataset**