

Proyecto Final: Avance 1 Solución

Carlos Tabares

2022

El objetivo de este avance es realizar una asignación aleatoria estratificada de los clientes inactivos. Esto con la finalidad de determinar los grupos de clientes que recibirán un tipo de comunicación específica en la estrategia experimental.

Instalación de librerías

```
required_pkgs <- c('tidyverse','RCT','fastDummies','kableExtra')

installed_pkgs <- installed.packages()

missing_pkgs <- required_pkgs[!(required_pkgs %in% installed_pkgs[, 1])]

if (length(missing_pkgs) == 0 ) {
  message("Librerías cargadas")
} else {
  install.packages(missing_pkgs)

  message("Instalacion completa")
}

rm(installed_pkgs,missing_pkgs)

invisible(lapply(required_pkgs, library, character.only = TRUE))

rm(required_pkgs)
```

Pregunta 1

Explora la base de datos de clientes nuevos. ¿A qué nivel de desagregación está la base? ¿Cuántos clientes únicos? ¿Qué variables tienen valores vacíos? Decide si debes excluir a esas observaciones o mantenerlas y justifica tu decisión.

Respuesta: Si bien se tienen valores vacíos en las variables tipo de dispositivo, canal de marketing, productos de interés, y tipo de producto, la falta de información no impide el proceso de asignación

```
#Cargamos la base
load("Bases input/base_inactivos.RData")

#Exploramos la base
glimpse(inactivos_db)
```

```
## Rows: 78,593
## Columns: 27
## $ numero_cliente      <dbl> 18802, 49828, 78069, 51725, 90325, 10526, 52735~
## $ email                <chr> "yie@gmail.com", "nauskqpgj@gmail.com", "ocezsb~
## $ organico             <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 1, 0, 0, 1,~
## $ compra_previa        <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 1, 0, 0, 0, 1,~
## $ monto_compra         <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 7, 0, 40, 0, 0, 0, 7~
## $ registro_newsletter  <dbl> 1, 1, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,~
## $ semanas_desde_contacto <dbl> 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3,~
## $ abrio_mail           <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0,~
## $ descargo_app         <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,~
## $ visitas_web          <dbl> 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,~
## $ articulos_carrito     <dbl> 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0, 1, 0,~
## $ correos_enviados      <dbl> 0, 0, 0, 0, 1, 0, 0, 0, 0, 1, 0, 1, 0, 0, 0, 2,~
## $ valor_carrito         <dbl> 0, 0, 96, 0, 0, 0, 0, 0, 0, 0, 0, 10, 0, 0, 7, ~
## $ login_app            <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,~
## $ promocion_previa     <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,~
## $ info_contacto        <dbl> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,~
## $ opt_in_promos        <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,~
## $ genero               <chr> "Mujer", "Mujer", "Mujer", "Mujer", "Hombre", "~
## $ dispositivo          <chr> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA,~
## $ referido             <dbl> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 0, 1, 1, 0,~
## $ edad                 <dbl> 69, 37, 29, 20, 64, 48, 75, 50, 30, 23, 59, 55,~
## $ canal_marketing      <chr> "Facebook", NA, "Google", "Facebook", "Google",~
## $ productos_interes    <chr> "Electronica", "Electronica", NA, "Electronica"~
## $ tipo_producto        <chr> NA, NA, NA, NA, "Producto Estrella", NA, NA, NA~
## $ minutos_pagina       <dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,~
## $ localidad            <chr> "ciudad", "suburbio", "suburbio", "suburbio", "~
## $ costo_adquisicion    <dbl> 5, 4, 2, 6, 12, 4, 5, 9, 2, 2, 10, 0, 0, 5, 9, ~
```

```
# La base está a nivel cliente (numero_cliente). Con 78,593 observaciones únicas.
length(unique(inactivos_db$numero_cliente))
```

```
## [1] 78593
```

```
# Valores faltantes
missings<-map_dbl(inactivos_db %>% select_all(),
                  ~100*sum(is.na(.))/nrow(inactivos_db))

missings[missings>0]
```

```
##      dispositivo  canal_marketing productos_interes    tipo_producto
##      98.54694      52.96019      16.78394      81.97677
```

Pregunta 2:

¿Qué variables crees que puedan estar más correlacionadas con el impacto del tratamiento?

Respuesta: Las variables organico, registro_newsletter, abrio_mail, descargo_app y visitas web podrían indicar el interés del usuario por adquirir un producto y por lo tanto vale considerarlas en la estratificación para eliminar cualquier sesgo posible.

Existen otras variables que resultarían interesantes de analizar en la evaluación de impacto tales como el costo de adquisición y la indicadora de productos en el carrito de compra. La estratificación por costo de

adquisicion nos permitirá conocer si aquellos usuarios con un mayor costo de adquisicion son más rentables. Por su parte, la variable de si el cliente agregó un artículo al carrito nos dirá si la promoción resultó el empujón necesario para completar la compra.

Pregunta 3:

Realiza una asignación aleatoria de la población de clientes en 3 grupos de tamaño similar. El grupo asignado determinará el tipo de promoción que recibirán.

```
#Variables estratificadoras
inactivos_db <- inactivos_db %>%
  mutate(grupo_cac = ntile(costo_adquisicion, 4),
         agrego_articulo = if_else(articulos_carrito==0,0,1))

# Asignación
asignacion <- treatment_assign(inactivos_db,
                              share_control = .33,
                              n_t=2,
                              strata_varlist = vars(organico, registro_newsletter,
                                                    abrio_mail, descargo_app,
                                                    agrego_articulo, grupo_cac),
                              seed = 2000,
                              key = 'numero_cliente')

list2env(asignacion, envir = .GlobalEnv)
```

```
## <environment: R_GlobalEnv>
```

```
# Juntamos la base de clientes inactivos con el universo asignado.
inactivos_db <- left_join(inactivos_db,
                          data,
                          by = "numero_cliente")
```

Pregunta 4.

Realiza las pruebas de balance sobre todas las variables. ¿Están balanceadas las variables entre los 3 grupos?

```
# Transformar las variables de texto en categóricas para poder incluirlas
# en la prueba de balance

inactivos_db_aux <- inactivos_db %>%
  dummy_cols(select_columns = c("genero","dispositivo","canal_marketing",
                              "productos_interes","tipo_producto","localidad"),
            ignore_na = T, remove_selected_columns = T) %>%
  mutate_at(vars(c(starts_with(c("genero","dispositivo","canal_marketing",
                              "productos_interes","tipo_producto","localidad")))),
            function(x) x = if_else(is.na(x),0,as.double(x)))

# Tabla de balance
balance_tab <- balance_table(data = inactivos_db_aux %>%
                             select(-c(numero_cliente, email)),
```

```

      treatment = 'treat')

kable(balance_tab,
      caption = "Tabla de balance",
      digits = 2)

```

Guardamos la base en un archivo csv para poder compartirlo con el equipo de marketing y programar los envios. El archivo especifica el grupo de pertenencia del usuario y el tratamiento a recibir.

```

inactivos_marketing <- inactivos_db%>%
  select(numero_cliente,email,treat)%>%
  mutate(treat = case_when(treat==0 ~ "Sin Comunicacion",
                           treat==1 ~ "Promocion Cashback",
                           treat==2 ~ "Promocion Descuento"))

#fwrite(inactivos_marketing, file = "Bases output/inactivos_marketing.csv")

```

Table 1: Tabla de balance

variables1	Media_control1	Media_trat1	Media_trat2	p_value1	p_value2
abrio_mail	0.15	0.15	0.15	0.84	0.72
agrego_articulo	0.20	0.20	0.21	0.78	0.70
articulos_carrito	0.39	0.39	0.40	0.82	0.36
canal_marketing_Facebook	0.35	0.35	0.35	0.37	0.60
canal_marketing_Google	0.08	0.08	0.08	0.95	0.90
canal_marketing_Instagram	0.04	0.03	0.04	0.07	0.57
canal_marketing_Otro	0.01	0.01	0.00	0.55	0.05
compra_previa	0.12	0.12	0.12	0.05	0.03
correos_enviados	0.24	0.24	0.24	0.28	0.76
costo_adquisicion	3.54	3.55	3.54	0.69	0.97
descargo_app	0.01	0.01	0.02	0.62	0.30
dispositivo_Android	0.00	0.00	0.00	0.05	0.17
dispositivo_IOS	0.01	0.01	0.01	0.57	0.69
edad	50.40	50.44	50.50	0.85	0.55
genero_Hombre	0.36	0.35	0.36	0.73	0.83
genero_Mujer	0.64	0.65	0.64	0.73	0.83
grupo_cac	2.50	2.50	2.50	0.98	0.85
info_contacto	0.56	0.57	0.56	0.14	0.42
localidad_ciudad	0.45	0.45	0.45	0.63	0.74
localidad_comunidad	0.34	0.34	0.34	0.41	0.83
localidad_extranjero	0.01	0.01	0.01	0.19	0.75
localidad_suburbio	0.20	0.20	0.20	0.23	0.55
login_app	0.11	0.10	0.11	0.63	0.97
minutos_pagina	0.44	0.45	0.45	0.46	0.20
missfit	0.06	0.06	0.06	0.99	0.99
monto_compra	25.83	25.94	26.52	0.92	0.58
opt_in_promos	0.33	0.33	0.33	0.17	0.62
organico	0.49	0.49	0.49	0.98	0.73
productos_interes_Deportes	0.03	0.03	0.02	0.22	0.02
productos_interes_Electronica	0.59	0.59	0.59	0.62	0.62
productos_interes_Gaming	0.19	0.19	0.19	0.82	0.99
productos_interes_Hogar	0.03	0.03	0.03	0.35	0.33
promocion_previa	0.00	0.00	0.00	0.67	0.82
referido	0.51	0.51	0.51	0.99	0.81
registro_newsletter	0.83	0.83	0.83	0.97	0.94
semanas_desde_contacto	2.24	2.23	2.23	0.41	0.28
strata	27.45	27.45	27.50	0.98	0.74
tipo_producto_Producto Estrella	0.18	0.17	0.17	0.22	0.14
tipo_producto_Producto Nuevo	0.00	0.00	0.00	0.22	0.55
tipo_producto_Producto Temporada	0.00	0.00	0.00	0.40	0.78
valor_carrito	10.43	10.56	10.81	0.74	0.33
visitas_web	0.55	0.56	0.56	0.54	0.73