

Análisis de campañas de E-mail Marketing

27/01/2022

Contents

INSTRUCCIONES	2
PASO 1	2
PASO 2	2
PASO 3	3
PASO 4	4
PASO 5	5
PASO 6	6

INSTRUCCIONES

Ante todo deben cargar las librerías que necesitarán para trabajar: tidyverse, lubridate

```
library(tidyverse)
library(lubridate)
```

PASO 1

Generar un dataset llamado email_analysis cargando el archivo “dataset_email-mkt.csv” cumpliendo con las siguientes consignas:

- No incluir la primera columna (es un index)
- Modificar la columna de sendout_date para que sea fecha y no character

```
## Importar data
email_analysis <- read.csv("dataset_email-mkt.csv")

email_analysis <- email_analysis %>%
  #Retirar la primera columna
  select(-1) %>%
  #Modificar la columna de sendout_date
  mutate(sendout_date = as_date(sendout_date))
```

PASO 2

Utilizar 3 funciones para explorar los datos Desarrolle un texto de análisis con las principales observaciones sobre el dataset en no menos de 5 líneas

```
## Explorar data
str(email_analysis)
glimpse(email_analysis)
summary(email_analysis)
```

Table 1: Filas y columnas

15000	20
-------	----

Table 2: Data de trabajo

sendout_date	email_scope	description	total_email_sent
2020-11-13	Campaign	NA	1
2020-08-13	Campaign	NA	1
2020-09-25	Campaign	NA	1
2019-12-17	Campaign	NA	1
2020-04-01	Campaign	NA	1

sendout_date	email_scope	description	total_email_sent
2020-10-08	Campaign	NA	1

Interpretación: El archivo está compuesto por cerca de 15000 registros (correos) y 20 variables. Estas variables brindan información acerca de la estrategia de emails: la fecha de envío (de octubre de 2019 a noviembre de 2020), objetivo del email, total de correos enviados, total de correos recibidos, entregados, abiertos, entre otros aspectos relacionados. Asimismo, algunas variables contienen información sobre los identificadores de la campaña, del usuario y sexo del usuario. Adicionalmente, se registran 90 campañas y 500 usuarios a quienes se les enviaron los 15000 correos. Finalmente, se puede indicar que las variables presentan bastantes valores perdidos (NA), incluso hay casos como el de la variable “descripción” donde todos los casos son valores faltantes.

PASO 3

Generar un segundo dataset email_campaign filtrando la columna email_scope == “Campaign”

```
email_campaign <- email_analisis %>%
  filter(email_scope == "Campaign")
```

Table 3: Data filtrada

sendout_date	email_scope	description	total_email_sent
2020-11-13	Campaign	NA	1
2020-08-13	Campaign	NA	1
2020-09-25	Campaign	NA	1
2019-12-17	Campaign	NA	1
2020-04-01	Campaign	NA	1
2020-10-08	Campaign	NA	1

Calculen los datos agregados de todas las columnas que comienzan con “total_” agrupando por journeys_id y teniendo cuidado de no incluir celdas con valores NA

```
agregados <- email_campaign %>%
  #agrupar por journeys_id
  group_by(journeys_ids) %>%
  #calcular datos agregados
  summarise(across(total_email_sent:total_pageviews,
    list(promedio = mean),
    na.rm = TRUE)) %>%
  round(2)
```

Table 4: Datos agregados (continued below)

journeys_ids	total_email_sent_promedio	total_email_delivered_promedio
1	1	1
2	1	0.98

journeys_ids	total_email_sent_promedio	total_email_delivered_promedio
3	1	1
4	1	1
5	1	1
6	1	1

total_email_clicks_promedio	total_email_open_promedio
0.08	1.32
0.03	1.61
0.11	1.59
0.04	1.39
0.13	1.58
0.12	1.61

PASO 4

Realizar los cálculos de open_rate y ctor para cada journeys_id OR: el porcentaje de emails que fueron abiertos por los destinatarios sobre el total de emails entregados. Click to Open Rate (CTOR): El porcentaje de usuarios que recibieron el mail, lo abrieron y realizaron clic en el link deseado.

```
# Calcular Open_Rate
OR <- email_campaign %>%
  mutate(total_email_open = case_when(total_email_open >= 1 ~ 1,
                                       TRUE ~ 0)) %>%

#agrupar por journeys_id
group_by(journeys_ids) %>%
  summarise(Open_Rate = sum(total_email_open)/sum(total_email_sent)*100) %>%
  mutate(Open_Rate = round(Open_Rate,1))

# Calcular Click_to_Open_Rate
COR <- email_campaign %>%
  mutate(total_email_open = case_when(total_email_open >= 1 ~ 1,
                                       TRUE ~ 0)) %>%

  mutate(email_received_and_open = case_when(total_email_delivered == 1 &
                                             total_email_open == 1 ~ 1,
                                             TRUE ~ 0)) %>%

  mutate(total_email_clicks = case_when(total_email_clicks >= 1 ~ 1,
                                         TRUE ~ 0)) %>%

#agrupar por journeys_id
group_by(journeys_ids) %>%
  summarise(Click_to_Open_Rate = sum(total_email_clicks)/sum(email_received_and_open)*100) %>%
  mutate(Click_to_Open_Rate = round(Click_to_Open_Rate,1))

# Unir indicadores en un solo database
indicadores <- merge(OR, COR, by="journeys_ids")
```

Table 6: Indicadores OR & COR

journeys_ids	Open_Rate	Click_to_Open_Rate
1	12.2	3.5
2	27.9	3.1
3	28.1	8.3
4	30	3
5	31.8	10.3
6	24.5	9.2

Interpretación: El análisis nos presenta una vista con los valores más altos de COR (Click_to_Open_Rate), que también corresponden a los valores más elevados de OR (Open_Rate). Al observar la tabla se puede concluir que las campañas 86 y 89 obtuvieron ratios de 100% en ambos indicadores, es decir, todos los usuarios que recibieron los correos de esa campaña los abrieron y realizaron clic en el link deseado. Además, la campaña 76 fue la tercera con mayor rendimiento, siendo que 7 de cada ## 10 usuarios abrieron los correos e hicieron clic.

PASO 5

Indicar cuáles han sido las 3 campañas que mejor han performado de acuerdo a OR y CTOR. Desarrolle un texto de análisis con su respuesta en no menos de 5 líneas (2,5 pts)

```
indicadores %>%
  arrange(desc(Click_to_Open_Rate)) %>%
  head()
```

Table 7: Campañas más exitosas

journeys_ids	Open_Rate	Click_to_Open_Rate
86	100	100
88	100	100
76	77.8	71.4
50	40	50
55	73.3	45.5
72	37.5	33.3

Interpretación: El análisis nos presenta una vista con los valores más altos de COR (Click_to_Open_Rate), que también corresponden a los valores más elevados de OR (Open_Rate). Al observar la tabla se puede concluir que las campañas 86 y 88 obtuvieron ratios de 100% en ambos indicadores, es decir, todos los usuarios que recibieron los correos de esa campaña los abrieron y realizaron clic en el link deseado. Además, la campaña 76 fue la tercera con mayor rendimiento, siendo que 7 de cada 10 usuarios abrieron los correos e hicieron clic. No obstante, un aspecto importante a destacar es que las campañas 86 y 88 solo enviaron un correo, mientras que la campaña 76 registra 9 correos enviados durante todo un año; haciendo más probable que registren ratios elevados. Por todo lo anterior, se sugiere segmentar el análisis según número de correos enviados; para así realizar una comparativa más justa y evitar resultados sesgados.

PASO 6

Realizar análisis de los usuarios según su género, realizando un nuevo dataset que agregue los datos según género (solo m y f) Calcular métricas de OR y CTOR para cada género e identificar si se perciben diferencias de comportamiento en relación a la tasa de apertura y clics. Desarrolle un texto de análisis con su respuesta en no menos de 5 líneas (2,5 pts)

```
# Calcular Open_Rate
OR_g <- email_campaign %>%
  filter(gender %in% c("m","f")) %>%
  mutate(total_email_open = case_when(total_email_open >= 1 ~ 1,
                                       TRUE ~ 0)) %>%

#agrupar por journeys_id
group_by(journeys_ids,gender) %>%
  summarise(Open_Rate = sum(total_email_open)/sum(total_email_sent)*100) %>%
  mutate(Open_Rate = round(Open_Rate,1)) %>%
  ungroup()

# Calcular Click_to_Open_Rate
COR_g <- email_analylisis %>%
  filter(gender %in% c("m","f")) %>%
  mutate(total_email_open = case_when(total_email_open >= 1 ~ 1,
                                       TRUE ~ 0)) %>%

  mutate(total_email_received_and_open = case_when(total_email_delivered == 1 & total_email_open == 1 ~ 1,
                                                    TRUE ~ 0)) %>%

  mutate(total_email_clicks = case_when(total_email_clicks >= 1 ~ 1,
                                       TRUE ~ 0)) %>%

#agrupar por journeys_id
group_by(journeys_ids,gender) %>%
  summarise(Click_to_Open_Rate = sum(total_email_clicks)/sum(total_email_received_and_open)*100) %>%
  mutate(Click_to_Open_Rate = round(Click_to_Open_Rate,1)) %>%
  ungroup()

indicadores_g <- merge(OR_g, COR_g, by=c("journeys_ids","gender"))

indicadores_g %>%
  group_by(gender) %>%
  summarise(across(Open_Rate:Click_to_Open_Rate,
                  list(promedio = mean),
                  na.rm = TRUE))
```

Table 8: Análisis según género

gender	Open_Rate_promedio	Click_to_Open_Rate_promedio
f	27.51	9.921
m	34.18	13.83

INTERPRETACIÓN: Al analizar los valores promedios de los indicadores OR y COR, según género, observamos que los hombres suelen presentar porcentajes más altos de correos abiertos y de clics en el enlace deseado. Sin embargo, en ambos grupos los valores son aceptables con un ratio de 30% en el caso de OR y de 12% en COR. Asimismo, no se observan diferencias en cuanto al mínimo y máximo valor de los indicadores, puesto que en ambos los ratios desde el 0 al 100%.