

TRABAJO GRUPAL

TEST A/B DE MARKETING

ANÁLISIS EXPLORATORIO

INTEGRANTES:

Angie Katherine Gonzalez Gonzales
Oscar Alejandro Castillo Naveda
Mateo Andrade Ballen



OBJETIVOS

BRINDAR LAS CONCLUSIONES SOBRE EL **ANÁLISIS EXPLORATORIO**
REALIZADO A LA BASE DE DATOS DE **TEST A/B**, QUE PERMITAN A LA
EMPRESA TENER LA INFORMACIÓN NECESARIA PARA REALIZAR SUS
PUBLICACIONES CON "ADS" Y "PSA" DE MANERA EFICIENTE.

**¡GRACIAS
ANALISIS EXPLORATORIO!**





MARKETING DIGITAL

El marketing digital es el conjunto de estrategias y acciones que se realizan en medios digitales para promocionar productos o servicios, utilizando canales como redes sociales, motores de búsqueda, correo electrónico y sitios web. Es esencialmente la adaptación del marketing tradicional a un entorno digital.

Philip Kotler, autor de renombre en marketing, ha abordado este tema en varios de sus libros, incluyendo "Marketing 4.0: Moving from Traditional to Digital".

CONTENIDO ORGÀNICO

Son aquellas publicaciones en redes sociales que se realizan sin pagar por su promoción. Estas publicaciones buscan generar interacción y engagement de manera natural con la audiencia. Hay diferentes tipos de contenidos orgánicos y dentro de ellos tenemos el “psa”.

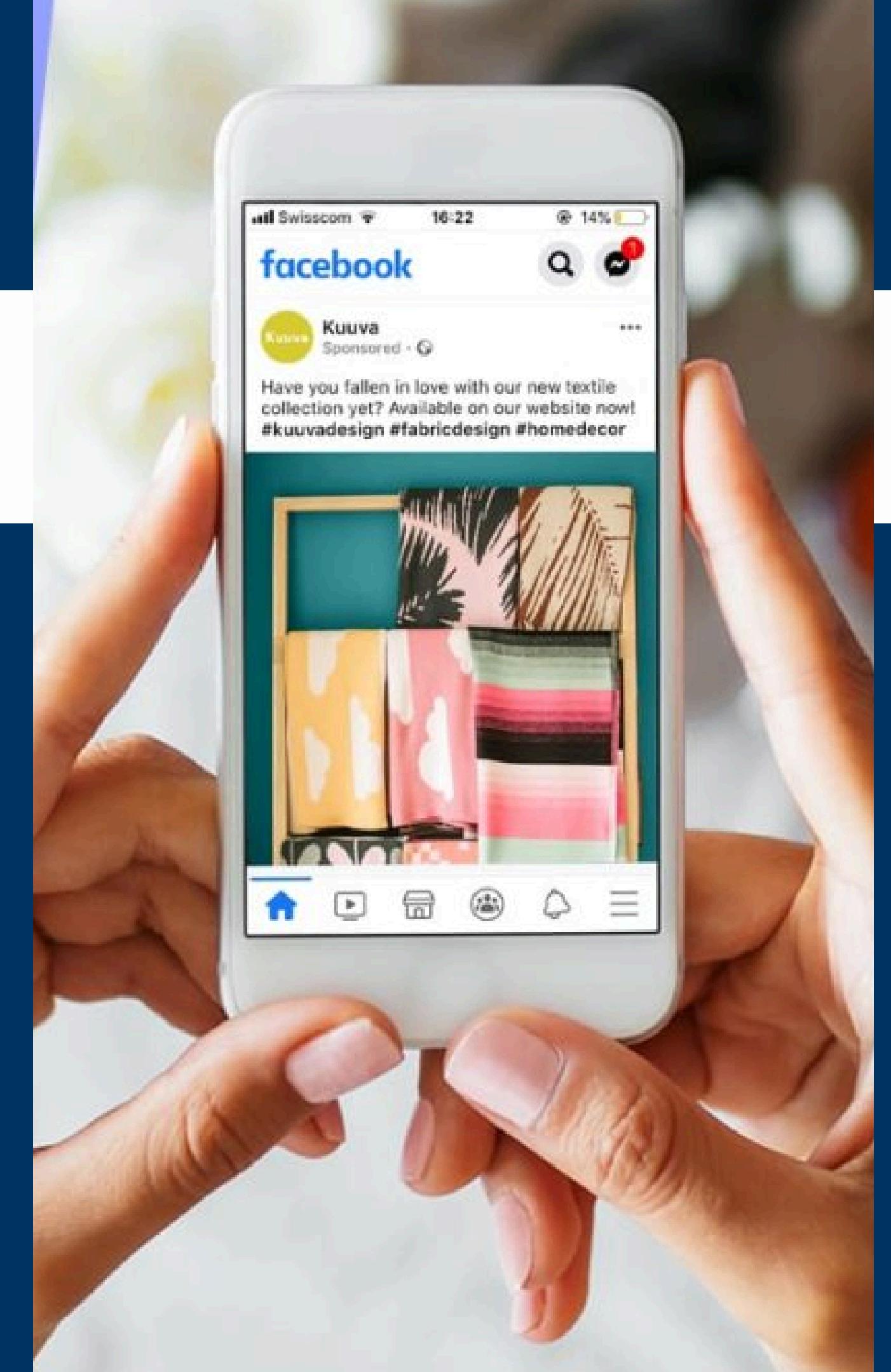
Autores como Gary Vaynerchuk, en su libro "Jab, Jab, Jab, Right Hook: How to Tell Your Story in a Noisy Social World", destacan la importancia de crear contenidos relevantes y auténticos para las redes sociales.



ANUNCIOS PAGADOS “ADS”

Un anuncio pagado en redes sociales es una publicación patrocinada que se muestra a los usuarios según sus intereses y comportamientos en la plataforma. Estos anuncios suelen tener objetivos específicos, como generar ventas, aumentar la visibilidad de la marca o impulsar la participación del usuario.

Ryan Deiss, en su libro "Digital Marketing for Dummies", ofrece una visión general sobre cómo crear y gestionar anuncios efectivos en redes sociales.

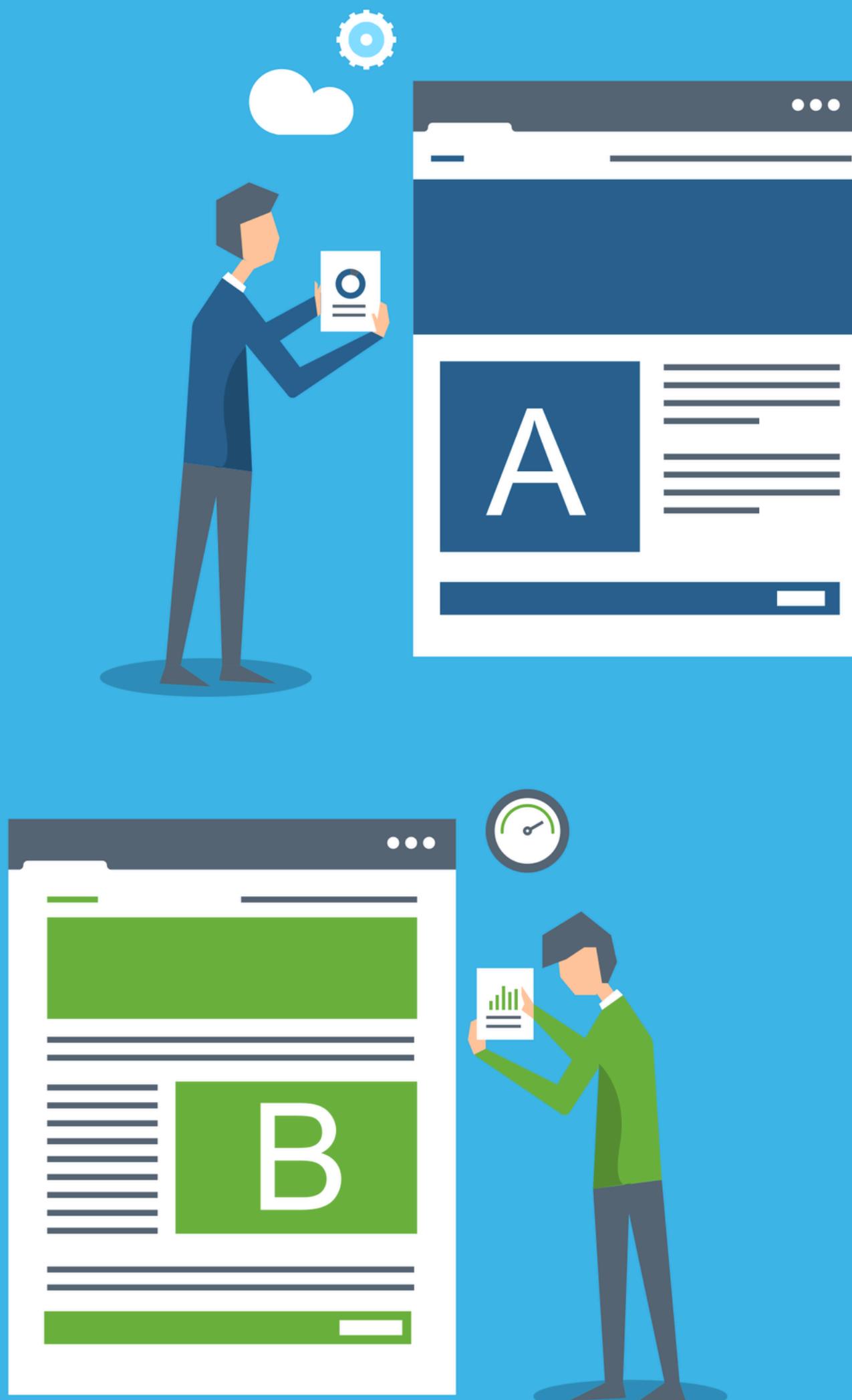




PUBLICIDAD “PSA”

La publicidad PSA (Public Service Announcement) en redes sociales se refiere a anuncios pagados que tienen como objetivo difundir información de interés público, como campañas de concientización, causas sociales o mensajes de utilidad general. No están orientados a la venta directa, sino a promover el bienestar social.

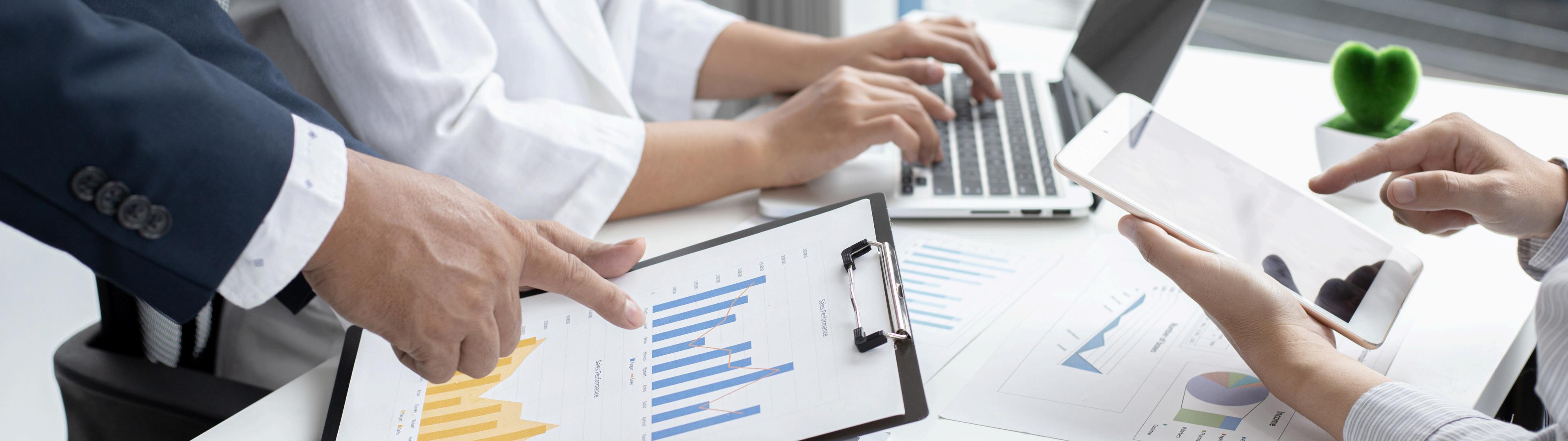
Autores como Dave Evans y Joe Pulizzi en su libro "Epic Content Marketing: How to Tell a Different Story, Break through the Clutter, and Win More Customers by Marketing Less" abordan la importancia de la relevancia y el propósito en la publicidad.



TEST A/B

Un test A/B, también conocido como test de división, es una técnica utilizada para comparar dos versiones diferentes de una publicación en redes sociales para determinar cuál tiene un mejor rendimiento en términos de interacción, clics, conversiones, etc. Esto se logra mostrando ambas versiones a grupos de usuarios diferentes y analizando los resultados.

En su libro "Influence: The Psychology of Persuasion", Robert Cialdini ofrece insights sobre cómo optimizar estrategias de marketing a través de la experimentación y la comprensión del comportamiento humano.



TEMÁTICA

Las empresas de marketing quieren ejecutar campañas exitosas, pero el mercado es complejo y hay varias opciones que pueden funcionar. Por lo tanto, normalmente ajustan las **pruebas A/B**, que es un proceso de experimentación aleatoria en el que dos o más versiones de una variable se muestran a diferentes segmentos de personas al mismo tiempo para determinar qué versión deja el máximo impacto e impulsa las métricas comerciales.

PREGUNTAS A RESPONDER

¿Mayor éxito los anuncios “ad” o “psa”?

¿Cuál es el mejor horario para anuncios?

¿Cuál es la mejor fecha para anuncios?

¿Cuál es la cantidad de anuncios vistos para la compra de un producto?



MAYOR CONTEXTO

¿Mayor éxito los anuncios "ad" o "psa"?

¿Cuál es el mejor horario para anuncios?

¿Cuál es el mejor dia para anuncios?

**¿Cuál es la cantidad de anuncios vistos
para la compra de un producto?**

Conversion por total "ad" y total "psa"

Relación entre "psa" y "ad" vistas con más
"ad" en el día en función al grupo testeado.

Distribución de Conversión.

Las dias con mas "ads" con conversión.



Detail Compact Column

About this file

Basic dataset about a marketing campaign with an A/B testing

# Index	user id User ID (unique)	test group If "ad" the person saw the advertisement, if "psa" they only saw the public service announcement	converted If a person bought the product then True, else is False	# total ads Amount of ads person
0	588k	ad	96%	
1	900k	psa	4%	
2	1.65m			
3	1069124	ad	False	138
4	1119715	ad	False	93
5	1144181	ad	False	21
6	1435133	ad	False	355
7	1015700	ad	False	276
8	1137664	ad	False	734
9	1116205	ad	False	264
10	1496843	ad	False	17
11	1448851	ad	False	21
12	1446284	ad	False	142
13	1257223	ad	False	209
14	1637531	ad	False	47
15	1081965	ad	False	61
16	1837215	ad	False	48
17	1535652	ad	False	28
18	1461774	ad	True	9
19	1492276	ad	False	64
20	1118924	ad	False	26
21	908681	psa	False	248
22	1053783	ad	False	73
23	1381767	ad	False	281
24	1478526	ad	False	389
25	1436823	ad	False	136
26	1274572	ad	False	47
27	1441228	ad	False	264
28	1025697	ad	False	87

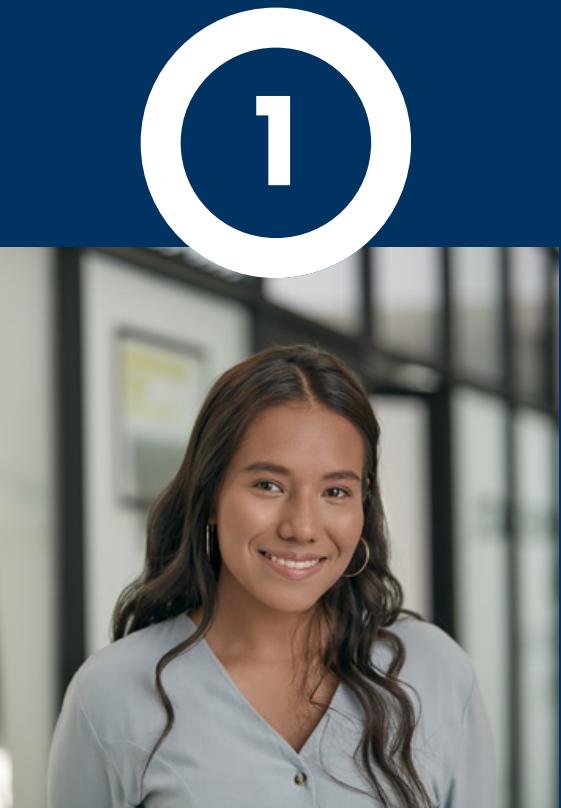
DATASET

Dataset marketing_AB.csv
 588 K datos
 07 columnas

Diccionario de datos:

- Índice:** índice de filas
- ID de usuario:** ID de usuario (único)
- grupo de prueba:** Si "ad" la persona vio el anuncio, si "psa" solo vio el anuncio de servicio público
- convertido:** si una persona compró el producto, entonces es Verdadero, de lo contrario es Falso
- Anuncios totales:** cantidad de anuncios vistos por persona.
- Día de la mayoría de los anuncios:** Día en que la persona vio la mayor cantidad de anuncios.
- Hora de mayor publicidad:** Hora del día en la que la persona vio la mayor cantidad de anuncios.

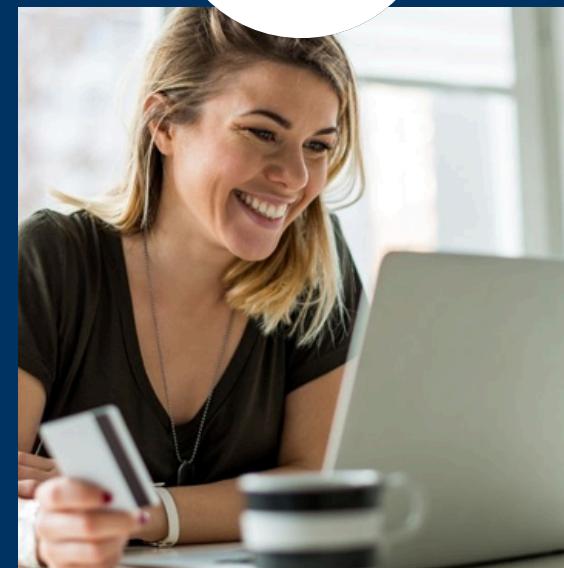
DATASET EN USUARIO



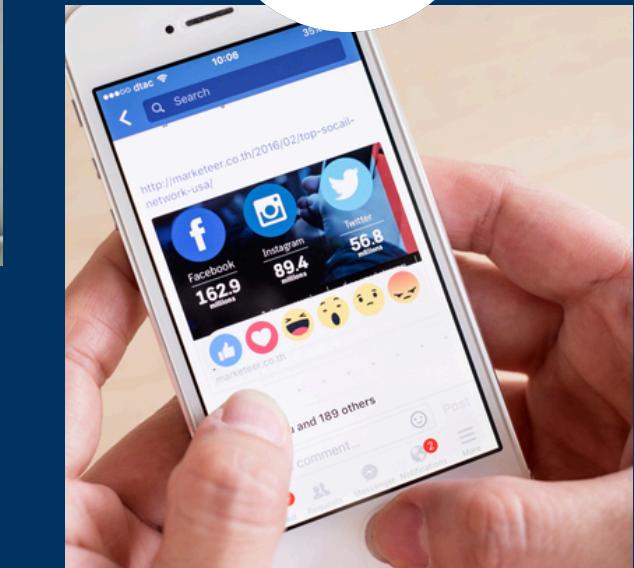
ID DE USUARIO



GRUPO DE PRUEBA



CONVERTIDO



ANUNCIOS TOTALES



DÍA DE LA MAYORÍA
DE LOS ANUNCIOS



HORA DE MAYOR
PUBLICIDAD

PROCESO INICIAL

CARGAR DATOS

```
import pandas as pd
import statistics as st
import numpy as np
from matplotlib import pyplot as plt
import seaborn as sns
import statsmodels.stats.api as sms
from scipy.stats import skew
from scipy.stats import kurtosis
from scipy.stats import ttest_isamp, shapiro, levene, ttest_ind, mannwhitneyu, pearsonr
from statsmodels.stats.proportion import proportions_ztest
import warnings
warnings.filterwarnings("ignore")
```

```
path_data = 'marketing_AB.csv' #ruta de la data
data_proyect = pd.read_csv(path_data) #cargar data
data_proyect.head() #data oficial
```

```
Unnamed: 0 user id test group converted total ads most ads day most ads hour
0 0 1069124 ad False 130 Monday 20
1 1 1119715 ad False 93 Tuesday 22
2 2 1144181 ad False 21 Tuesday 18
3 3 1435133 ad False 355 Tuesday 10
4 4 1015700 ad False 276 Friday 14
```

LIMPIEZA DE DATOS

```
data_proyect.columns #campos de la data
Index(['Unnamed: 0', 'user id', 'test group', 'converted', 'total ads',
       'most ads day', 'most ads hour'],
      dtype='object')
```

```
[ ] df_workProyect = data_proyect.copy() #crear data de trabajo
df_workProyect = df_workProyect.drop('Unnamed: 0', axis=1) #eliminar campo no necesario
df_workProyect = df_workProyect.set_index('user id') #establece ID como indice de la data
df_workProyect.head() #mostrar cabecera
```

	test group	converted	total ads	most ads day	most ads hour
user id					
1069124	ad	False	130	Monday	20
1119715	ad	False	93	Tuesday	22
1144181	ad	False	21	Tuesday	18
1435133	ad	False	355	Tuesday	10
1015700	ad	False	276	Friday	14

```
[ ] df_workProyect.shape #forma data
(588101, 5)
```

```
[ ] df_workProyect.info() #info de la data
#NO HAY VALORES FALTANTES, TODOS LOS CAMPOS SE ENCUENTRAN COMPLETOS
```

```
<class 'pandas.core.frame.DataFrame'>
Index: 588101 entries, 1069124 to 1237779
Data columns (total 5 columns):
 #   Column            Non-Null Count  Dtype  
--- 
 0   test group        588101 non-null  object 
 1   converted         588101 non-null  bool   
 2   total ads         588101 non-null  int64  
 3   most ads day     588101 non-null  object 
 4   most ads hour    588101 non-null  int64 
```

FILTRADO

```
[ ] print(f'Antes de filas repetidas: {df_workProyect.shape}')
df_workProyect.drop_duplicates(inplace=True)
print(f'Despues de filas repetidas: {df_workProyect.shape}')

#Existian filas de igual valor
```

```
[ ] Antes de filas repetidas: (588101, 5)
Despues de filas repetidas: (44478, 5)
```

```
[ ] #FILTRADO
df_workProyect.loc[(df_workProyect.converted == True), 'converted'] = 1
df_workProyect.loc[(df_workProyect.converted == False), 'converted'] = 0
df_workProyect["converted"] = df_workProyect["converted"].astype(int)
```

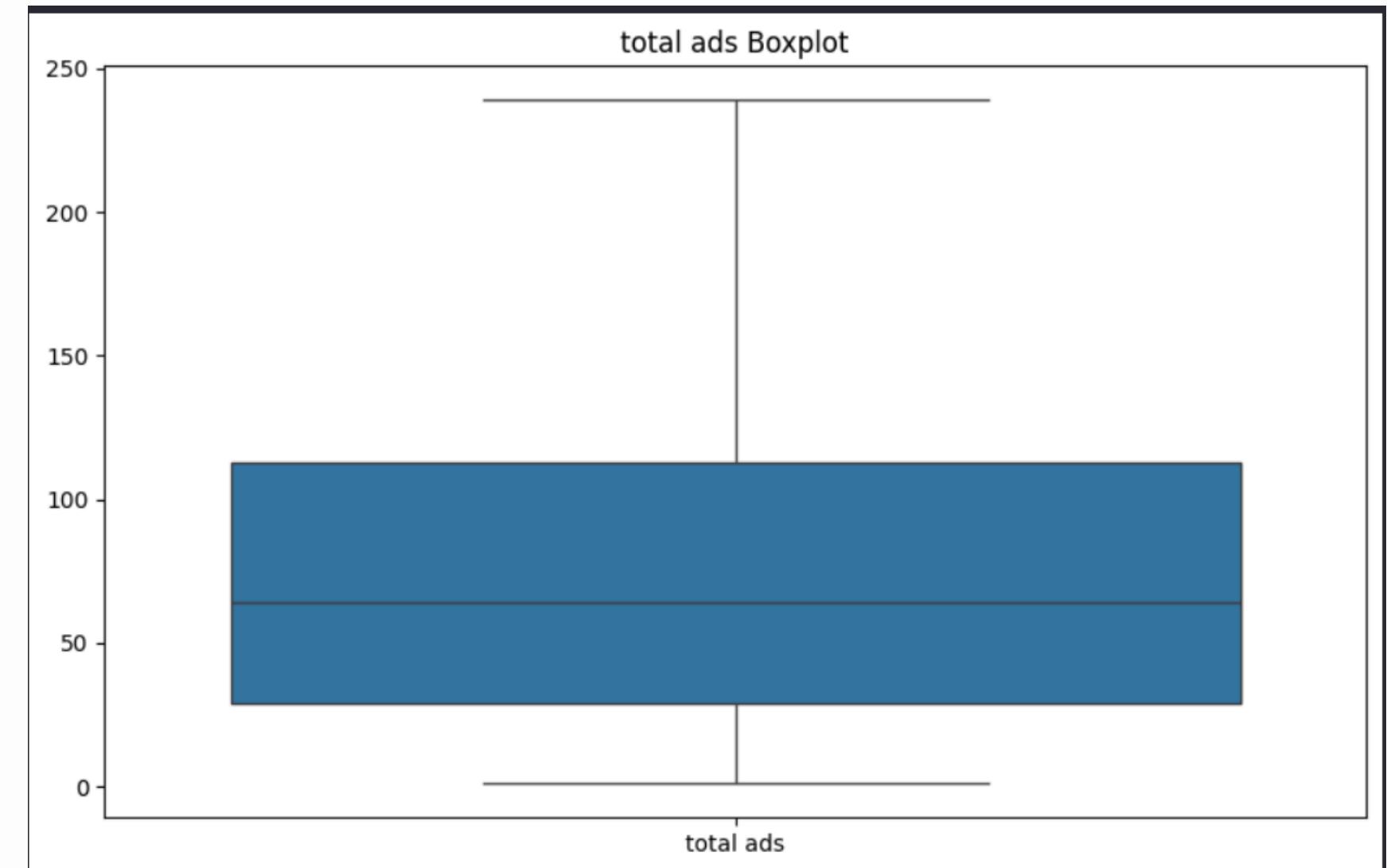
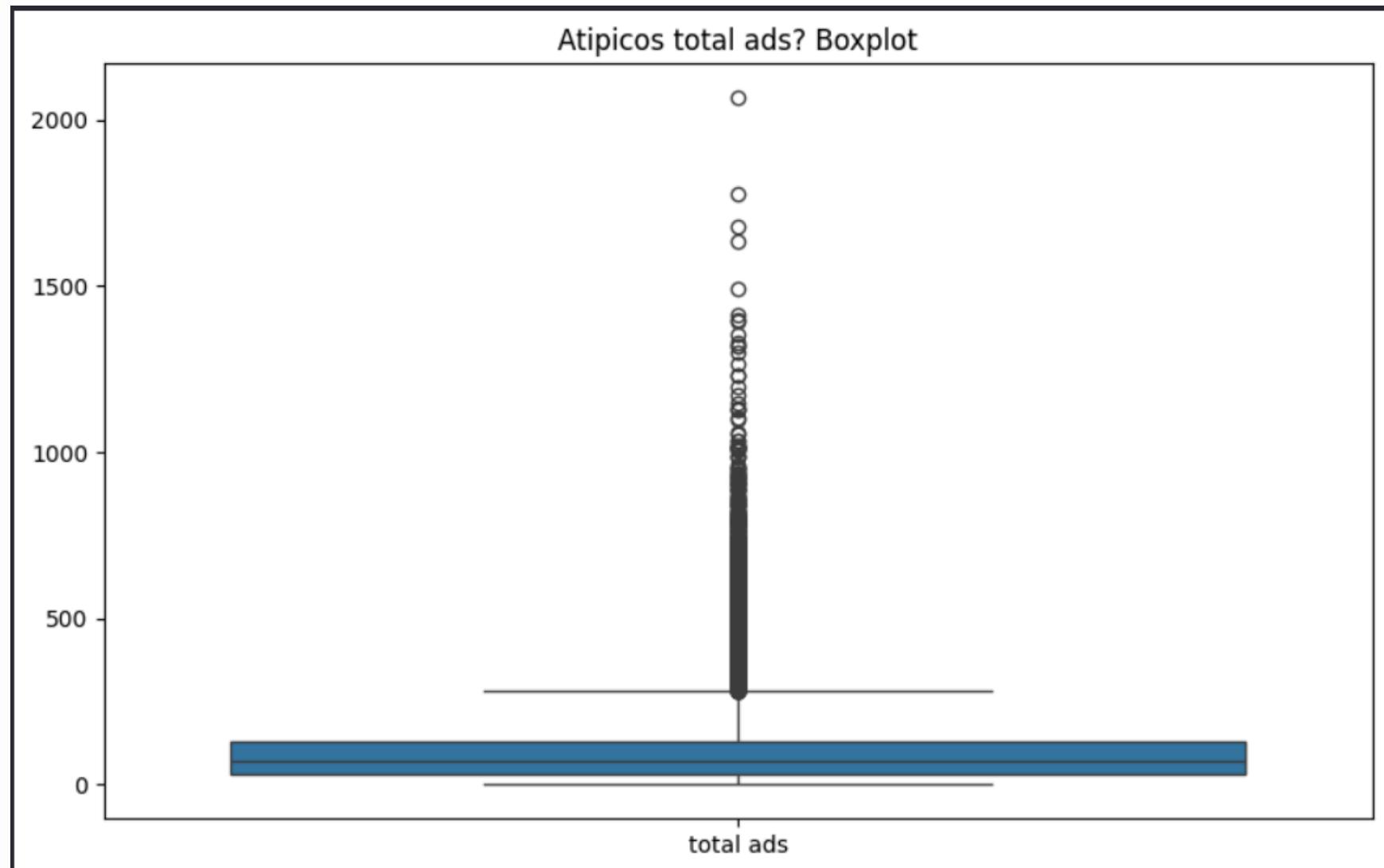
```
► #PRIMERO VEMOS LA GRAFICA PARA LUEGO SELECCIONAR LA SECCION QUE NOS SERVIRA COMO DATOS
def drop_outlier(df_workProyect, col):
    plt.figure(figsize=(10, 6))
    sns.boxplot(data=df_workProyect[[col]])
    plt.title(f'{col} Boxplot')
    plt.show()

    Q1 = df_workProyect[col].quantile(0.25)
    Q3 = df_workProyect[col].quantile(0.75)
    IQR = Q3 - Q1
    down_limit = Q1 - 1.5 * IQR
    up_limit = Q3 + 1.5 * IQR

    df_workProyect = df_workProyect[~((df_workProyect[col] < down_limit) | (df_workProyect[col] > up_limit))]
    return df_workProyect
```

```
[ ] df_workProyect = drop_outlier(df_workProyect, 'total ads')
```

BOXPLOT DEL TOTAL “ADS”



→
Filtro x3

CALCULO ESTADISTICO

MEDIDAS DE TENDENCIA CENTRAL

```
: dataframe=df_workProyect.select_dtypes(include = [float, int])
dataframe.describe().T

:          count      mean       std    min   25%   50%   75%   max
: converted    40702.0  0.241389  0.427931  0.0    0.0    0.0    0.0    1.0
: total ads    40702.0  76.751609  57.977489  1.0   29.0   64.0  113.0  239.0
: most ads hour 40702.0  13.608226  5.908874  0.0   10.0   14.0  18.0  23.0

: # Media
media= st.mean(df_workProyect['total ads'])
print('Media:', media)

# Mediana
mediana= st.median(df_workProyect['total ads'])
print('Mediana:', mediana)

# Moda
moda= st.mode(df_workProyect['total ads'])
print('Moda:', moda)

Media: 76.75160925753035
Mediana: 64.0
Moda: 24

: # Media
media= st.mean(df_workProyect['most ads hour'])
print('Media:', media)

# Mediana
mediana= st.median(df_workProyect['most ads hour'])
print('Mediana:', mediana)

# Moda
moda= st.mode(df_workProyect['most ads hour'])
print('Moda:', moda)

Media: 13.60822564001769
Mediana: 14.0
Moda: 13
```

MEDIDAS DE POSICION

Medidas de Posición

```
# Cuartiles
df_workProyect['total ads'].quantile([0.25,0.5,0.75])

0.25    29.0
0.50    64.0
0.75   113.0
Name: total ads, dtype: float64

# Deciles
df_workProyect['total ads'].quantile([0.1,0.2,0.3,0.4,0.5,0.6,0.7,0.8,0.90])

0.1     12.0
0.2     23.0
0.3     35.0
0.4     49.0
0.5     64.0
0.6     81.0
0.7    101.0
0.8    126.0
0.9    165.0
Name: total ads, dtype: float64

# Percentiles

# Percentil 5
print("Percentil 5:",np.percentile( df_workProyect['total ads'], 5))
# Percentil 50
print("Percentil 50:",np.percentile(df_workProyect['total ads'], 50))
# Percentil 90
print("Percentil 95:",np.percentile(df_workProyect['total ads'], 95))

Percentil 5: 6.0
Percentil 50: 64.0
Percentil 95: 194.0

# Cuartiles
df_workProyect['most ads hour'].quantile([0.25,0.5,0.75])

0.25    10.0
0.50    14.0
0.75    18.0
Name: most ads hour, dtype: float64
```

MEDIDAS DE DISPERSION

Medidas de Dispersion

```
print(f"VARIANZA: {np.var(df_workProyect['total ads'])}")
print(f"DESVIACIÓN ESTÁNDAR: {np.std(df_workProyect['total ads'])}")
print("MAX:",max(df_workProyect['total ads']))
print("MIN:",min(df_workProyect['total ads']))
print("RANGO:",max(df_workProyect['total ads'])-min(df_workProyect['total ads']))

VARIANZA: 3361.3065886098157
DESVIACIÓN ESTÁNDAR: 57.97677628680139
MAX: 239
MIN: 1
RANGO: 238

print(f"VARIANZA: {np.var(df_workProyect['most ads hour'])}")
print(f"DESVIACIÓN ESTÁNDAR: {np.std(df_workProyect['most ads hour'])}")
print("MAX:",max(df_workProyect['most ads hour']))
print("MIN:",min(df_workProyect['most ads hour']))
print("RANGO:",max(df_workProyect['most ads hour'])-min(df_workProyect['most ads hour']))

VARIANZA: 34.91392968541422
DESVIACIÓN ESTÁNDAR: 5.908801036201356
MAX: 23
MIN: 0
RANGO: 23
```

CALCULO ESTADISTICO

MEDIDAS DE FORMA

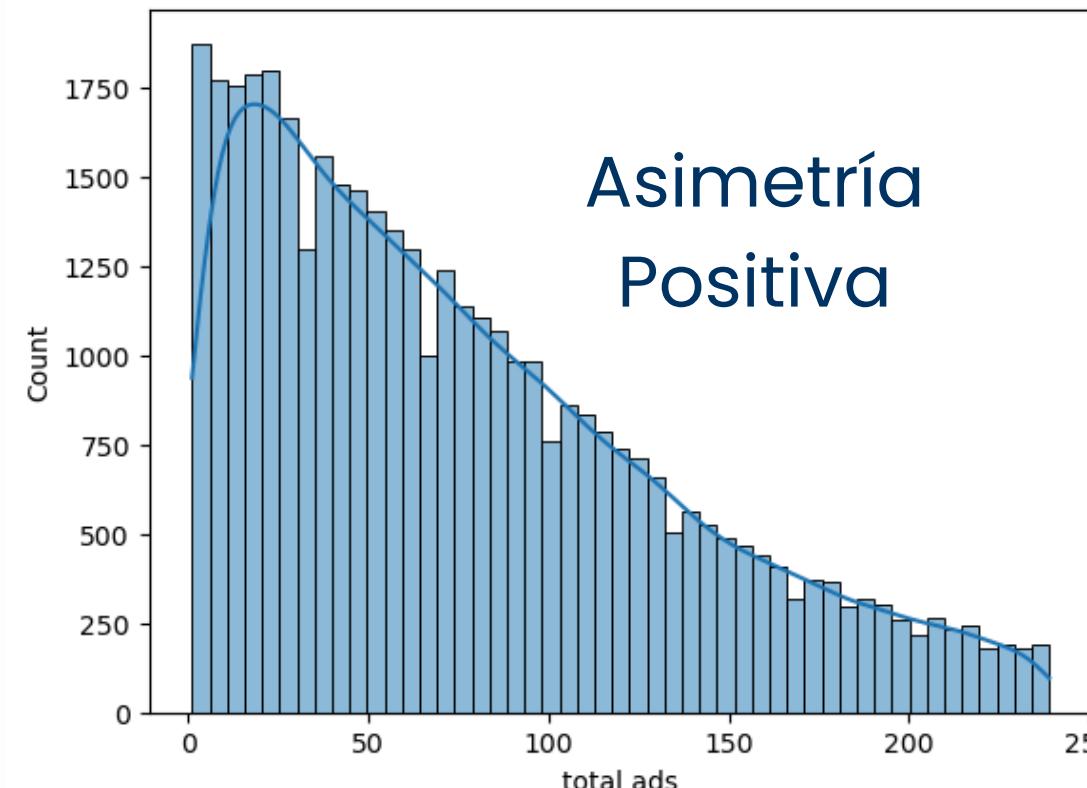
```
print(skew(df_workProyect['total ads'], bias=False))
print(skew(df_workProyect['most ads hour'], bias=False))
```

```
0.7948662722286618
-0.42890914611789416
```

```
print(kurtosis(df_workProyect['total ads'], fisher=False))#segun Pearson
print(kurtosis(df_workProyect['total ads'], fisher=True))#Segun Feasher
```

```
2.8249137464151697
-0.17508625358483032
```

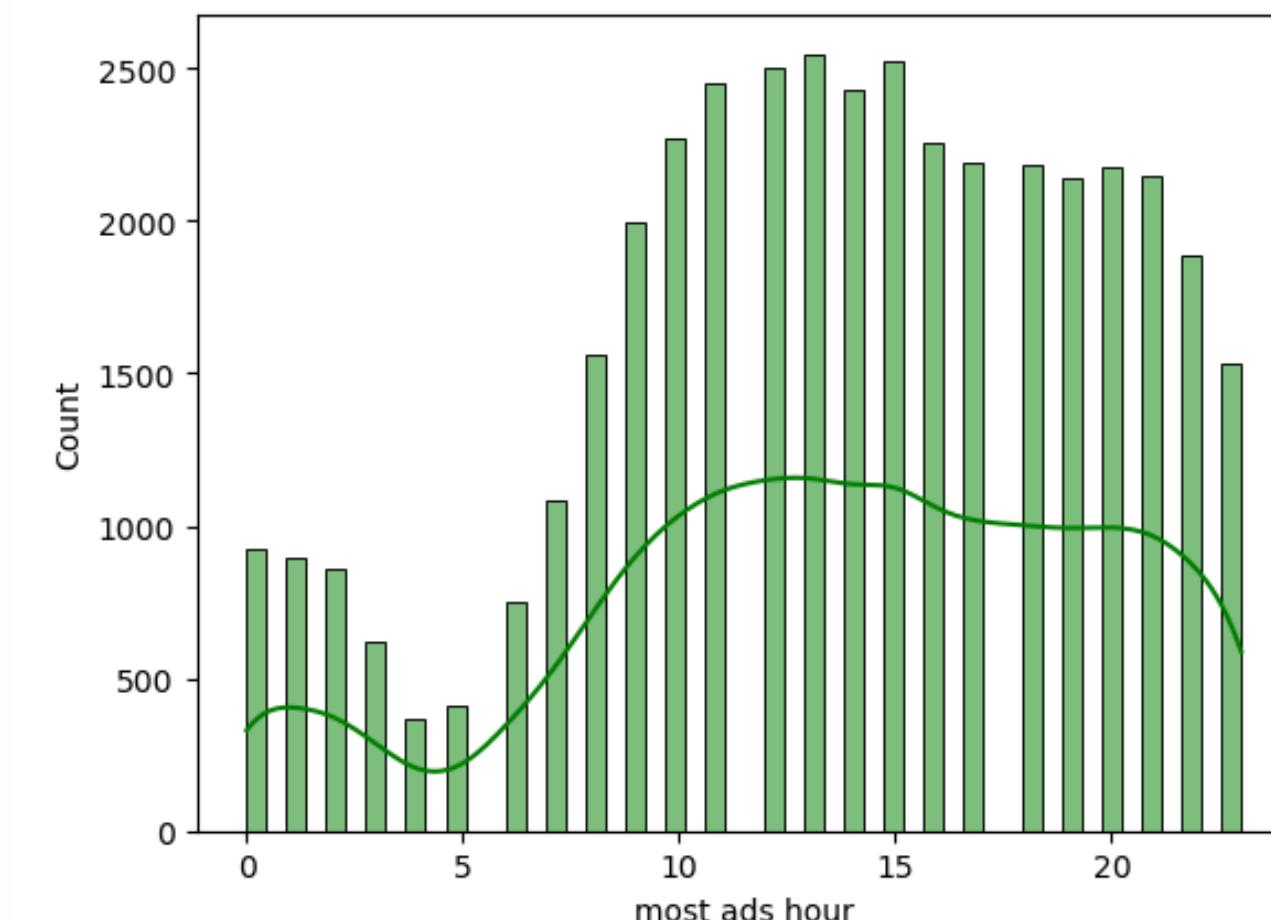
```
sns.histplot(data=df_workProyect,x='total ads', kde=True)
plt.show()
```



```
print(kurtosis(df_workProyect['most ads hour'], fisher=False))#segun Pearson
print(kurtosis(df_workProyect['most ads hour'], fisher=True))#Segun Feasher
```

```
2.5188607643604275
-0.4811392356395725
```

```
sns.histplot(data=df_workProyect,x='most ads hour', kde=True, color='green')
plt.show()
```



Dispersión de los Datos

1. ¿MAYOR ÉXITO LOS ANUNCIOS “AD” O “PSA”?

```
df_workProyect['test group'].value_counts()
```

```
test group  
ad      33130  
psa     7572  
Name: count, dtype: int64
```

```
df_workProyect.groupby('test group')['total ads'].mean().reset_index().set_index('test group')
```

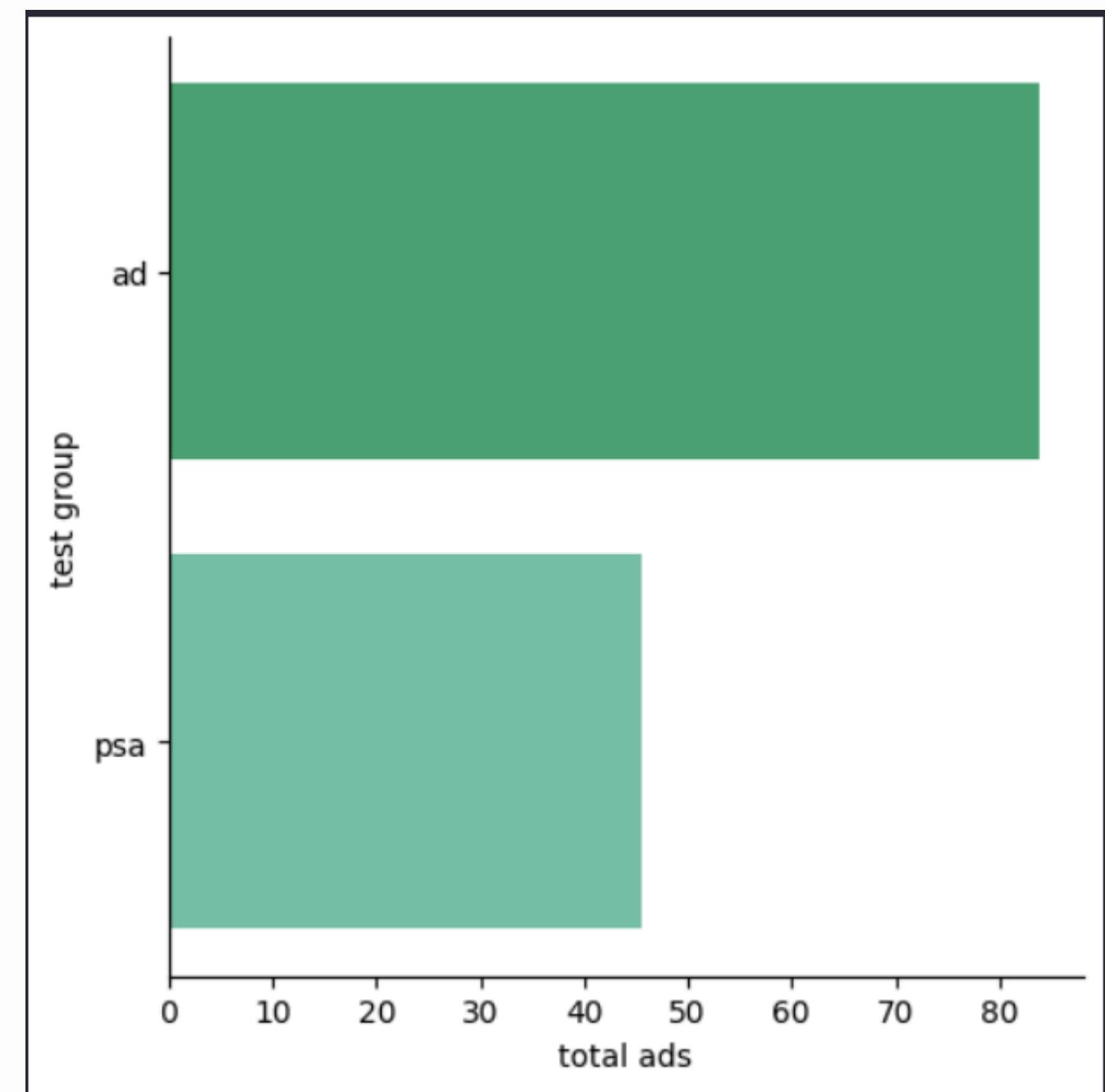
```
total ads
```

```
test group
```

```
ad  83.898219
```

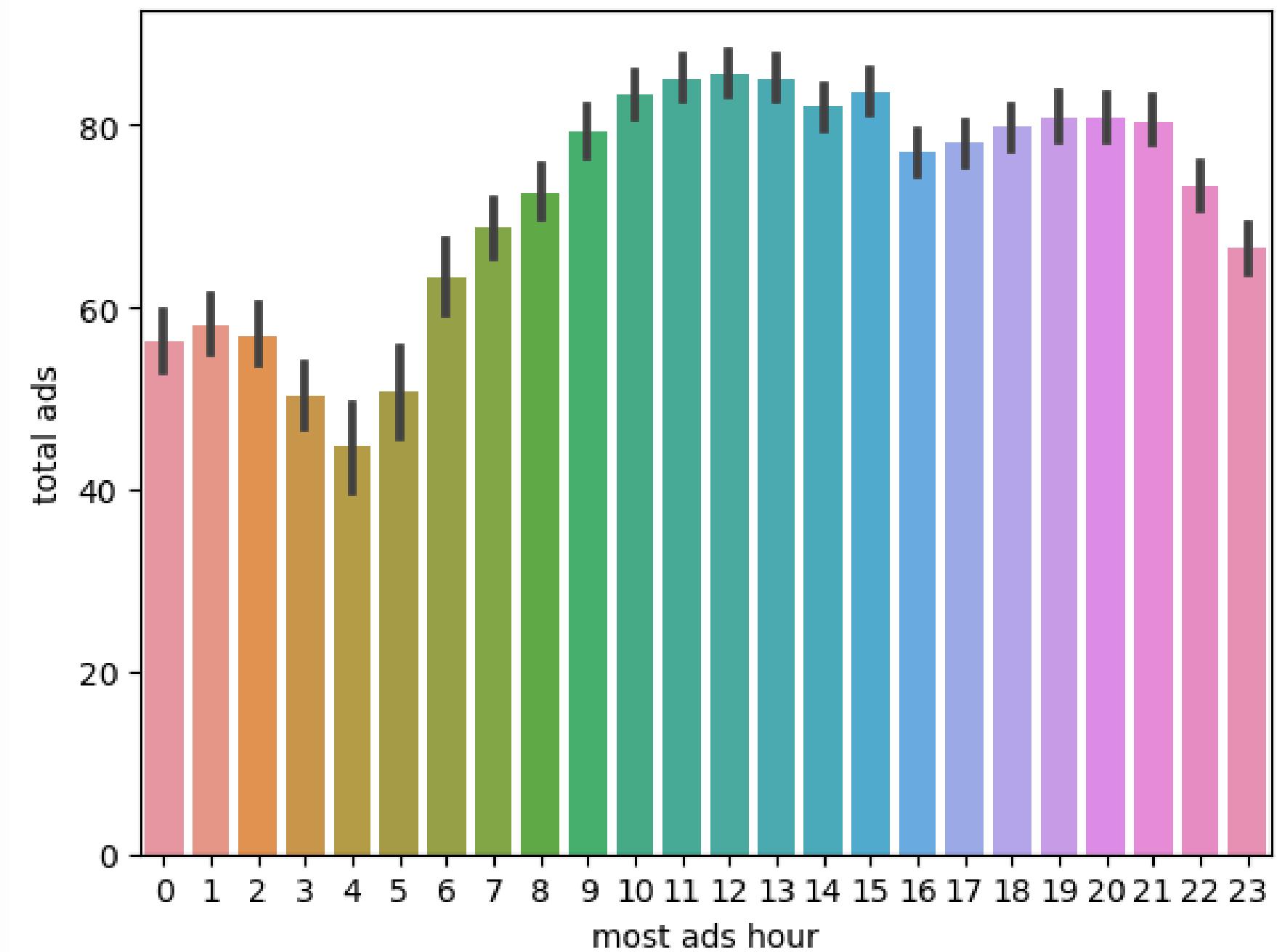
```
psa 45.482831
```

```
sns.catplot(x="total ads",  
            y="test group",  
            data=df_workProyect,  
            kind="bar",  
            ci=None,  
            palette=['mediumseagreen','mediumaquamarine'])  
plt.show()
```



2 ¿CUÁL ES EL MEJOR HORARIO PARA ANUNCIOS?

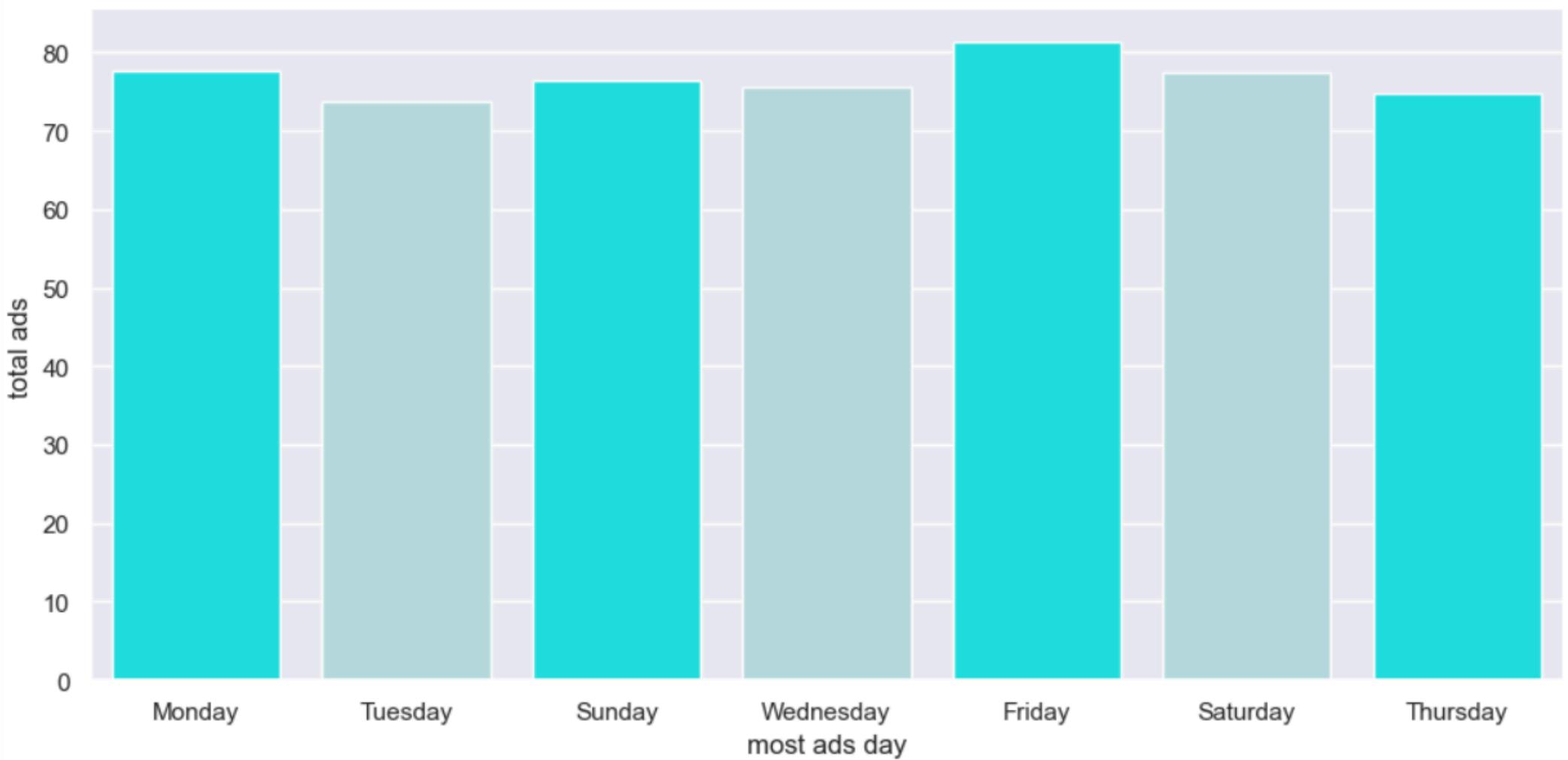
total ads	
most ads hour	
0	55.940605
1	58.005605
2	56.613318
3	50.059581
4	44.670270
5	50.353659
6	62.773635
7	68.491682
8	72.452915
9	78.822645
10	82.954205
11	84.991035
12	85.389932
13	84.887711
14	81.493405
15	83.512485
16	76.708703



3. ¿CUÁL ES EL MEJOR DÍA PARA ANUNCIOS?

```
df_workProyect.groupby('most ads day')['total ads'].mean().reset_index().set_index('most ads day')

total ads
most ads day
Friday 81.344728
Monday 77.608429
Saturday 77.422171
Sunday 76.318848
Thursday 74.737550
Tuesday 73.719035
Wednesday 75.517622
```

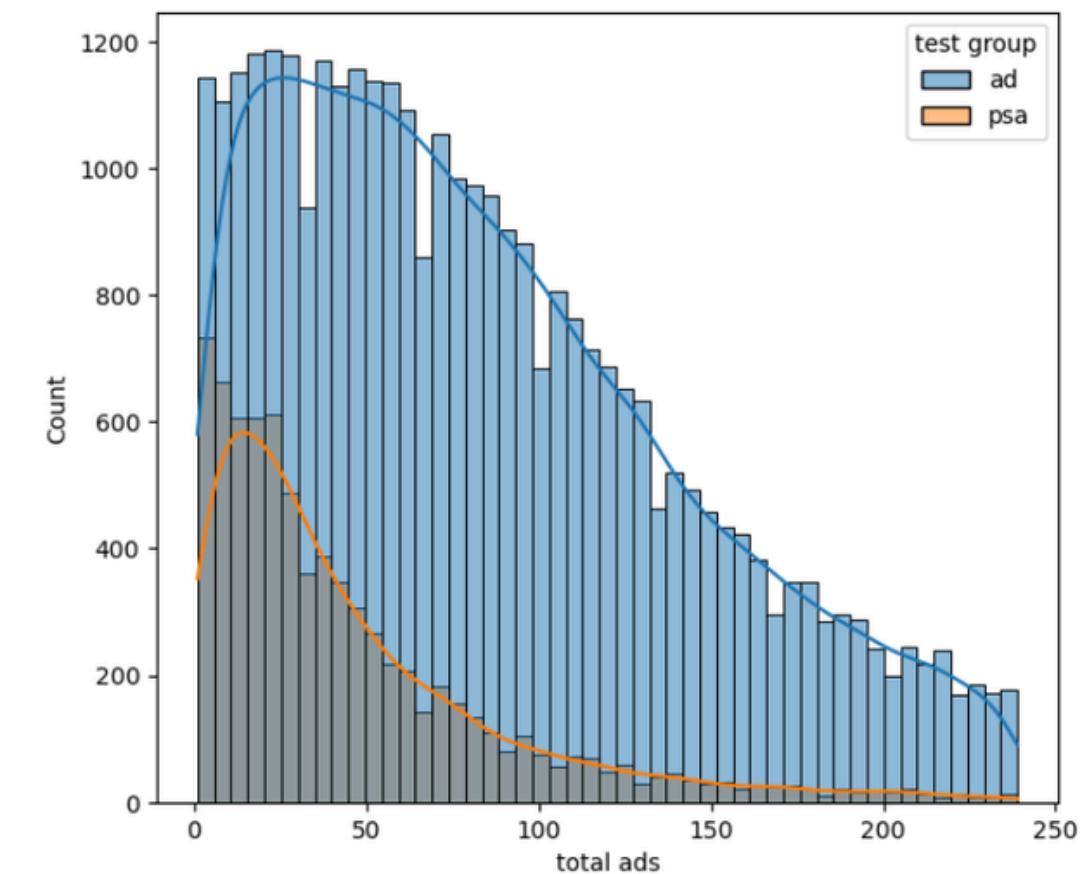
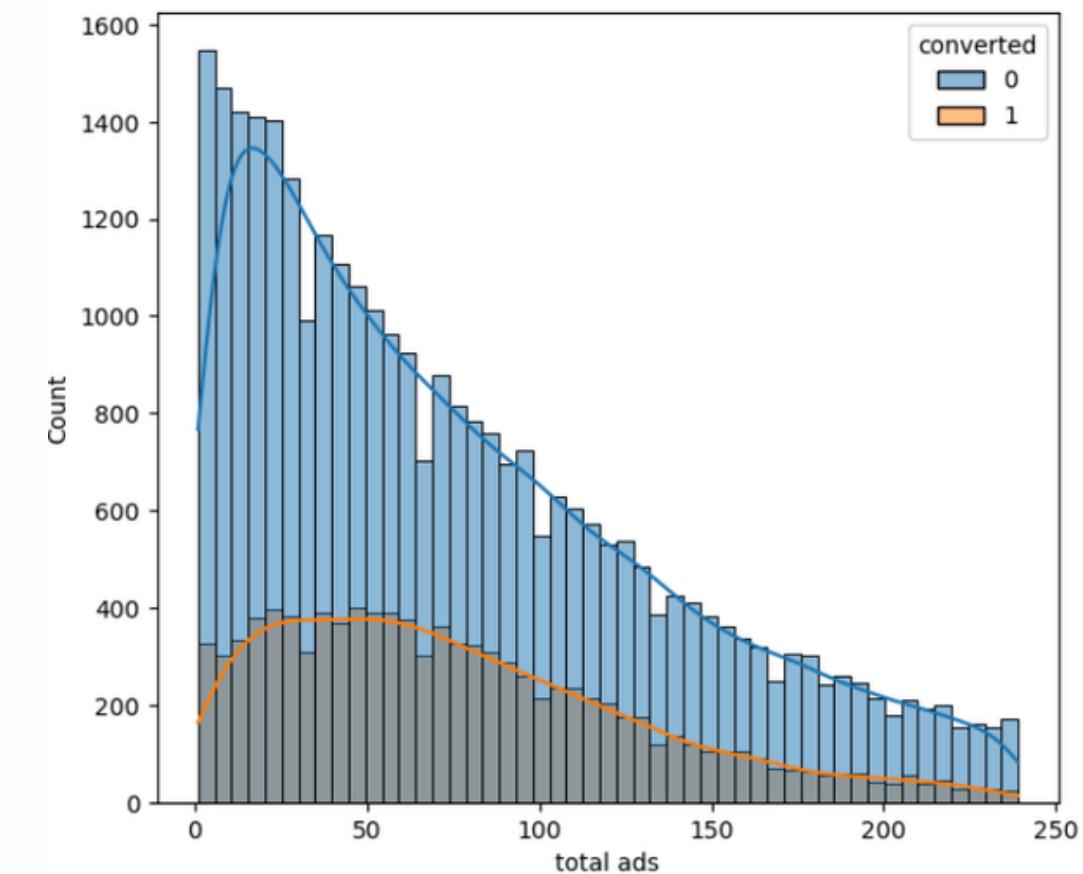


4. ¿CUAL ES LA CANTIDAD DE ANUNCIOS ("AD" Y "PSA") VISTOS DEL GRUPO DE PRUEBA PARA LA COMPRA DE UN PRODUCTO?

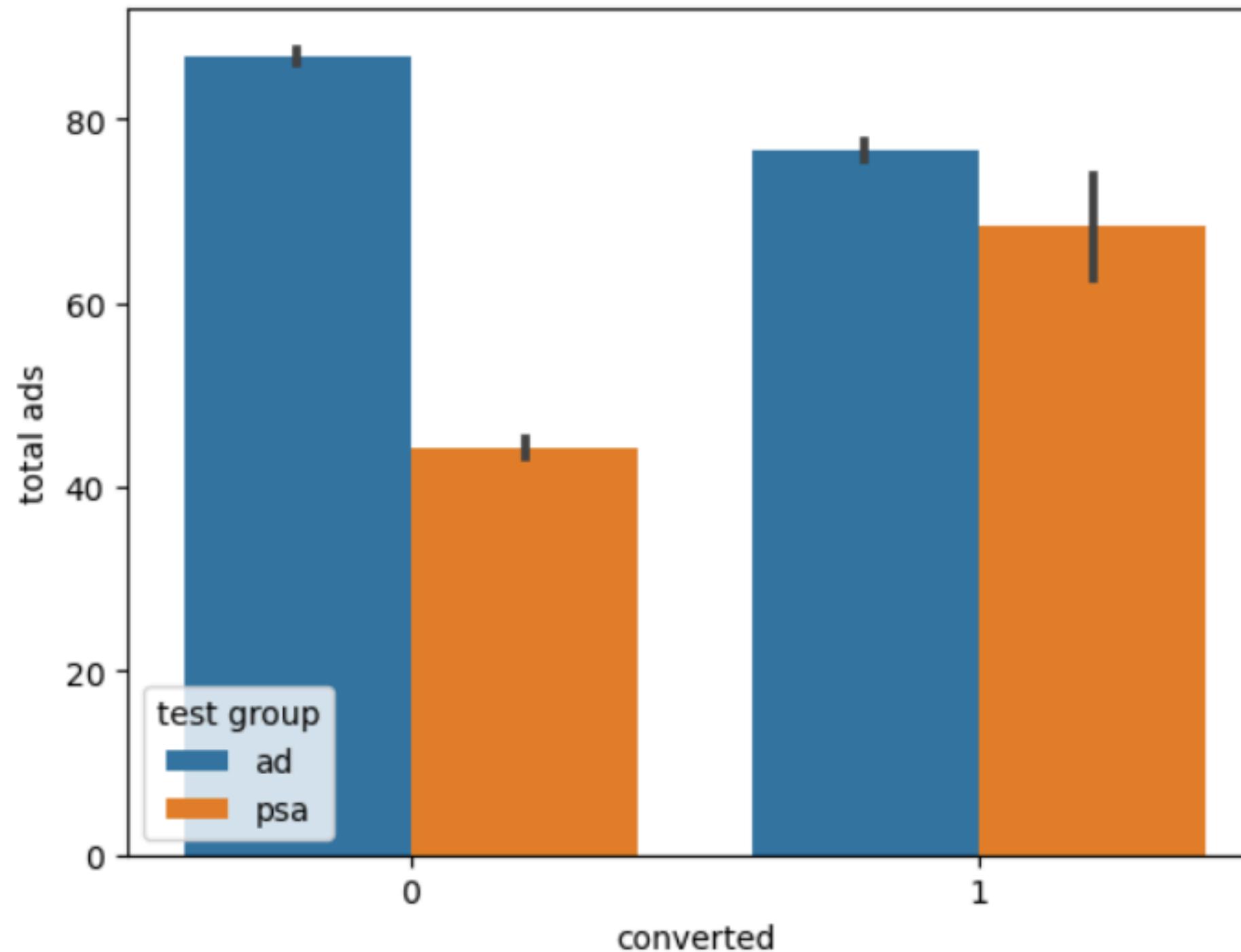
```
] df_workProyect.groupby(['converted','test group'])['total ads'].count().reset_index()  
]  


| converted | test group | total ads |
|-----------|------------|-----------|
| 0         | ad         | 23693     |
| 1         | psa        | 7184      |
| 2         | ad         | 9437      |
| 3         | psa        | 388       |

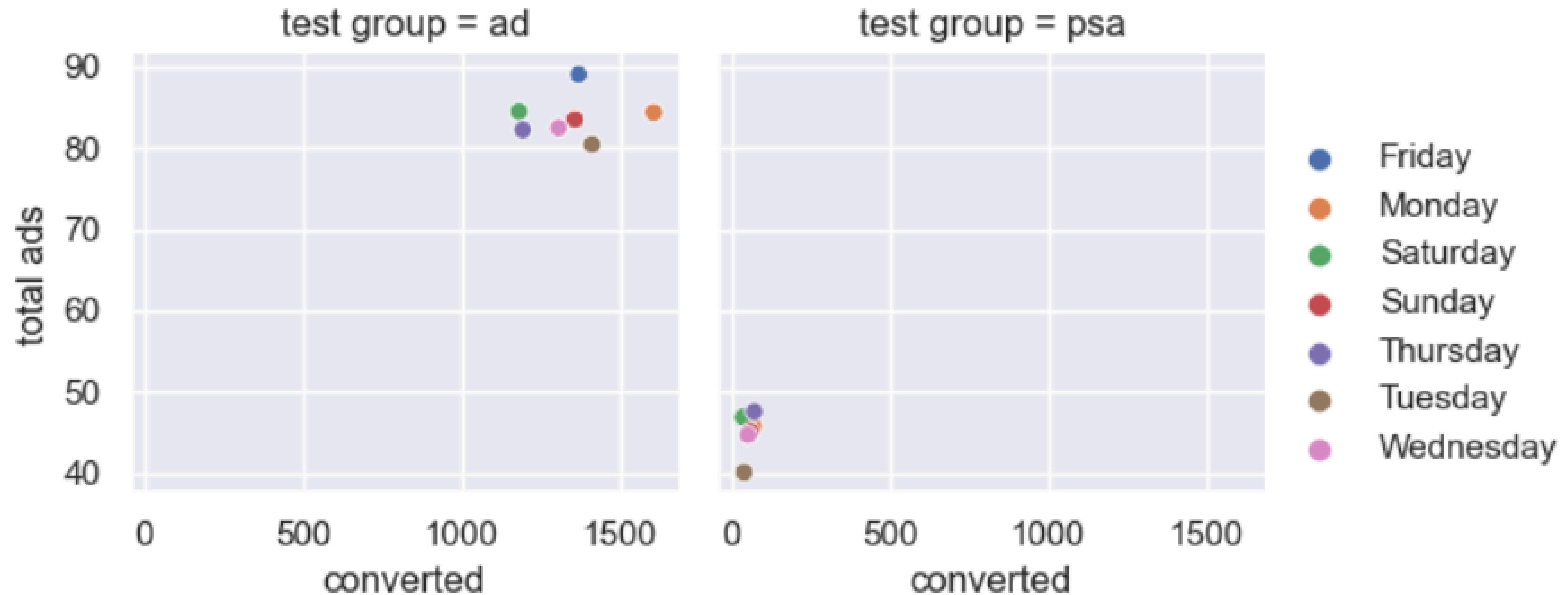

```



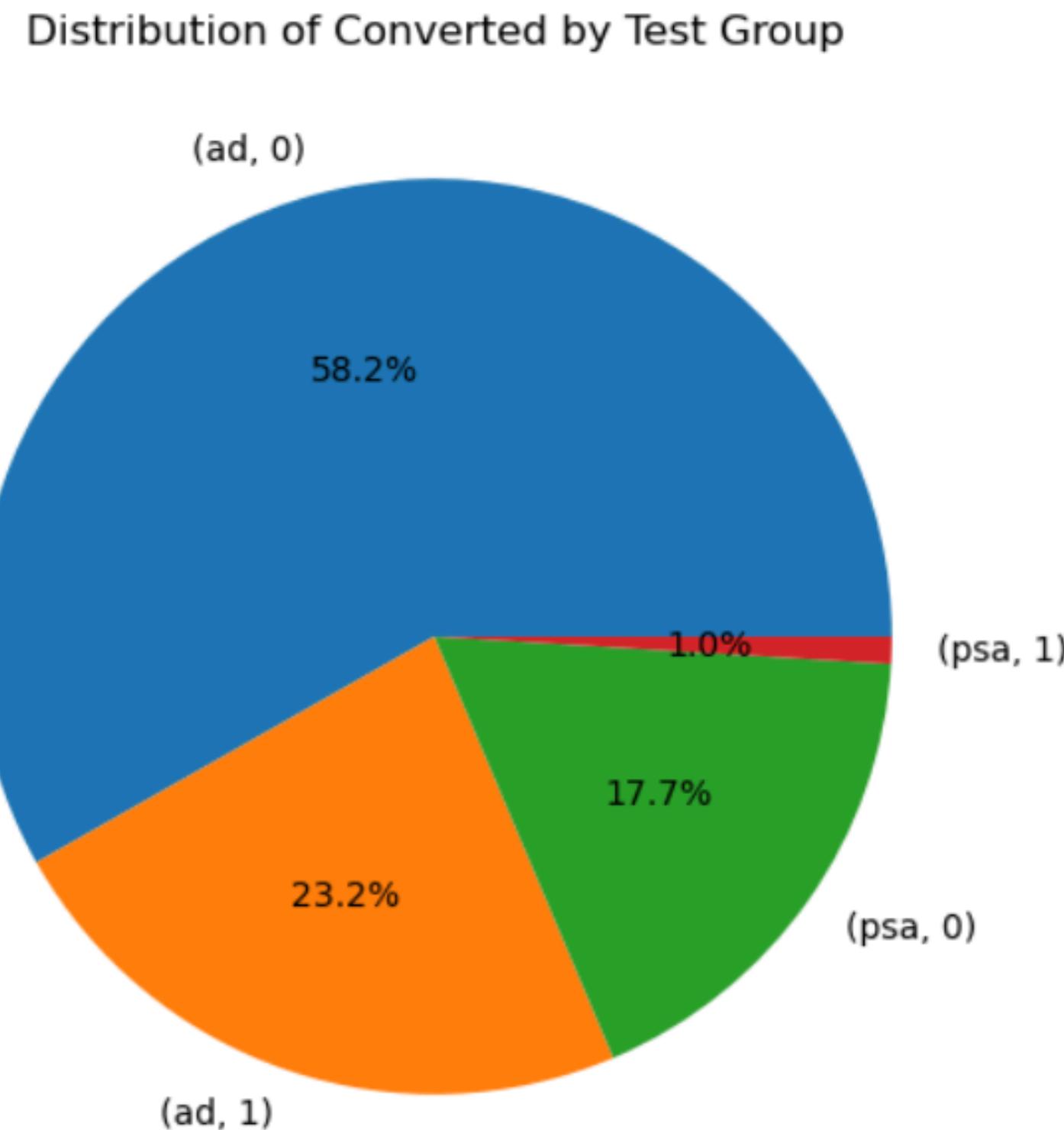
4.1. CONVERSION POR TOTAL “AD” Y TOTAL “PSA”



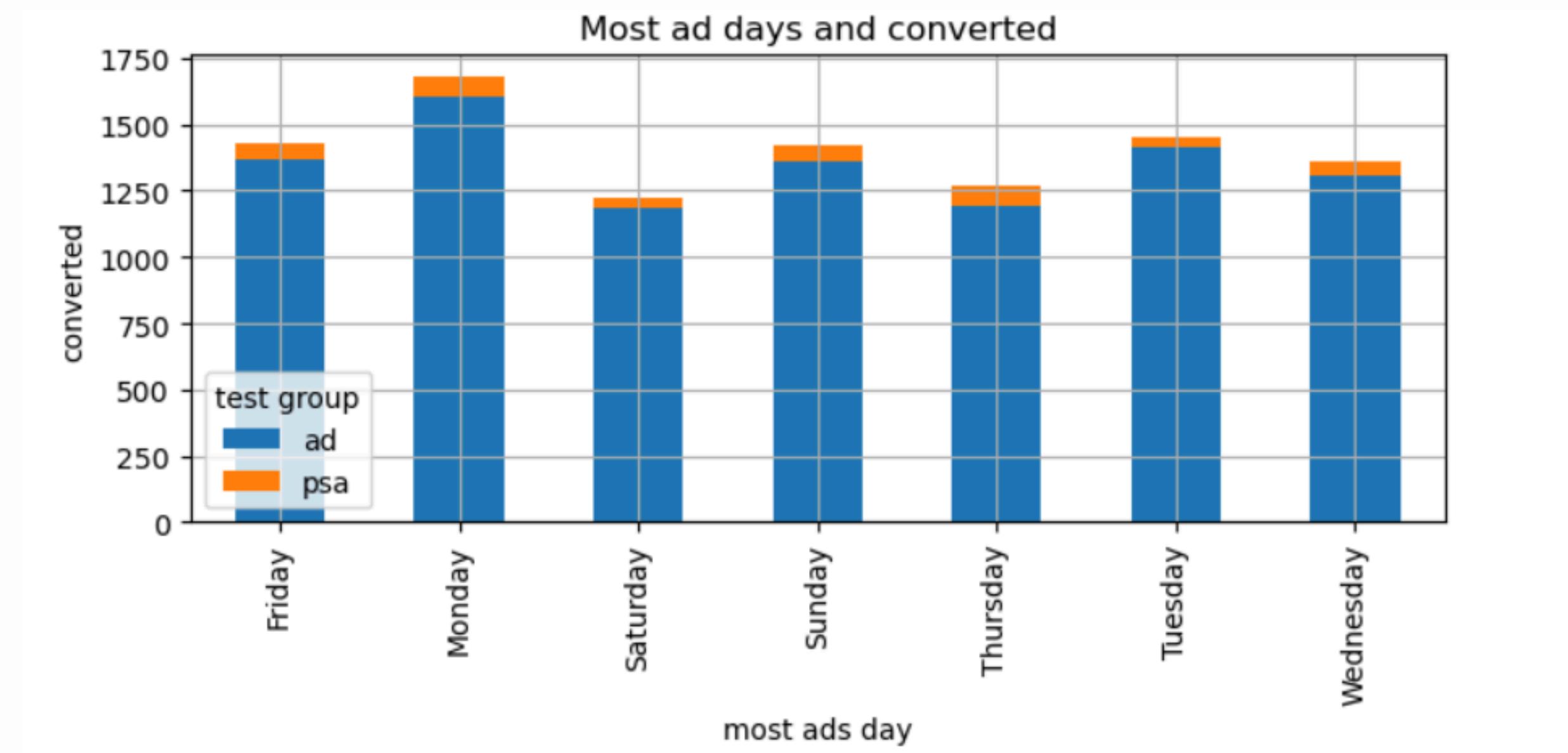
4.2. RELACION ENTRE “PSA” Y “AD” VISTAS CON MAS “AD” EN EL DIA EN FUNCION AL GRUPO TESTEADO



4.3. DISTRIBUCIÓN DE CONVERSION EN FUNCION AL GRUPO TESTEADO



4.4. LOS DIAS CON MAS “AD” EN CONVERSION





CONCLUSIONES

LOGRAMOS BRINDAR LA INFORMACIÓN NECESARIA BASADOS EN LA BASE DE DATOS ADQUIRIDAS, DE MANERA QUE PODEMOS RESALTAR QUE LOS "AD" SON LOS MÁS VISTOS, EL MEJOR HORARIO ES AL MEDIO DÍA, EL MEJOR DÍA PARA PUBLICITAR ES EL VIERNES, NO EXISTE DIFERENCIA DE RESULTADOS DE CONVERSIÓN ENTRE "AD" Y "PSA", ASIMISMO, SI EXISTE UNA GRAN DIFERENCIA DE VISUALIZACIÓN, DE TODA LA INVERSIÓN EN ANUNCIOS EL 40% OBTUVO CONVERSIÓN, EL DÍA CON MÁS "AD" DE CONVERSIÓN FUE EL LUNES Y FINALMENTE, LA CANTIDAD NECESARIA DE VISUALIZACIONES PARA LA COMPRA DE UN PRODUCTO EN EL GRUPO DE PRUEBA QUE HICIERON CONVERSIÓN FUE MAYOR A 9825.

ESTA INFORMACIÓN PUEDE AYUDAR A CUALQUIER EMPRESA A DESARROLLAR UNA MEJOR ESTRATEGIA DE MARKETING DIGITAL BASADO EN ANUNCIOS "AD" Y "PSA".

¡GRACIAS!

