

In [1]: pwd

Out[1]: '/Users/yaha/Desktop'

In [2]: import pandas as pd

In [3]: df = pd.read\_csv('/Users/yaha/Desktop/AdSmartABdata - AdSmartABdata.csv', sep=',')  
df

Out[3]:

	auction_id	experiment	date	hour	device_make	platform_os	browser	yes	no
0	0008ef63-77a7-448b-bd1e-075f42c55e39	exposed	2020-07-10	8	Generic Smartphone	6	Chrome Mobile	0	0
1	000eabc5-17ce-4137-8efe-44734d914446	exposed	2020-07-07	10	Generic Smartphone	6	Chrome Mobile	0	0
2	0016d14a-ae18-4a02-a204-6ba53b52f2ed	exposed	2020-07-05	2	E5823	6	Chrome Mobile WebView	0	1
3	00187412-2932-4542-a8ef-3633901c98d9	control	2020-07-03	15	Samsung SM-A705FN	6	Facebook	0	0
4	001a7785-d3fe-4e11-a344-c8735acacc2c	control	2020-07-03	15	Generic Smartphone	6	Chrome Mobile	0	0
...	...	...	...	...	...	...	...	...	...
8072	ffea24ec-cec1-43fb-b1d1-8f93828c2be2	exposed	2020-07-05	7	Generic Smartphone	6	Chrome Mobile	0	0
8073	ffea3210-2c3e-426f-a77d-0aa72e73b20f	control	2020-07-03	15	Generic Smartphone	6	Chrome Mobile	0	0
8074	ffea0f1-1d72-4ba9-afb4-314b3b00a7c7	control	2020-07-04	9	Generic Smartphone	6	Chrome Mobile	0	0
8075	ffeed62-3f7c-4a6e-8ba7-95d303d40969	exposed	2020-07-05	15	Samsung SM-A515F	6	Samsung Internet	0	0
8076	fffb9ff-568a-41a5-a0c3-6866592f80d8	control	2020-07-10	14	Samsung SM-G960F	6	Facebook	0	0

8077 rows × 9 columns

In [4]: *#Informations sur la dataframe*  
df.info()

```
<class 'pandas.core.frame.DataFrame'>  
RangeIndex: 8077 entries, 0 to 8076  
Data columns (total 9 columns):  
#   Column      Non-Null Count  Dtype  
---  -  
0   auction_id  8077 non-null   object  
1   experiment  8077 non-null   object  
2   date        8077 non-null   object  
3   hour        8077 non-null   int64  
4   device_make 8077 non-null   object  
5   platform_os 8077 non-null   int64  
6   browser     8077 non-null   object  
7   yes         8077 non-null   int64  
8   no          8077 non-null   int64  
dtypes: int64(4), object(5)  
memory usage: 568.0+ KB
```

In [5]: *#Détection des valeurs manquantes*  
df.isnull().sum()

Out[5]: auction\_id 0  
experiment 0  
date 0  
hour 0  
device\_make 0  
platform\_os 0  
browser 0  
yes 0  
no 0  
dtype: int64

In [6]: df.duplicated().sum()

Out[6]: 0

In [7]: df1 = df.sort\_values(by='date', ascending=True)  
df1.head()

Out[7]:

	auction_id	experiment	date	hour	device_make	platform_os	browser	yes	no
4038	818f37f3-dae8-4a33-a22e-e54056e0ba7d	control	2020-07-03	15	Samsung SM-A600FN	6	Chrome Mobile WebView	0	0
3345	6be2edc7-cf48-4831-8edf-3694eb241e09	control	2020-07-03	15	Generic Smartphone	6	Chrome Mobile	0	0
3348	6bfed6f0-66e1-4cf8-9871-ff948b5a2afe	exposed	2020-07-03	11	Generic Smartphone	6	Chrome Mobile	0	0
3354	6c3987e1-42ab-4bca-b1c9-3c6e4446f7eb	control	2020-07-03	15	Generic Smartphone	6	Chrome Mobile	0	0
3356	6c3cd274-3fba-4a35-97c8-85af0385517a	exposed	2020-07-03	8	Generic Smartphone	6	Chrome Mobile	0	0

In [8]: *#Suppression de la colonne "auction\_id" pour mieux l'exploiter sur Tableau*  
df2 = df1.drop(['auction\_id'], axis=1)  
df2.head()

Out[8]:

	experiment	date	hour	device_make	platform_os	browser	yes	no
4038	control	2020-07-03	15	Samsung SM-A600FN	6	Chrome Mobile WebView	0	0
3345	control	2020-07-03	15	Generic Smartphone	6	Chrome Mobile	0	0
3348	exposed	2020-07-03	11	Generic Smartphone	6	Chrome Mobile	0	0
3354	control	2020-07-03	15	Generic Smartphone	6	Chrome Mobile	0	0
3356	exposed	2020-07-03	8	Generic Smartphone	6	Chrome Mobile	0	0

In [9]: df2.to\_csv('fichier\_AdSmartABdata.csv')

In [ ]: