

```
In [1]: #import pandas library
import pandas as pd
```

```
In [2]: #import data
data = pd.read_csv('c:\\users\\maagalu\\Desktop\\traffic.csv')
```

```
In [3]: data
```

```
Out[3]:
```

	event	date	country	city	artist	album	track	isrc
0	click	2021-08-21	Saudi Arabia	Jeddah	Tesher	Jalebi Baby	Jalebi Baby	QZNWQ2070741
1	click	2021-08-21	Saudi Arabia	Jeddah	Tesher	Jalebi Baby	Jalebi Baby	QZNWQ2070741
2	click	2021-08-21	India	Ludhiana	Reyanna Maria	So Pretty	So Pretty	USUM72100871
3	click	2021-08-21	France	Unknown	Simone & Simaria, Sebastian Yatra	No Llores Más	No Llores Más	BRUM72003904
4	click	2021-08-21	Maldives	Malé	Tesher	Jalebi Baby	Jalebi Baby	QZNWQ2070741
...
226273	pageview	2021-08-24	Kuwait	Kuwait City	Sean Paul	The Trinity	Temperature	USAT20505520
226274	pageview	2021-08-24	India	Chennai	Miscél	when you left	when you left	QM42K1907890
226275	pageview	2021-08-24	India	Jaipur	Trippie Redd, Lil Uzi Vert	Holy Smokes (feat. Lil Uzi Vert)	Holy Smokes	QZJ842001118
226276	pageview	2021-08-24	France	Unknown	Young Thug	Tick Tock	Tick Tock	USAT22104514

	event	date	country	city	artist	album	track	isrc
226277	pageview	2021-08-24	Iraq	Duhok	Tesher	Jalebi Baby	Jalebi Baby	QZNWQ2070741
								2d8997b6-1c5fb9c1

226278 rows × 9 columns

```
In [4]: #number of rows
len(data)
```

Out[4]: 226278

```
In [5]: #column headers
data.columns
```

Out[5]: Index(['event', 'date', 'country', 'city', 'artist', 'album', 'track', 'isrc', 'linkid'], dtype='object')

```
In [6]: #number of columns
len(data.columns)
```

Out[6]: 9

```
In [7]: #top rows
data.head()
```

	event	date	country	city	artist	album	track	isrc	linkid
0	click	2021-08-21	Saudi Arabia	Jeddah	Tesher	Jalebi Baby	Jalebi Baby	QZNWQ2070741	2d896d31-97b6-4869-967b-1c5fb9cd4bb8
1	click	2021-08-21	Saudi Arabia	Jeddah	Tesher	Jalebi Baby	Jalebi Baby	QZNWQ2070741	2d896d31-97b6-4869-967b-1c5fb9cd4bb8
2	click	2021-08-21	India	Ludhiana	Reyanna Maria	So Pretty	So Pretty	USUM72100871	23199824-9cf5-4b98-942a-34965c3b0cc2
3	click	2021-08-21	France	Unknown	Simone & Simaria, Sebastian Yatra	No Llores Más	No Llores Más	BRUM72003904	35573248-4e49-47c7-af80-08a960fa74cd

	event	date	country	city	artist	album	track	isrc	linkid
4	click	2021-08-21	Maldives	Malé	Tesher	Jalebi Baby	Jalebi Baby	QZNWQ2070741	2d896d31-97b6-4869-967b-1c5fb9cd4bb8

In [8]:

```
#bottom rows
data.tail()
```

Out[8]:

	event	date	country	city	artist	album	track	isrc	linkid
226273	pageview	2021-08-24	Kuwait	Kuwait City	Sean Paul	The Trinity	Temperature	USAT20505520	04ba6b71-1055-4f0a-7f51-3cdb0d0d41
226274	pageview	2021-08-24	India	Chennai	Miscél	when you left	when you left	QM42K1907890	2fc15e81-83aa-40e1-9f51-5d96c6d435
226275	pageview	2021-08-24	India	Jaipur	Trippie Redd, Lil Uzi Vert	Holy Smokes (feat. Lil Uzi Vert)	Holy Smokes	QZJ842001118	eedc8701-6bd2-4997-260c35f72e
226276	pageview	2021-08-24	France	Unknown	Young Thug	Tick Tock	Tick Tock	USAT22104514	e0a76d01-7ccd-408e-a2c55c0e0e
226277	pageview	2021-08-24	Iraq	Duhok	Tesher	Jalebi Baby	Jalebi Baby	QZNWQ2070741	2d896d31-97b6-4869-967b-1c5fb9cd4bb8

In [9]:

```
#information on the columns
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 226278 entries, 0 to 226277
Data columns (total 9 columns):
#   Column      Non-Null Count  Dtype
---  -
0   event       226278 non-null object
1   date        226278 non-null object
2   country     226267 non-null object
3   city        226267 non-null object
4   artist      226241 non-null object
5   album       226273 non-null object
6   track       226273 non-null object
7   isrc        219157 non-null object
8   linkid      226278 non-null object
dtypes: object(9)
memory usage: 15.5+ MB
```

In [10]:

#summary
data.describe()

Out[10]:

	event	date	country	city	artist	album	track	isrc	linkid
count	226278	226278	226267	226267	226241	226273	226273	219157	226278
unique	3	7	211	11993	2419	3254	3562	709	3839
top	pageview	2021-08-19	Saudi Arabia	Jeddah	Tesher	Jalebi Baby	Jalebi Baby	QZNWQ2070741	2d896d31-97b6-4869-967b-1c5fb9cd4bb8
freq	142015	35361	47334	22791	40841	40841	40841	40841	40841

In [43]:

data = pd.read_csv('c:\\users\\maagalu\\Desktop\\traffic.csv')

In [45]:

data['event'].count()

Out[45]: 226278

In [25]:

data.date[data.event=='pageview'].count()

Out[25]: 142015

In [2]:

import pandas as pd
data = pd.read_csv('c:\\users\\maagalu\\Desktop\\traffic.csv')

In [3]:

data

Out[3]:

	event	date	country	city	artist	album	track	isrc
0	click	2021-08-21	Saudi Arabia	Jeddah	Tesher	Jalebi Baby	Jalebi Baby	QZNWQ2070741
1	click	2021-08-21	Saudi Arabia	Jeddah	Tesher	Jalebi Baby	Jalebi Baby	QZNWQ2070741
2	click	2021-08-21	India	Ludhiana	Reyanna Maria	So Pretty	So Pretty	USUM72100871
3	click	2021-08-21	France	Unknown	Simone & Simaria, Sebastian Yatra	No Llores Más	No Llores Más	BRUM72003904

	event	date	country	city	artist	album	track	isrc
4	click	2021-08-21	Maldives	Malé	Tesher	Jalebi Baby	Jalebi Baby	QZNWQ2070741
...
226273	pageview	2021-08-24	Kuwait	Kuwait City	Sean Paul	The Trinity	Temperature	USAT20505520
226274	pageview	2021-08-24	India	Chennai	Miscél	when you left	when you left	QM42K1907890
226275	pageview	2021-08-24	India	Jaipur	Trippie Redd, Lil Uzi Vert	Holy Smokes (feat. Lil Uzi Vert)	Holy Smokes	QZJ842001118
226276	pageview	2021-08-24	France	Unknown	Young Thug	Tick Tock	Tick Tock	USAT22104514
226277	pageview	2021-08-24	Iraq	Duhok	Tesher	Jalebi Baby	Jalebi Baby	QZNWQ2070741

226278 rows × 9 columns

```
In [4]: data.date[data.event=='pageview'].count()
```

Out[4]: 142015

```
In [5]: data.date[data.event=='click'].count()
```

Out[5]: 55732

```
In [59]: #click rate
clickrate=data.date[data.event=='click'].count()/data.date[data.event=='pageview'].count()
```

```
In [60]: clickrate
```

Out[60]: 0.3924374185825441

```
In [12]: # pageview countries
```

```
data.loc[data['event'] == 'pageview', 'country'].drop_duplicates()
```

```
Out[12]: 84043          Saudi Arabia
84044          United States
84046          Ireland
84047          United Kingdom
84051          France
...
165434          Afghanistan
176541  Central African Republic
200553          Guernsey
216014          Sint Maarten
223904          Saint Martin
Name: country, Length: 212, dtype: object
```

```
In [14]: data.groupby('event')['date'].count()
```

```
Out[14]: event
click          55732
pageview       142015
preview        28531
Name: date, dtype: int64
```

```
In [20]: data.groupby('date')['event'].count()
```

```
Out[20]: date
2021-08-19    35361
2021-08-20    34112
2021-08-21    34083
2021-08-22    32633
2021-08-23    29808
2021-08-24    29834
2021-08-25    30447
Name: event, dtype: int64
```

```
In [38]: events=data.groupby(['date', 'event'])['event'].count()
```

```
In [39]: events
```

```
Out[39]: date    event
2021-08-19  click          9207
           pageview       22366
           preview        3788
2021-08-20  click          8508
           pageview       21382
           preview        4222
2021-08-21  click          8071
           pageview       21349
           preview        4663
2021-08-22  click          7854
           pageview       20430
           preview        4349
2021-08-23  click          7315
           pageview       18646
           preview        3847
2021-08-24  click          7301
           pageview       18693
           preview        3840
2021-08-25  click          7476
```

```
pageview    19149
preview     3822
Name: event, dtype: int64
```

```
In [42]: # pageviews per day
df=data.groupby('date')['event'].apply(lambda x: (x=='pageview').sum()).reset_index(nam
```

```
In [43]: df
```

```
Out[43]:
```

	date	pageviews
0	2021-08-19	22366
1	2021-08-20	21382
2	2021-08-21	21349
3	2021-08-22	20430
4	2021-08-23	18646
5	2021-08-24	18693
6	2021-08-25	19149

```
In [44]: df1=data.groupby('date')['event'].apply(lambda x: (x=='click').sum()).reset_index(name=
```

```
In [45]: df1
```

```
Out[45]:
```

	date	clicks_per_day
0	2021-08-19	9207
1	2021-08-20	8508
2	2021-08-21	8071
3	2021-08-22	7854
4	2021-08-23	7315
5	2021-08-24	7301
6	2021-08-25	7476

```
In [46]: df2=data.groupby('date')['event'].apply(lambda x: (x=='preview').sum()).reset_index(nam
```

```
In [47]: df2
```

```
Out[47]:
```

	date	previews_per_day
0	2021-08-19	3788
1	2021-08-20	4222

	date	previews_per_day
2	2021-08-21	4663
3	2021-08-22	4349
4	2021-08-23	3847
5	2021-08-24	3840
6	2021-08-25	3822

In [56]: `data.date[data.event=='pageview'].count()`

Out[56]: 142015

In [53]: `data.date[data.event=='click'].count()`

Out[53]: 55732

In [54]: `data.date[data.event=='preview'].count()`

Out[54]: 28531

In [58]: `data.head()`

Out[58]:

	event	date	country	city	artist	album	track	isrc	linkid
0	click	2021-08-21	Saudi Arabia	Jeddah	Tesher	Jalebi Baby	Jalebi Baby	QZNWQ2070741	2d896d31-97b6-4869-967b-1c5fb9cd4bb8
1	click	2021-08-21	Saudi Arabia	Jeddah	Tesher	Jalebi Baby	Jalebi Baby	QZNWQ2070741	2d896d31-97b6-4869-967b-1c5fb9cd4bb8
2	click	2021-08-21	India	Ludhiana	Reyanna Maria	So Pretty	So Pretty	USUM72100871	23199824-9cf5-4b98-942a-34965c3b0cc2
3	click	2021-08-21	France	Unknown	Simone & Simaria, Sebastian Yatra	No Llores Más	No Llores Más	BRUM72003904	35573248-4e49-47c7-af80-08a960fa74cd
4	click	2021-08-21	Maldives	Malé	Tesher	Jalebi Baby	Jalebi Baby	QZNWQ2070741	2d896d31-97b6-4869-967b-1c5fb9cd4bb8

In [68]: `data.groupby('event')['linkid'].count()`


```
Out[68]: event
click      55732
pageview   142015
preview    28531
Name: linkid, dtype: int64
```

```
In [69]: data['clickrate']=data.date[data.event=='click'].count()/data.date[data.event=='pagevie
```

```
In [70]: data.head()
```

Out[70]:

	event	date	country	city	artist	album	track	isrc	linkid	clickrate
0	click	2021-08-21	Saudi Arabia	Jeddah	Tesher	Jalebi Baby	Jalebi Baby	QZNWQ2070741	2d896d31-97b6-4869-967b-1c5fb9cd4bb8	0.392437
1	click	2021-08-21	Saudi Arabia	Jeddah	Tesher	Jalebi Baby	Jalebi Baby	QZNWQ2070741	2d896d31-97b6-4869-967b-1c5fb9cd4bb8	0.392437
2	click	2021-08-21	India	Ludhiana	Reyanna Maria	So Pretty	So Pretty	USUM72100871	23199824-9cf5-4b98-942a-34965c3b0cc2	0.392437
3	click	2021-08-21	France	Unknown	Simone & Simaria, Sebastian Yatra	No Llores Más	No Llores Más	BRUM72003904	35573248-4e49-47c7-af80-08a960fa74cd	0.392437
4	click	2021-08-21	Maldives	Malé	Tesher	Jalebi Baby	Jalebi Baby	QZNWQ2070741	2d896d31-97b6-4869-967b-1c5fb9cd4bb8	0.392437

```
In [72]: del data['clickrate']
```

```
In [73]: data.head()
```

Out[73]:

	event	date	country	city	artist	album	track	isrc	linkid
0	click	2021-08-21	Saudi Arabia	Jeddah	Tesher	Jalebi Baby	Jalebi Baby	QZNWQ2070741	2d896d31-97b6-4869-967b-1c5fb9cd4bb8
1	click	2021-08-21	Saudi Arabia	Jeddah	Tesher	Jalebi Baby	Jalebi Baby	QZNWQ2070741	2d896d31-97b6-4869-967b-1c5fb9cd4bb8

	event	date	country	city	artist	album	track	isrc	linkid
2	click	2021-08-21	India	Ludhiana	Reyanna Maria	So Pretty	So Pretty	USUM72100871	23199824-9cf5-4b98-942a-34965c3b0cc2
3	click	2021-08-21	France	Unknown	Simone & Simaria, Sebastian Yatra	No Llores Más	No Llores Más	BRUM72003904	35573248-4e49-47c7-af80-08a960fa74cd
4	click	2021-08-21	Maldives	Malé	Tesher	Jalebi Baby	Jalebi Baby	QZNWQ2070741	2d896d31-97b6-4869-967b-1c5fb9cd4bb8

In [79]: `d=data.groupby(['linkid','event'])['event'].count()`

In [80]: `d`

Out[80]:

linkid	event	
00073307-ae96-5089-a117-4783afb42f8e	pageview	2
00126b32-0c35-507b-981c-02c80d2aa8e7	click	2
	pageview	2
0018cfff-50a1-5984-9715-01ef2d11a49a	pageview	1
0033934b-5d16-5a06-af58-d087bcdd3680	pageview	1
	..	
fff84c0e-90a1-59d8-9997-adc909d50e16	pageview	1
fffc17a7-f935-5d3e-bd3e-d761fd80d479	click	1
	pageview	2
fffd0045-29de-522b-b5d8-35786363bf07	click	1
	pageview	2

Name: event, Length: 6587, dtype: int64

In [84]: `data.describe()`

Out[84]:

	event	date	country	city	artist	album	track	isrc	linkid
count	226278	226278	226267	226267	226241	226273	226273	219157	226278
unique	3	7	211	11993	2419	3254	3562	709	3839
top	pageview	2021-08-19	Saudi Arabia	Jeddah	Tesher	Jalebi Baby	Jalebi Baby	QZNWQ2070741	2d896d31-97b6-4869-967b-1c5fb9cd4bb8
freq	142015	35361	47334	22791	40841	40841	40841	40841	40841

In [2]: `import pandas as pd`
`data=pd.read_csv('c:\\users\\maagalu\\Desktop\\traffic.csv')`

In [3]: `data.head()`

Out[3]:

	event	date	country	city	artist	album	track	isrc	linkid
0	click	2021-08-21	Saudi Arabia	Jeddah	Tesher	Jalebi Baby	Jalebi Baby	QZNWQ2070741	2d896d31-97b6-4869-967b-1c5fb9cd4bb8
1	click	2021-08-21	Saudi Arabia	Jeddah	Tesher	Jalebi Baby	Jalebi Baby	QZNWQ2070741	2d896d31-97b6-4869-967b-1c5fb9cd4bb8
2	click	2021-08-21	India	Ludhiana	Reyanna Maria	So Pretty	So Pretty	USUM72100871	23199824-9cf5-4b98-942a-34965c3b0cc2
3	click	2021-08-21	France	Unknown	Simone & Simaria, Sebastian Yatra	No Llores Más	No Llores Más	BRUM72003904	35573248-4e49-47c7-af80-08a960fa74cd
4	click	2021-08-21	Maldives	Malé	Tesher	Jalebi Baby	Jalebi Baby	QZNWQ2070741	2d896d31-97b6-4869-967b-1c5fb9cd4bb8

In [49]: `pv=data.groupby('linkid')['event'].apply(lambda x: (x=='pageview').sum()).reset_index(n`

In [50]: `pv`

Out[50]:

	linkid	pageviews
0	00073307-ae96-5089-a117-4783afb42f8e	2
1	00126b32-0c35-507b-981c-02c80d2aa8e7	2
2	0018cff-50a1-5984-9715-01ef2d11a49a	1
3	0033934b-5d16-5a06-af58-d087bccdd3680	1
4	0034d6cf-3bd8-5ffe-aa-afc-b3959fc48608	1
...
3834	fff38ca0-8043-50cd-a5f1-f65ebb7105c5	1
3835	fff4e5f0-4ee5-5fe7-aa30-e870edaf6ed7	2
3836	fff84c0e-90a1-59d8-9997-adc909d50e16	1
3837	fffc17a7-f935-5d3e-bd3e-d761fd80d479	2
3838	fffd0045-29de-522b-b5d8-35786363bf07	2

3839 rows × 2 columns

In [24]: `data['pageviews']=pv`

In [25]:

```
data.tail()
```

Out[25]:

	event	date	country	city	artist	album	track	isrc	linkid
226273	pageview	2021-08-24	Kuwait	Kuwait City	Sean Paul	The Trinity	Temperature	USAT20505520	04ba6b1-1055-4f0a75-3cdb0d0d41
226274	pageview	2021-08-24	India	Chennai	Miscél	when you left	when you left	QM42K1907890	2fc15e8-83aa-40e9f5-5d96c6d435
226275	pageview	2021-08-24	India	Jaipur	Trippie Redd, Lil Uzi Vert	Holy Smokes (feat. Lil Uzi Vert)	Holy Smokes	QZJ842001118	eedc870-6bd2-49e97-260c35f72a
226276	pageview	2021-08-24	France	Unknown	Young Thug	Tick Tock	Tick Tock	USAT22104514	e0a76dc-7ccd-408e-a2c55c0e0e
226277	pageview	2021-08-24	Iraq	Duhok	Tesher	Jalebi Baby	Jalebi Baby	QZNWQ2070741	2d896d31-97b6-4869-967b-1c5fb9cd4b

In [26]:

```
del data['pageviews']
```

In [27]:

```
data.head()
```

Out[27]:

	event	date	country	city	artist	album	track	isrc	linkid
0	click	2021-08-21	Saudi Arabia	Jeddah	Tesher	Jalebi Baby	Jalebi Baby	QZNWQ2070741	2d896d31-97b6-4869-967b-1c5fb9cd4bb8
1	click	2021-08-21	Saudi Arabia	Jeddah	Tesher	Jalebi Baby	Jalebi Baby	QZNWQ2070741	2d896d31-97b6-4869-967b-1c5fb9cd4bb8
2	click	2021-08-21	India	Ludhiana	Reyanna Maria	So Pretty	So Pretty	USUM72100871	23199824-9cf5-4b98-942a-34965c3b0cc2
3	click	2021-08-21	France	Unknown	Simone & Simaria, Sebastian Yatra	No Llores Más	No Llores Más	BRUM72003904	35573248-4e49-47c7-af80-08a960fa74cd

	event	date	country	city	artist	album	track	isrc	linkid
4	click	2021-08-21	Maldives	Malé	Tesher	Jalebi Baby	Jalebi Baby	QZNWQ2070741	2d896d31-97b6-4869-967b-1c5fb9cd4bb8

```
In [28]: c=data.groupby('linkid')['event'].apply(lambda x: (x=='click').sum()).reset_index(name=
```

```
In [29]: c
```

Out[29]:

	linkid	clicks
0	00073307-ae96-5089-a117-4783afb42f8e	0
1	00126b32-0c35-507b-981c-02c80d2aa8e7	2
2	0018cff-50a1-5984-9715-01ef2d11a49a	0
3	0033934b-5d16-5a06-af58-d087bcdd3680	0
4	0034d6cf-3bd8-5ffe-aafe-b3959fc48608	0
...
3834	fff38ca0-8043-50cd-a5f1-f65ebb7105c5	1
3835	fff4e5f0-4ee5-5fe7-aa30-e870edaf6ed7	0
3836	fff84c0e-90a1-59d8-9997-adc909d50e16	1
3837	fffc17a7-f935-5d3e-bd3e-d761fd80d479	1
3838	fffd0045-29de-522b-b5d8-35786363bf07	1

3839 rows × 2 columns

```
In [57]: b=data.date[data.event=='click'].count()/data.date[data.event=='pageview'].count()
```

```
In [58]: b
```

Out[58]: 0.3924374185825441

```
In [66]: pv=data.groupby(['linkid','event'])['event'].apply(lambda x: (x=='pageview').sum()).res
```

```
In [61]: pv
```

Out[61]:

	linkid	event	pageviews
0	00073307-ae96-5089-a117-4783afb42f8e	pageview	2
1	00126b32-0c35-507b-981c-02c80d2aa8e7	click	0

	linkid	event	pageviews
2	00126b32-0c35-507b-981c-02c80d2aa8e7	pageview	2
3	0018cfff-50a1-5984-9715-01ef2d11a49a	pageview	1
4	0033934b-5d16-5a06-af58-d087bcdd3680	pageview	1
...
6582	fff84c0e-90a1-59d8-9997-adc909d50e16	pageview	1
6583	fffc17a7-f935-5d3e-bd3e-d761fd80d479	click	0
6584	fffc17a7-f935-5d3e-bd3e-d761fd80d479	pageview	2
6585	fffd0045-29de-522b-b5d8-35786363bf07	click	0
6586	fffd0045-29de-522b-b5d8-35786363bf07	pageview	2

6587 rows × 3 columns

In [69]: `c=data.groupby(['linkid','event'])['event'].apply(lambda x: (x=='click').sum()).reset_i`

In [70]: `c`

Out[70]:

	linkid	event	clicks
0	00073307-ae96-5089-a117-4783afb42f8e	pageview	0
1	00126b32-0c35-507b-981c-02c80d2aa8e7	click	2
2	00126b32-0c35-507b-981c-02c80d2aa8e7	pageview	0
3	0018cfff-50a1-5984-9715-01ef2d11a49a	pageview	0
4	0033934b-5d16-5a06-af58-d087bcdd3680	pageview	0
...
6582	fff84c0e-90a1-59d8-9997-adc909d50e16	pageview	0
6583	fffc17a7-f935-5d3e-bd3e-d761fd80d479	click	1
6584	fffc17a7-f935-5d3e-bd3e-d761fd80d479	pageview	0
6585	fffd0045-29de-522b-b5d8-35786363bf07	click	1
6586	fffd0045-29de-522b-b5d8-35786363bf07	pageview	0

6587 rows × 3 columns

In [72]: `frames=[pv,c]`

In [73]: `result = pd.concat(frames)`

In [74]: `result`

Out[74]:

	linkid	event	pageviews	clicks
0	00073307-ae96-5089-a117-4783afb42f8e	pageview	2.0	NaN
1	00126b32-0c35-507b-981c-02c80d2aa8e7	click	0.0	NaN
2	00126b32-0c35-507b-981c-02c80d2aa8e7	pageview	2.0	NaN
3	0018cff-50a1-5984-9715-01ef2d11a49a	pageview	1.0	NaN
4	0033934b-5d16-5a06-af58-d087bccdd3680	pageview	1.0	NaN
...
6582	fff84c0e-90a1-59d8-9997-adc909d50e16	pageview	NaN	0.0
6583	fffc17a7-f935-5d3e-bd3e-d761fd80d479	click	NaN	1.0
6584	fffc17a7-f935-5d3e-bd3e-d761fd80d479	pageview	NaN	0.0
6585	fffd0045-29de-522b-b5d8-35786363bf07	click	NaN	1.0
6586	fffd0045-29de-522b-b5d8-35786363bf07	pageview	NaN	0.0

13174 rows × 4 columns

In [4]: `q=data.groupby('linkid')['event'].apply(lambda x: (x=='click').sum()).reset_index(name=`

In [76]: `w=data.groupby('linkid')['event'].apply(lambda x: (x=='pageview').sum()).reset_index(na`

In [5]: `q`

Out[5]:

	linkid	clicks
0	00073307-ae96-5089-a117-4783afb42f8e	0
1	00126b32-0c35-507b-981c-02c80d2aa8e7	2
2	0018cff-50a1-5984-9715-01ef2d11a49a	0
3	0033934b-5d16-5a06-af58-d087bccdd3680	0
4	0034d6cf-3bd8-5ffe-aa5c-b3959fc48608	0
...
3834	fff38ca0-8043-50cd-a5f1-f65ebb7105c5	1
3835	fff4e5f0-4ee5-5fe7-aa30-e870edaf6ed7	0
3836	fff84c0e-90a1-59d8-9997-adc909d50e16	1
3837	fffc17a7-f935-5d3e-bd3e-d761fd80d479	1
3838	fffd0045-29de-522b-b5d8-35786363bf07	1

3839 rows × 2 columns

```
In [77]: frames = (q,w)
```

```
In [82]: final = pd.merge(q,w, on='linkid')
```

```
In [83]: final
```

```
Out[83]:
```

	linkid	clicks	pageviews
0	00073307-ae96-5089-a117-4783afb42f8e	0	2
1	00126b32-0c35-507b-981c-02c80d2aa8e7	2	2
2	0018cff-50a1-5984-9715-01ef2d11a49a	0	1
3	0033934b-5d16-5a06-af58-d087bcdd3680	0	1
4	0034d6cf-3bd8-5ffe-aafc-b3959fc48608	0	1
...
3834	fff38ca0-8043-50cd-a5f1-f65ebb7105c5	1	1
3835	fff4e5f0-4ee5-5fe7-aa30-e870edaf6ed7	0	2
3836	fff84c0e-90a1-59d8-9997-adc909d50e16	1	1
3837	fffc17a7-f935-5d3e-bd3e-d761fd80d479	1	2
3838	fffd0045-29de-522b-b5d8-35786363bf07	1	2

3839 rows × 3 columns

```
In [96]: df.drop('crate', axis=1, inplace=True)
```

```
In [97]: final
```

```
Out[97]:
```

	linkid	clicks	pageviews	crate	CR_across_link	clickrate
0	00073307-ae96-5089-a117-4783afb42f8e	0	2	0.0	0.0	0.0
1	00126b32-0c35-507b-981c-02c80d2aa8e7	2	2	1.0	1.0	1.0
2	0018cff-50a1-5984-9715-01ef2d11a49a	0	1	0.0	0.0	0.0
3	0033934b-5d16-5a06-af58-d087bcdd3680	0	1	0.0	0.0	0.0
4	0034d6cf-3bd8-5ffe-aafc-b3959fc48608	0	1	0.0	0.0	0.0
...
3834	fff38ca0-8043-50cd-a5f1-f65ebb7105c5	1	1	1.0	1.0	1.0
3835	fff4e5f0-4ee5-5fe7-aa30-e870edaf6ed7	0	2	0.0	0.0	0.0
3836	fff84c0e-90a1-59d8-9997-adc909d50e16	1	1	1.0	1.0	1.0
3837	fffc17a7-f935-5d3e-bd3e-d761fd80d479	1	2	0.5	0.5	0.5
3838	fffd0045-29de-522b-b5d8-35786363bf07	1	2	0.5	0.5	0.5

3839 rows × 6 columns

In [98]:

df

Out[98]:

	linkid	clicks	pageviews	CR_across_link	clickrate
0	00073307-ae96-5089-a117-4783afb42f8e	0	2	0.0	0.0
1	00126b32-0c35-507b-981c-02c80d2aa8e7	2	2	1.0	1.0
2	0018cff-50a1-5984-9715-01ef2d11a49a	0	1	0.0	0.0
3	0033934b-5d16-5a06-af58-d087bcdd3680	0	1	0.0	0.0
4	0034d6cf-3bd8-5ffe-aafe-b3959fc48608	0	1	0.0	0.0
...
3834	fff38ca0-8043-50cd-a5f1-f65ebb7105c5	1	1	1.0	1.0
3835	fff4e5f0-4ee5-5fe7-aa30-e870edaf6ed7	0	2	0.0	0.0
3836	fff84c0e-90a1-59d8-9997-adc909d50e16	1	1	1.0	1.0
3837	fffc17a7-f935-5d3e-bd3e-d761fd80d479	1	2	0.5	0.5
3838	fffd0045-29de-522b-b5d8-35786363bf07	1	2	0.5	0.5

3839 rows × 5 columns

In [99]:

df.drop('CR_across_link', axis=1, inplace=True)

In [100]:

df

Out[100]:

	linkid	clicks	pageviews	clickrate
0	00073307-ae96-5089-a117-4783afb42f8e	0	2	0.0
1	00126b32-0c35-507b-981c-02c80d2aa8e7	2	2	1.0
2	0018cff-50a1-5984-9715-01ef2d11a49a	0	1	0.0
3	0033934b-5d16-5a06-af58-d087bcdd3680	0	1	0.0
4	0034d6cf-3bd8-5ffe-aafe-b3959fc48608	0	1	0.0
...
3834	fff38ca0-8043-50cd-a5f1-f65ebb7105c5	1	1	1.0
3835	fff4e5f0-4ee5-5fe7-aa30-e870edaf6ed7	0	2	0.0
3836	fff84c0e-90a1-59d8-9997-adc909d50e16	1	1	1.0
3837	fffc17a7-f935-5d3e-bd3e-d761fd80d479	1	2	0.5
3838	fffd0045-29de-522b-b5d8-35786363bf07	1	2	0.5

3839 rows × 4 columns

```
In [7]: df.clickrate[df.clickrate==1].count()
```

```
-----
NameError                                Traceback (most recent call last)
<ipython-input-7-cdbadfba3e90> in <module>
----> 1 df.clickrate[df.clickrate==1].count()

NameError: name 'df' is not defined
```

```
In [112]: df.clickrate[df.clickrate==0.5].count()
```

```
Out[112]: 318
```

```
In [113]: df.clickrate[df.clickrate==0].count()
```

```
Out[113]: 1584
```

```
In [120]: prev=data.groupby('linkid')['event'].apply(lambda x: (x=='preview').sum()).reset_index()
```

```
In [121]: click=data.groupby('linkid')['event'].apply(lambda x: (x=='click').sum()).reset_index(n
```

```
In [122]: df1 = pd.merge(prev,click, on='linkid')
```

```
In [123]: df1
```

```
Out[123]:
```

	linkid	preview	click
0	00073307-ae96-5089-a117-4783afb42f8e	0	0
1	00126b32-0c35-507b-981c-02c80d2aa8e7	0	2
2	0018cff-50a1-5984-9715-01ef2d11a49a	0	0
3	0033934b-5d16-5a06-af58-d087bccdd3680	0	0
4	0034d6cf-3bd8-5ffe-aafe-b3959fc48608	0	0
...
3834	fff38ca0-8043-50cd-a5f1-f65ebb7105c5	0	1
3835	fff4e5f0-4ee5-5fe7-aa30-e870edaf6ed7	0	0
3836	fff84c0e-90a1-59d8-9997-adc909d50e16	0	1
3837	fffc17a7-f935-5d3e-bd3e-d761fd80d479	0	1
3838	fffd0045-29de-522b-b5d8-35786363bf07	0	1

3839 rows × 3 columns

```
In [124]: df1.preview[df1.preview==1].count()
```

Out[124...] 123

In [125...] `df1.click[df1.click==1].count()`

Out[125...] 1428

In [126...]

```
import numpy as np
import scipy.stats
```

In [130...]

```
import matplotlib
import matplotlib.pyplot as pp

import pandas.plotting

from IPython import display

%matplotlib inline
```

In [131...]

```
import re
import mailbox
import csv
```

In [132...] `df1.head()`

	linkid	preview	click
0	00073307-ae96-5089-a117-4783afb42f8e	0	0
1	00126b32-0c35-507b-981c-02c80d2aa8e7	0	2
2	0018cfff-50a1-5984-9715-01ef2d11a49a	0	0
3	0033934b-5d16-5a06-af58-d087bcdd3680	0	0
4	0034d6cf-3bd8-5ffe-aafe-b3959fc48608	0	0

In [137...] `df1.tail()`

	linkid	preview	click
3834	fff38ca0-8043-50cd-a5f1-f65ebb7105c5	0	1
3835	fff4e5f0-4ee5-5fe7-aa30-e870edaf6ed7	0	0
3836	fff84c0e-90a1-59d8-9997-adc909d50e16	0	1
3837	fffc17a7-f935-5d3e-bd3e-d761fd80d479	0	1
3838	fffd0045-29de-522b-b5d8-35786363bf07	0	1

In [140...] `pd.DataFrame(df1.preview.value_counts())`

Out[140...

preview	
0	3344
1	123
2	47
3	29
4	19
...	...
212	1
166	1
152	1
150	1
1605	1

120 rows × 1 columns

In [138...

```
df1.preview[df1.preview==1]
```

Out[138...

11	1
21	1
46	1
62	1
94	1
..	
3728	1
3753	1
3770	1
3778	1
3808	1

Name: preview, Length: 123, dtype: int64

In [141...

```
df1.value_counts()
```

Out[141...

linkid	preview	click
00073307-ae96-5089-a117-4783afb42f8e	0	0
a7b383bb-c406-5c86-899f-b8580e596021	0	0
a6c33729-486e-5b43-8a5d-976a3cbba0ab	0	1
a6f76bd9-6065-4d02-88f4-158d783e05b0	13	19
a6f7963f-4813-5bbb-b76c-225065dd5f32	0	1
..		
548b6882-44dd-446a-b6bc-46252c5ff5f6	0	0
54e0a47a-c0da-5b7f-a9ae-d26a04e7bf7e	0	0
5503691c-6a30-5b19-b499-425f28351583	0	2
550c17a8-235a-510a-aeb5-616426a0a67d	0	0
fffd0045-29de-522b-b5d8-35786363bf07	0	1

Length: 3839, dtype: int64

In [145...

```
pd.DataFrame(df1.click.value_counts(normalize=True))
```

Out[145...

	click
0	0.412607
1	0.371972
2	0.070331
3	0.025267
4	0.015108
...	...
1364	0.000260
878	0.000260
468	0.000260
434	0.000260
1429	0.000260

171 rows × 1 columns

In [146...

```
pd.DataFrame(df1.click.value_counts())
```

Out[146...

	click
0	1584
1	1428
2	270
3	97
4	58
...	...
1364	1
878	1
468	1
434	1
1429	1

171 rows × 1 columns

In [147...

```
bypreview = df1.groupby('preview').click.value_counts(normalize=True)
```

In [148...

```
bypreview
```

Out[148...

preview	click
0	0.464414

```
1      0.412380
2      0.074761
3      0.023026
4      0.009868
...
849    1111    1.000000
1001   1429    1.000000
1571   2198    1.000000
1605   2109    1.000000
5974   9692    1.000000
Name: click, Length: 347, dtype: float64
```

In [150...

bypreview.unstack()

Out[150...

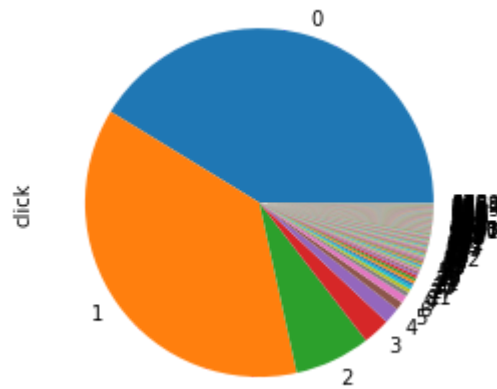
click	0	1	2	3	4	5	6	7	8	
preview										
0	0.464414	0.412380	0.074761	0.023026	0.009868	0.005084	0.001794	0.000299	0.000897	0.0014
1	0.195122	0.300813	0.121951	0.097561	0.081301	0.040650	0.065041	0.008130	0.008130	0.0243
2	0.106383	0.191489	0.063830	0.085106	0.148936	0.042553	0.063830	NaN	0.042553	N
3	0.034483	0.068966	0.034483	0.103448	0.034483	0.103448	0.103448	0.103448	0.068966	0.1034
4	0.052632	NaN	0.052632	NaN	0.263158	0.052632	0.105263	0.052632	NaN	0.0526
...
849	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	N
1001	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	N
1571	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	N
1605	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	N
5974	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	N

120 rows × 171 columns

In [152...

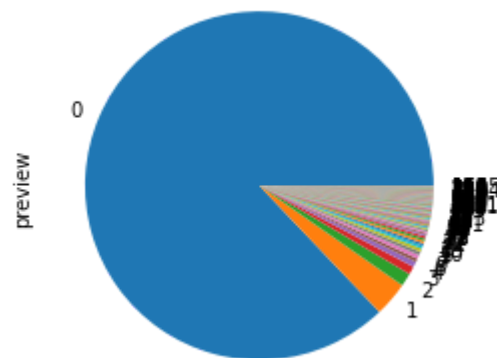
df1.click.value_counts().plot(kind='pie')

Out[152... <AxesSubplot:ylabel='click'>



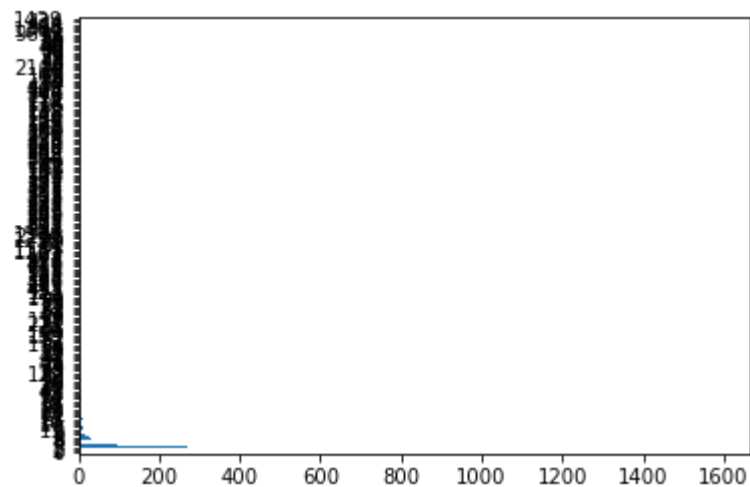
```
In [153... df1.preview.value_counts().plot(kind='pie')
```

```
Out[153... <AxesSubplot:ylabel='preview'>
```



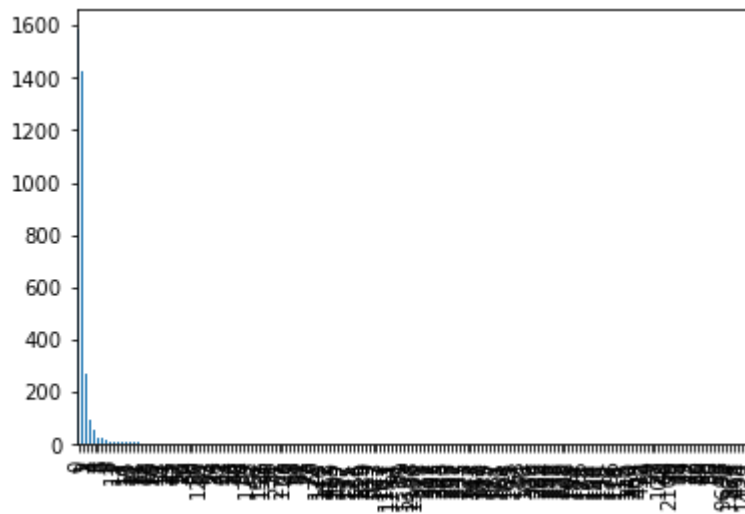
```
In [156... df1.click.value_counts().plot(kind='barh')
```

```
Out[156... <AxesSubplot:>
```



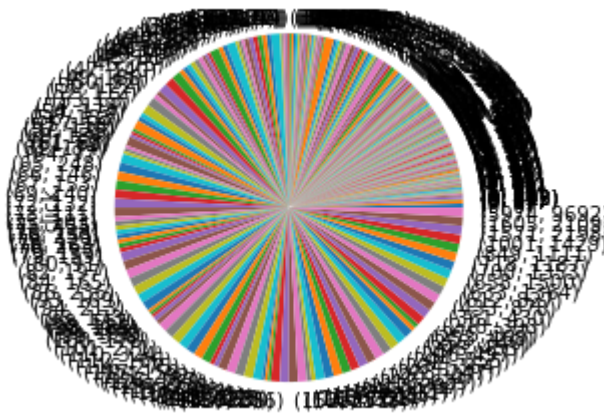
```
In [157... df1.click.value_counts().plot(kind='bar')
```

Out[157... <AxesSubplot:>



In [158... bypreview.plot(kind='pie')

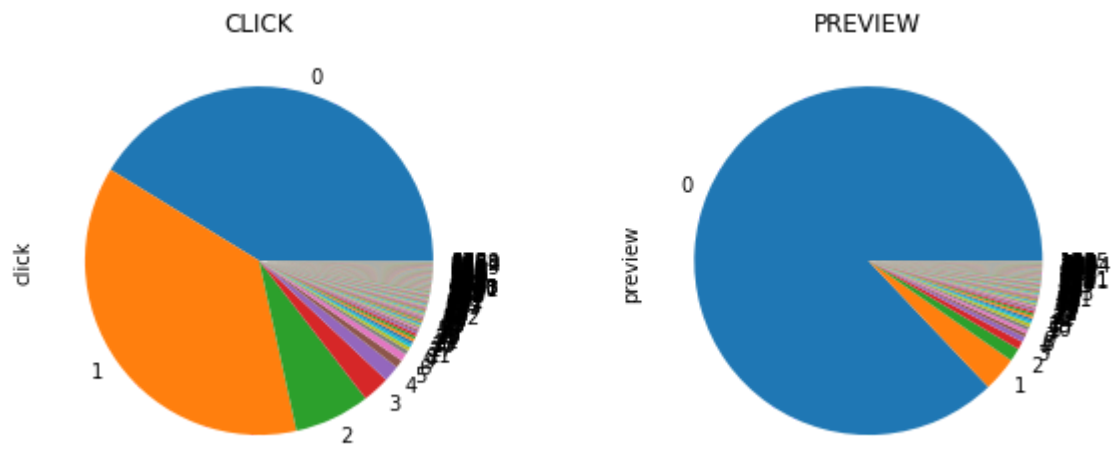
Out[158... <AxesSubplot:ylabel='click'>



```
In [161... pp.figure(figsize=(10,4))

pp.subplot(1,2,1);df1.click.value_counts().plot(kind='pie');pp.title('CLICK')
pp.subplot(1,2,2);df1.preview.value_counts().plot(kind='pie');pp.title('PREVIEW')
```

Out[161... Text(0.5, 1.0, 'PREVIEW')



	click
3	97
4	58
...	...
1364	1
878	1
468	1
434	1
1429	1

171 rows × 1 columns

In [168...

```
import numpy as np
r = np.corrcoef(x, y)
```

In [169...

```
r
```

Out[169... array([[nan, nan, nan, ..., nan, nan, nan],
[nan, nan, nan, ..., nan, nan, nan],
[nan, nan, nan, ..., nan, nan, nan],
...,
[nan, nan, nan, ..., nan, nan, nan],
[nan, nan, nan, ..., nan, nan, nan],
[nan, nan, nan, ..., nan, nan, nan]])

In [172...

```
df1
```

Out[172...

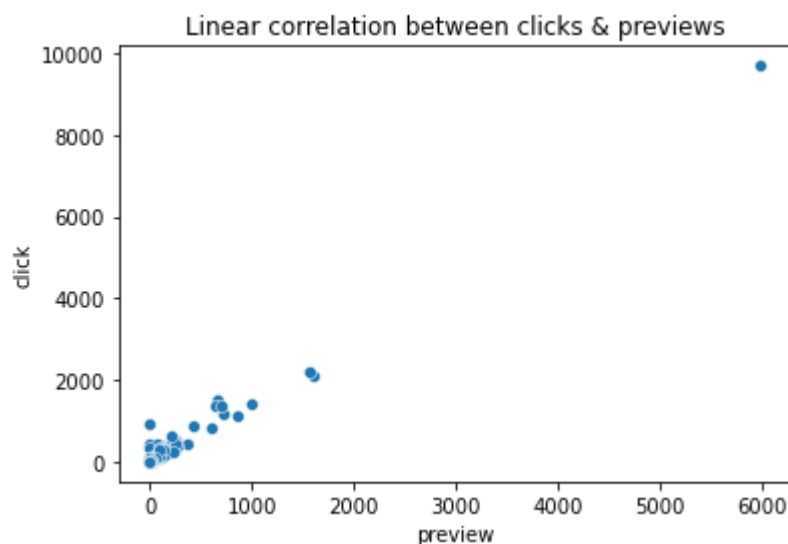
	linkid	preview	click
0	00073307-ae96-5089-a117-4783afb42f8e	0	0
1	00126b32-0c35-507b-981c-02c80d2aa8e7	0	2
2	0018cfff-50a1-5984-9715-01ef2d11a49a	0	0
3	0033934b-5d16-5a06-af58-d087bcdd3680	0	0
4	0034d6cf-3bd8-5ffe-aafc-b3959fc48608	0	0
...
3834	fff38ca0-8043-50cd-a5f1-f65ebb7105c5	0	1
3835	fff4e5f0-4ee5-5fe7-aa30-e870edaf6ed7	0	0
3836	fff84c0e-90a1-59d8-9997-adc909d50e16	0	1
3837	fffc17a7-f935-5d3e-bd3e-d761fd80d479	0	1
3838	fffd0045-29de-522b-b5d8-35786363bf07	0	1

3839 rows × 3 columns

In [175...

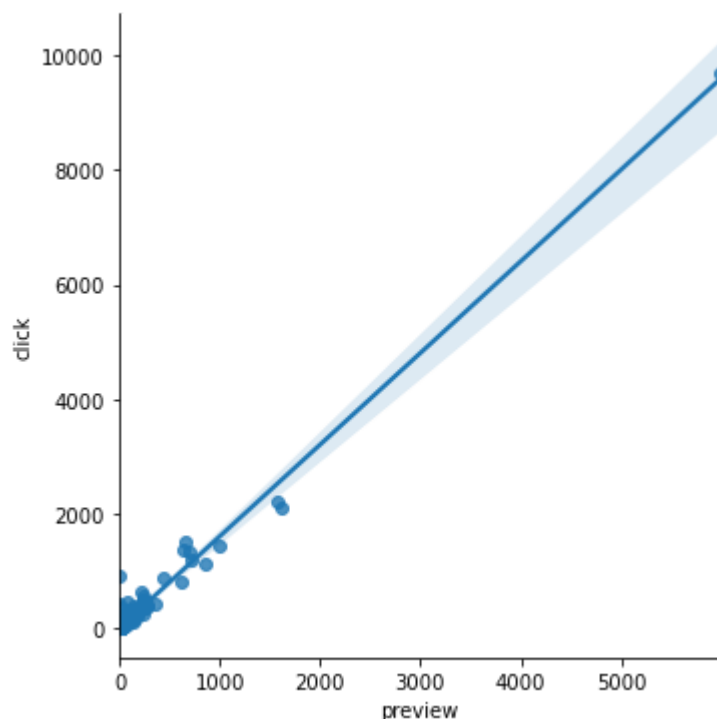
```
import seaborn as sns
pc=sns.scatterplot(x="preview", y="click", data=df1);
pc.set_title("Linear correlation between clicks & previews")
```

Out[175... Text(0.5, 1.0, 'Linear correlation between clicks & previews')



In [177...

```
sns.lmplot(x="preview", y="click", data=df1);
```



In [180...

```
from scipy import stats
stats.pearsonr(df1['preview'], df1['click'])
```

Out[180... (0.9886586274883709, 0.0)

In [181...

```
data.describe()
```

Out[181]...

	event	date	country	city	artist	album	track	isrc	linkid
count	226278	226278	226267	226267	226241	226273	226273	219157	226278
unique	3	7	211	11993	2419	3254	3562	709	3839
top	pageview	2021-08-19	Saudi Arabia	Jeddah	Tesher	Jalebi Baby	Jalebi Baby	QZNWQ2070741	2d896d31-97b6-4869-967b-1c5fb9cd4bb8
freq	142015	35361	47334	22791	40841	40841	40841	40841	40841

In [26]:

```
import pandas as pd
import numpy as np
data = pd.read_csv('c:\\users\\maagalu\\Desktop\\traffic.csv')
```

In [31]:

```
data
```

Out[31]:

	event	date	country	city	artist	album	track	isrc	linkid
0	click	8/21/2021	Saudi Arabia	Jeddah	Tesher	Jalebi Baby	Jalebi Baby	QZNWQ2070741	2d896d31-97b6-4869-967b-1c5fb9cd4bb8
1	click	8/21/2021	Saudi Arabia	Jeddah	Tesher	Jalebi Baby	Jalebi Baby	QZNWQ2070741	2d896d31-97b6-4869-967b-1c5fb9cd4bb8
2	click	8/21/2021	India	Ludhiana	Reyanna Maria	So Pretty	So Pretty	USUM72100871	2d896d31-97b6-4869-967b-1c5fb9cd4bb8
3	click	8/21/2021	France	Unknown	Simone & Simaria, Sebastian Yatra	No Llores Más	No Llores Más	BRUM72003904	2d896d31-97b6-4869-967b-1c5fb9cd4bb8
4	click	8/21/2021	Maldives	Malé	Tesher	Jalebi Baby	Jalebi Baby	QZNWQ2070741	2d896d31-97b6-4869-967b-1c5fb9cd4bb8
...
226273	pageview	8/24/2021	Kuwait	Kuwait City	Sean Paul	The Trinity	Temperature	USAT20505520	2d896d31-97b6-4869-967b-1c5fb9cd4bb8
226274	pageview	8/24/2021	India	Chennai	Miscél	when you left	when you left	QM42K1907890	2d896d31-97b6-4869-967b-1c5fb9cd4bb8

	event	date	country	city	artist	album	track	isrc	
226275	pageview	8/24/2021	India	Jaipur	Trippie Redd, Lil Uzi Vert	Holy Smokes (feat. Lil Uzi Vert)	Holy Smokes	QZJ842001118	6 26
226276	pageview	8/24/2021	France	Unknown	Young Thug	Tick Tock	Tick Tock	USAT22104514	a2c
226277	pageview	8/24/2021	Iraq	Duhok	Tesher	Jalebi Baby	Jalebi Baby	QZNWQ2070741	9 1c5

226278 rows × 9 columns

In [32]:

data['pageview']=0

In [33]:

data.head()

Out[33]:

	event	date	country	city	artist	album	track	isrc	linkid	pageview
0	click	8/21/2021	Saudi Arabia	Jeddah	Tesher	Jalebi Baby	Jalebi Baby	QZNWQ2070741	2d896d31-97b6-4869-967b-1c5fb9cd4bb8	
1	click	8/21/2021	Saudi Arabia	Jeddah	Tesher	Jalebi Baby	Jalebi Baby	QZNWQ2070741	2d896d31-97b6-4869-967b-1c5fb9cd4bb8	
2	click	8/21/2021	India	Ludhiana	Reyanna Maria	So Pretty	So Pretty	USUM72100871	23199824-9cf5-4b98-942a-34965c3b0cc2	
3	click	8/21/2021	France	Unknown	Simone & Simaria, Sebastian Yatra	No Llores Más	No Llores Más	BRUM72003904	35573248-4e49-47c7-af80-08a960fa74cd	
4	click	8/21/2021	Maldives	Malé	Tesher	Jalebi Baby	Jalebi Baby	QZNWQ2070741	2d896d31-97b6-4869-967b-1c5fb9cd4bb8	

In [34]:

data.loc[data.event == "pageview", "pageview"] = 1

In [35]:

```
data.tail()
```

Out[35]:

	event	date	country	city	artist	album	track	isrc	
226273	pageview	8/24/2021	Kuwait	Kuwait City	Sean Paul	The Trinity	Temperature	USAT20505520	04b1053cdb0d
226274	pageview	8/24/2021	India	Chennai	Miscél	when you left	when you left	QM42K1907890	2fc83a55d96c6
226275	pageview	8/24/2021	India	Jaipur	Trippie Redd, Lil Uzi Vert	Holy Smokes (feat. Lil Uzi Vert)	Holy Smokes	QZJ842001118	eec6bd260c3
226276	pageview	8/24/2021	France	Unknown	Young Thug	Tick Tock	Tick Tock	USAT22104514	e0a7cc1a2c55c
226277	pageview	8/24/2021	Iraq	Duhok	Tesher	Jalebi Baby	Jalebi Baby	QZNWQ2070741	2d897b61c5fb9

In [36]:

```
data['click']=0
```

In [37]:

```
data['preview']=0
```

In [38]:

```
data.loc[data.event == "click", "click"] = 1
data.loc[data.event == "preview", "preview"] = 1
```

In [39]:

```
data
```

Out[39]:

	event	date	country	city	artist	album	track	isrc	
0	click	8/21/2021	Saudi Arabia	Jeddah	Tesher	Jalebi Baby	Jalebi Baby	QZNWQ2070741	91c5
1	click	8/21/2021	Saudi Arabia	Jeddah	Tesher	Jalebi Baby	Jalebi Baby	QZNWQ2070741	91c5

	event	date	country	city	artist	album	track	isrc	
2	click	8/21/2021	India	Ludhiana	Reyanna Maria	So Pretty	So Pretty	USUM72100871	349
3	click	8/21/2021	France	Unknown	Simone & Simaria, Sebastian Yatra	No Llores Más	No Llores Más	BRUM72003904	08a
4	click	8/21/2021	Maldives	Malé	Tesher	Jalebi Baby	Jalebi Baby	QZNWQ2070741	91c5
...
226273	pageview	8/24/2021	Kuwait	Kuwait City	Sean Paul	The Trinity	Temperature	USAT20505520	3cdt
226274	pageview	8/24/2021	India	Chennai	Miscél	when you left	when you left	QM42K1907890	5d9
226275	pageview	8/24/2021	India	Jaipur	Trippie Redd, Lil Uzi Vert	Holy Smokes (feat. Lil Uzi Vert)	Holy Smokes	QZJ842001118	626l
226276	pageview	8/24/2021	France	Unknown	Young Thug	Tick Tock	Tick Tock	USAT22104514	a2c
226277	pageview	8/24/2021	Iraq	Duhok	Tesher	Jalebi Baby	Jalebi Baby	QZNWQ2070741	91c5

226278 rows × 12 columns

```
In [43]: df=data.groupby('linkid')[['click','pageview','preview']].count()

In [44]: df1=data.groupby('linkid')[['click','pageview','preview']].sum()

In [45]: df

Out[45]:
```

	click	pageview	preview
--	-------	----------	---------

linkid	click	pageview	preview
linkid			
00073307-ae96-5089-a117-4783afb42f8e	2	2	2
00126b32-0c35-507b-981c-02c80d2aa8e7	4	4	4
0018cfff-50a1-5984-9715-01ef2d11a49a	1	1	1
0033934b-5d16-5a06-af58-d087bcdd3680	1	1	1
0034d6cf-3bd8-5ffe-aafc-b3959fc48608	1	1	1
...
fff38ca0-8043-50cd-a5f1-f65ebb7105c5	2	2	2
fff4e5f0-4ee5-5fe7-aa30-e870edaf6ed7	2	2	2
fff84c0e-90a1-59d8-9997-adc909d50e16	2	2	2
fffc17a7-f935-5d3e-bd3e-d761fd80d479	3	3	3
fffd0045-29de-522b-b5d8-35786363bf07	3	3	3

3839 rows × 3 columns

In [46]:

```
df1
```

Out[46]:

linkid	click	pageview	preview
00073307-ae96-5089-a117-4783afb42f8e	0	2	0
00126b32-0c35-507b-981c-02c80d2aa8e7	2	2	0
0018cfff-50a1-5984-9715-01ef2d11a49a	0	1	0
0033934b-5d16-5a06-af58-d087bcdd3680	0	1	0
0034d6cf-3bd8-5ffe-aafc-b3959fc48608	0	1	0
...
fff38ca0-8043-50cd-a5f1-f65ebb7105c5	1	1	0
fff4e5f0-4ee5-5fe7-aa30-e870edaf6ed7	0	2	0
fff84c0e-90a1-59d8-9997-adc909d50e16	1	1	0
fffc17a7-f935-5d3e-bd3e-d761fd80d479	1	2	0
fffd0045-29de-522b-b5d8-35786363bf07	1	2	0

3839 rows × 3 columns

In [49]:

```
x=df1.list('pageview')
```


AttributeError

Traceback (most recent call last)

<ipython-input-49-e7cd084e2525> in <module>

----> 1 x=df1.list('pageview')

~\Anaconda3\lib\site-packages\pandas\core\generic.py in __getattr__(self, name)

5463 if self._info_axis._can_hold_identifiers_and_holds_name(name):

5464 return self[name]

-> 5465 return object.__getattr__(self, name)

5466

5467 def __setattr__(self, name: str, value) -> None:

AttributeError: 'DataFrame' object has no attribute 'list'

In [50]:

data

Out[50]:

	event	date	country	city	artist	album	track	isrc	
0	click	8/21/2021	Saudi Arabia	Jeddah	Tesher	Jalebi Baby	Jalebi Baby	QZNWQ2070741	91c5
1	click	8/21/2021	Saudi Arabia	Jeddah	Tesher	Jalebi Baby	Jalebi Baby	QZNWQ2070741	91c5
2	click	8/21/2021	India	Ludhiana	Reyanna Maria	So Pretty	So Pretty	USUM72100871	349
3	click	8/21/2021	France	Unknown	Simone & Simaria, Sebastian Yatra	No Llores Más	No Llores Más	BRUM72003904	08a
4	click	8/21/2021	Maldives	Malé	Tesher	Jalebi Baby	Jalebi Baby	QZNWQ2070741	91c5
...
226273	pageview	8/24/2021	Kuwait	Kuwait City	Sean Paul	The Trinity	Temperature	USAT20505520	3cdt
226274	pageview	8/24/2021	India	Chennai	Miscél	when you left	when you left	QM42K1907890	5d9
226275	pageview	8/24/2021	India	Jaipur	Trippie Redd, Lil Uzi Vert	Holy Smokes (feat. Lil Uzi Vert)	Holy Smokes	QZJ842001118	626

	event	date	country	city	artist	album	track	isrc	
226276	pageview	8/24/2021	France	Unknown	Young Thug	Tick Tock	Tick Tock	USAT22104514	a2c
226277	pageview	8/24/2021	Iraq	Duhok	Tesher	Jalebi Baby	Jalebi Baby	QZNWQ2070741	9
									1c5

226278 rows × 12 columns

In [51]:

data.info()

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 226278 entries, 0 to 226277
Data columns (total 12 columns):
#   Column      Non-Null Count  Dtype
---  -
0   event       226278 non-null object
1   date        226278 non-null object
2   country     226267 non-null object
3   city        226267 non-null object
4   artist      226241 non-null object
5   album       226273 non-null object
6   track       226273 non-null object
7   isrc        219157 non-null object
8   linkid      226278 non-null object
9   pageview    226278 non-null int64
10  click       226278 non-null int64
11  preview     226278 non-null int64
dtypes: int64(3), object(9)
memory usage: 20.7+ MB
```

In [55]:

data.groupby('linkid')['click'].sum()

```
Out[55]: linkid
00073307-ae96-5089-a117-4783afb42f8e    0
00126b32-0c35-507b-981c-02c80d2aa8e7    2
0018cfff-50a1-5984-9715-01ef2d11a49a    0
0033934b-5d16-5a06-af58-d087bcd3680     0
0034d6cf-3bd8-5ffe-aafc-b3959fc48608    0
..
fff38ca0-8043-50cd-a5f1-f65ebb7105c5    1
fff4e5f0-4ee5-5fe7-aa30-e870edaf6ed7    0
fff84c0e-90a1-59d8-9997-adc909d50e16    1
fffc17a7-f935-5d3e-bd3e-d761fd80d479    1
fffd0045-29de-522b-b5d8-35786363bf07    1
Name: click, Length: 3839, dtype: int64
```

In [56]:

data.groupby('linkid')['pageview'].sum()

```
Out[56]: linkid
00073307-ae96-5089-a117-4783afb42f8e    2
00126b32-0c35-507b-981c-02c80d2aa8e7    2
0018cfff-50a1-5984-9715-01ef2d11a49a    1
```

```
0033934b-5d16-5a06-af58-d087bcdd3680    1
0034d6cf-3bd8-5ffe-aafc-b3959fc48608    1
..
fff38ca0-8043-50cd-a5f1-f65ebb7105c5    1
fff4e5f0-4ee5-5fe7-aa30-e870edaf6ed7    2
fff84c0e-90a1-59d8-9997-adc909d50e16    1
fffc17a7-f935-5d3e-bd3e-d761fd80d479    2
fffd0045-29de-522b-b5d8-35786363bf07    2
Name: pageview, Length: 3839, dtype: int64
```

```
In [77]: q=data.groupby('linkid')[['pageview','click','preview']].sum()
```

```
In [78]: q
```

Out[78]:

	pageview	click	preview
linkid			
00073307-ae96-5089-a117-4783afb42f8e	2	0	0
00126b32-0c35-507b-981c-02c80d2aa8e7	2	2	0
0018cfff-50a1-5984-9715-01ef2d11a49a	1	0	0
0033934b-5d16-5a06-af58-d087bcdd3680	1	0	0
0034d6cf-3bd8-5ffe-aafc-b3959fc48608	1	0	0
...
fff38ca0-8043-50cd-a5f1-f65ebb7105c5	1	1	0
fff4e5f0-4ee5-5fe7-aa30-e870edaf6ed7	2	0	0
fff84c0e-90a1-59d8-9997-adc909d50e16	1	1	0
fffc17a7-f935-5d3e-bd3e-d761fd80d479	2	1	0
fffd0045-29de-522b-b5d8-35786363bf07	2	1	0

3839 rows × 3 columns

```
In [79]: sorted_df = q.sort_values("pageview", ascending=False)
```

```
In [80]: sorted_df
```

Out[80]:

	pageview	click	preview
linkid			
2d896d31-97b6-4869-967b-1c5fb9cd4bb8	25175	9692	5974
522da5cc-8177-4140-97a7-a84fdb4caf1c	6600	2109	1605
e849515b-929d-44c8-a505-e7622f1827e9	5981	2198	1571
c2c876ab-b093-4750-9449-6b4913da6af3	4303	1429	1001
23199824-9cf5-4b98-942a-34965c3b0cc2	3532	1187	718

	pageview	click	preview
linkid			
...
8c71ba08-d449-521e-8092-5d4f7e14d759	1	0	0
8c7849a7-cb1f-5482-ae81-043546086f2e	1	0	0
2a20c79c-7578-5247-878b-a6b71fba3769	1	1	0
54166799-1895-4f35-9b2f-b249c2f7a351	0	1	0
aee2b83d-5f50-4309-9e62-200c404d4751	0	1	0

3839 rows × 3 columns

```
In [81]: sorted_df['clickrate']=sorted_df['click']/sorted_df['pageview']
print(sorted_df)
```

linkid	pageview	click	preview	clickrate
2d896d31-97b6-4869-967b-1c5fb9cd4bb8	25175	9692	5974	0.384985
522da5cc-8177-4140-97a7-a84fdb4caf1c	6600	2109	1605	0.319545
e849515b-929d-44c8-a505-e7622f1827e9	5981	2198	1571	0.367497
c2c876ab-b093-4750-9449-6b4913da6af3	4303	1429	1001	0.332094
23199824-9cf5-4b98-942a-34965c3b0cc2	3532	1187	718	0.336070
...
8c71ba08-d449-521e-8092-5d4f7e14d759	1	0	0	0.000000
8c7849a7-cb1f-5482-ae81-043546086f2e	1	0	0	0.000000
2a20c79c-7578-5247-878b-a6b71fba3769	1	1	0	1.000000
54166799-1895-4f35-9b2f-b249c2f7a351	0	1	0	inf
aee2b83d-5f50-4309-9e62-200c404d4751	0	1	0	inf

[3839 rows x 4 columns]

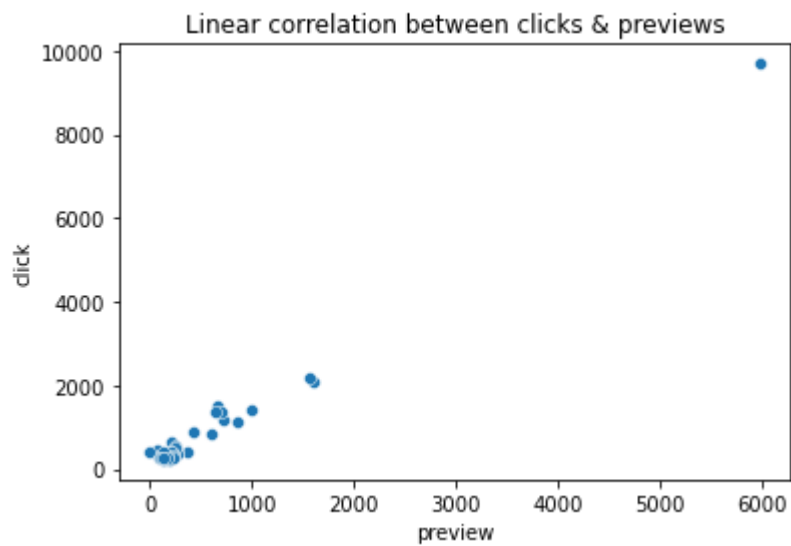
```
In [82]: sorted_df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Index: 3839 entries, 2d896d31-97b6-4869-967b-1c5fb9cd4bb8 to aee2b83d-5f50-4309-9e62-200c404d4751
Data columns (total 4 columns):
#   Column      Non-Null Count  Dtype
---  -
0   pageview    3839 non-null   int64
1   click       3839 non-null   int64
2   preview     3839 non-null   int64
3   clickrate   3839 non-null   float64
dtypes: float64(1), int64(3)
memory usage: 150.0+ KB
```

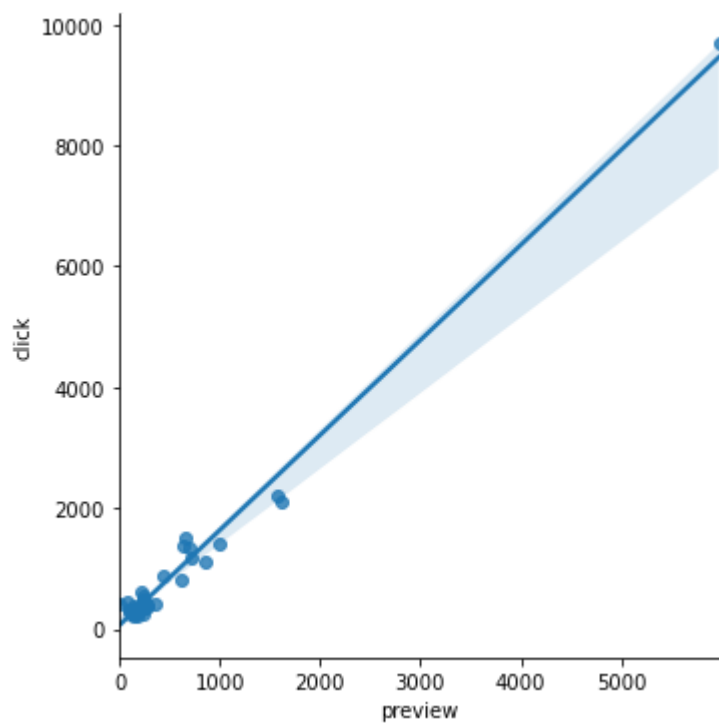
```
In [85]: lc=sorted_df.head(40)
```

```
In [86]: import seaborn as sns
pc=sns.scatterplot(x="preview", y="click", data=lc);
pc.set_title("Linear correlation between clicks & previews")
```

```
Out[86]: Text(0.5, 1.0, 'Linear correlation between clicks & previews')
```



```
In [87]: sns.lmplot(x="preview", y="click", data=lc);
```



```
In [88]: import scipy as sp
import matplotlib
import matplotlib.pyplot as pp

import pandas.plotting

from IPython import display

%matplotlib inline
```

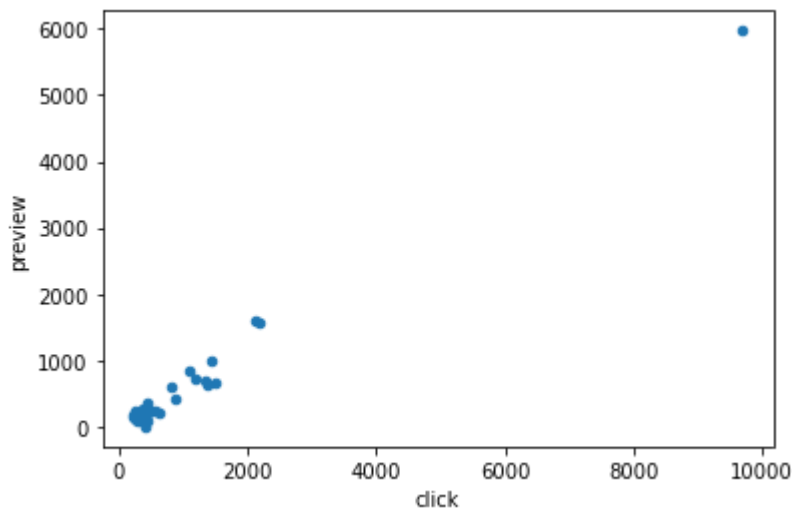
```
In [89]: lc.corr()
```

Out[89]:

	pageview	click	preview	clickrate
pageview	1.000000	0.997149	0.997690	-0.046261
click	0.997149	1.000000	0.992586	0.010673
preview	0.997690	0.992586	1.000000	-0.073922
clickrate	-0.046261	0.010673	-0.073922	1.000000

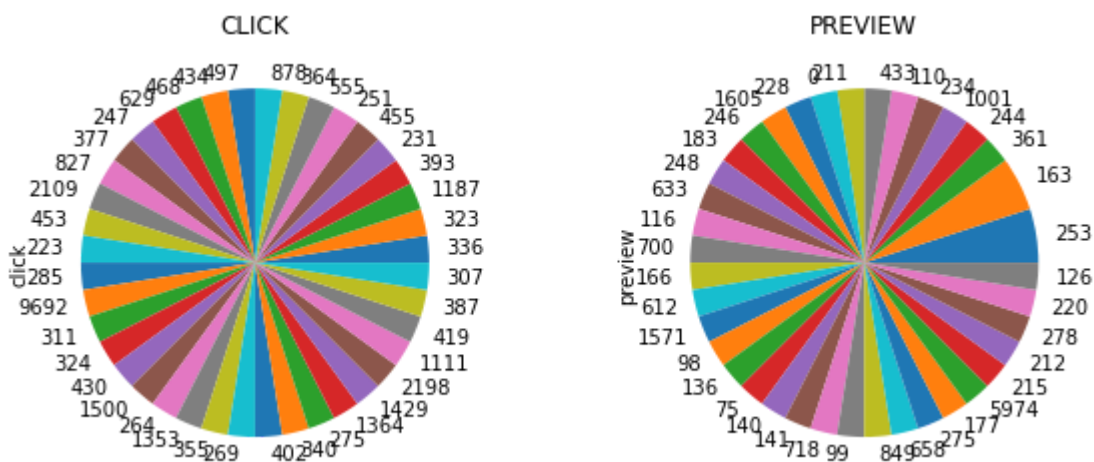
In [90]: `lc.plot.scatter(x='click',y='preview')`

Out[90]: `<AxesSubplot:xlabel='click', ylabel='preview'>`



In [91]: `pp.figure(figsize=(10,4))`
`pp.subplot(1,2,1);lc.click.value_counts().plot(kind='pie');pp.title('CLICK')`
`pp.subplot(1,2,2);lc.preview.value_counts().plot(kind='pie');pp.title('PREVIEW')`

Out[91]: `Text(0.5, 1.0, 'PREVIEW')`



In [92]: `from scipy.stats import pearsonr`

```
In [93]: #correlation coefficient and P-value
pearsonr(lc['click'], lc['preview'])
```

```
Out[93]: (0.9925857273425244, 2.150237831306553e-36)
```

```
In [94]: sorted_df.corr()
```

```
Out[94]:
```

	pageview	click	preview	clickrate
pageview	1.000000	0.994001	0.996691	-0.004248
click	0.994001	1.000000	0.988659	0.076821
preview	0.996691	0.988659	1.000000	-0.004378
clickrate	-0.004248	0.076821	-0.004378	1.000000

```
In [95]: pearsonr(sorted_df['click'], sorted_df['preview'])
```

```
Out[95]: (0.9886586274883703, 0.0)
```

```
In [96]: sorted_df
```

```
Out[96]:
```

	pageview	click	preview	clickrate
linkid				
2d896d31-97b6-4869-967b-1c5fb9cd4bb8	25175	9692	5974	0.384985
522da5cc-8177-4140-97a7-a84fdb4caf1c	6600	2109	1605	0.319545
e849515b-929d-44c8-a505-e7622f1827e9	5981	2198	1571	0.367497
c2c876ab-b093-4750-9449-6b4913da6af3	4303	1429	1001	0.332094
23199824-9cf5-4b98-942a-34965c3b0cc2	3532	1187	718	0.336070
...
8c71ba08-d449-521e-8092-5d4f7e14d759	1	0	0	0.000000
8c7849a7-cb1f-5482-ae81-043546086f2e	1	0	0	0.000000
2a20c79c-7578-5247-878b-a6b71fba3769	1	1	0	1.000000
54166799-1895-4f35-9b2f-b249c2f7a351	0	1	0	inf
aee2b83d-5f50-4309-9e62-200c404d4751	0	1	0	inf

3839 rows × 4 columns

```
In [99]: sorted_df.click.mean()
```

```
Out[99]: 14.51732221932795
```

```
In [100]: sorted_df.preview.mean()
```

Out[100... 7.431883302943475

In [101... `from scipy.stats import ttest_1samp`

In [104... `_, p_value=ttest_1samp(a=sorted_df.click, popmean=sorted_df.preview.mean())`

In [105... `p_value`

Out[105... 0.013765691480951268

In [1]:
`import pandas as pd`
`import numpy as np`
`import scipy as sp`
`import scipy.stats`
`from scipy import stats`
`from scipy.stats import chi2`

In [2]: `events=pd.read_csv('c:\\users\\maagalu\\Desktop\\traffic.csv')`

In [3]:
`events['pageview']=0`
`events['click']=0`
`events['preview']=0`

In [4]:
`events.loc[events.event == "pageview", "pageview"] = 1`
`events.loc[events.event == "click", "click"] = 1`
`events.loc[events.event == "preview", "preview"] = 1`

In [5]: `events.head()`

Out[5]:

	event	date	country	city	artist	album	track	isrc	linkid	pagev
0	click	8/21/2021	Saudi Arabia	Jeddah	Tesher	Jalebi Baby	Jalebi Baby	QZNWQ2070741	2d896d31-97b6-4869-967b-1c5fb9cd4bb8	
1	click	8/21/2021	Saudi Arabia	Jeddah	Tesher	Jalebi Baby	Jalebi Baby	QZNWQ2070741	2d896d31-97b6-4869-967b-1c5fb9cd4bb8	
2	click	8/21/2021	India	Ludhiana	Reyanna Maria	So Pretty	So Pretty	USUM72100871	23199824-9cf5-4b98-942a-34965c3b0cc2	

	event	date	country	city	artist	album	track	isrc	linkid	pageview
3	click	8/21/2021	France	Unknown	Simone & Simaria, Sebastian Yatra	No Llores Más	No Llores Más	BRUM72003904	35573248-4e49-47c7-af80-08a960fa74cd	
4	click	8/21/2021	Maldives	Malé	Tesher	Jalebi Baby	Jalebi Baby	QZNWQ2070741	2d896d31-97b6-4869-967b-1c5fb9cd4bb8	

```
In [6]: events_table=pd.crosstab(events['linkid'],events['event'])
print(events_table)
```

event linkid	click	pageview	preview
00073307-ae96-5089-a117-4783afb42f8e	0	2	0
00126b32-0c35-507b-981c-02c80d2aa8e7	2	2	0
0018cfff-50a1-5984-9715-01ef2d11a49a	0	1	0
0033934b-5d16-5a06-af58-d087bcdd3680	0	1	0
0034d6cf-3bd8-5ffe-aafc-b3959fc48608	0	1	0
...
fff38ca0-8043-50cd-a5f1-f65ebb7105c5	1	1	0
fff4e5f0-4ee5-5fe7-aa30-e870edaf6ed7	0	2	0
fff84c0e-90a1-59d8-9997-adc909d50e16	1	1	0
fffc17a7-f935-5d3e-bd3e-d761fd80d479	1	2	0
fffd0045-29de-522b-b5d8-35786363bf07	1	2	0

[3839 rows x 3 columns]

```
In [7]: events_table.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Index: 3839 entries, 00073307-ae96-5089-a117-4783afb42f8e to fffd0045-29de-522b-b5d8-35786363bf07
Data columns (total 3 columns):
#   Column      Non-Null Count  Dtype
---  -
0   click       3839 non-null   int64
1   pageview    3839 non-null   int64
2   preview     3839 non-null   int64
dtypes: int64(3)
memory usage: 120.0+ KB
```

```
In [8]: observed_values=events_table.values
```

```
In [9]: observed_values
```

```
Out[9]: array([[0, 2, 0],
               [2, 2, 0],
               [0, 1, 0],
               ...,
               [1, 1, 0],
               [1, 2, 0],
               [1, 2, 0]], dtype=int64)
```

```
In [10]: val=stats.chi2_contingency(events_table)
```

```
In [11]: val
```

```
Out[11]: (9210.213737609443,  
1.1847342971735872e-31,  
7676,  
array([[0.4925976 , 1.25522587, 0.25217653],  
       [0.9851952 , 2.51045175, 0.50435305],  
       [0.2462988 , 0.62761294, 0.12608826],  
       ...,  
       [0.4925976 , 1.25522587, 0.25217653],  
       [0.7388964 , 1.88283881, 0.37826479],  
       [0.7388964 , 1.88283881, 0.37826479]]))
```

```
In [12]: expected_values=val[3]
```

```
In [13]: chi_square=sum([(o-e)**2./e for o,e in zip(observed_values,expected_values)])  
chi_square_statistic=chi_square[0]+chi_square[1]
```

```
In [14]: print("chi-square statistic:-",chi_square_statistic)
```

```
chi-square statistic:- 6153.871493092706
```

```
In [16]: alpha=0.05
```

```
In [18]: ddof=1
```

```
In [19]: critical_value=chi2.ppf(q=1-alpha,df=ddof)  
print('critical_value:',critical_value)
```

```
critical_value: 3.841458820694124
```

```
In [20]: p_value=1-chi2.cdf(x=chi_square_statistic,df=ddof)  
print('p_value:',p_value)  
print('significance level:',alpha)
```

```
p_value: 0.0  
significance level: 0.05
```

```
In [21]: if chi_square_statistic>=critical_value:  
    print('Reject H0, There is a relationship between 2 categorical variables')  
else:  
    print('Retain H0, There is no relationship between 2 categorical variables')  
  
if p_value<=alpha:  
    print('Reject H0, There is a relationship between 2 categorical variables')  
else:  
    print('Retain H0, There is no relationship between 2 categorical variables')
```

Reject H_0 , There is a relationship between 2 categorical variables
Reject H_0 , There is a relationship between 2 categorical variables

In []: