

```
In [1]: #import Libraries
import pandas as pd
import numpy as np
```

```
In [2]: #load the dataset into the kernel and view how it looks
traffic = pd.read_csv("traffic.csv")
traffic
```

Out[2]:		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
	...	...	...	...	...	...	...	...	...	...
	226273	pageview	2021-08-24	Kuwait	Kuwait City	Sean Paul	The Trinity	Temperature	USAT20505520	04ba6b17-1055-4f0e-a75e-3cdb0d0d4106
	226274	pageview	2021-08-24	India	Chennai	Miscél	when you left	when you left	QM42K1907890	2fc15e8b-83aa-40ee-9f9d-5d96c6d4356c
	226275	pageview	2021-08-24	India	Jaipur	Trippie Redd, Lil Uzi Vert	Holy Smokes (feat. Lil Uzi Vert)	Holy Smokes	QZJ842001118	eedc8702-6bd2-499a-972f-260c35f72e0f
	226276	pageview	2021-08-24	France	Unknown	Young Thug	Tick Tock	Tick Tock	USAT22104514	e0a76dc3-7ccd-405f-8ee8-a2c55c0e0ed2
	226277	pageview	2021-08-24	Iraq	Duhok	Tesher	Jalebi Baby	Jalebi Baby	QZNWQ2070741	2d896d31-97b6-4869-967b-1c5fb9cd4bb8

226278 rows × 9 columns

In [3]: `traffic.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
4	click	2021-08-21	Maldives	Malé	Tesher	Jalebi Baby	Jalebi Baby	QZNWQ2070741	2d896d31-97b6-4869-967b-1c5fb9cd4bb8

In [4]: `traffic[traffic["event"] == "pageview"]`

## Linkfire Project

Out[4]:

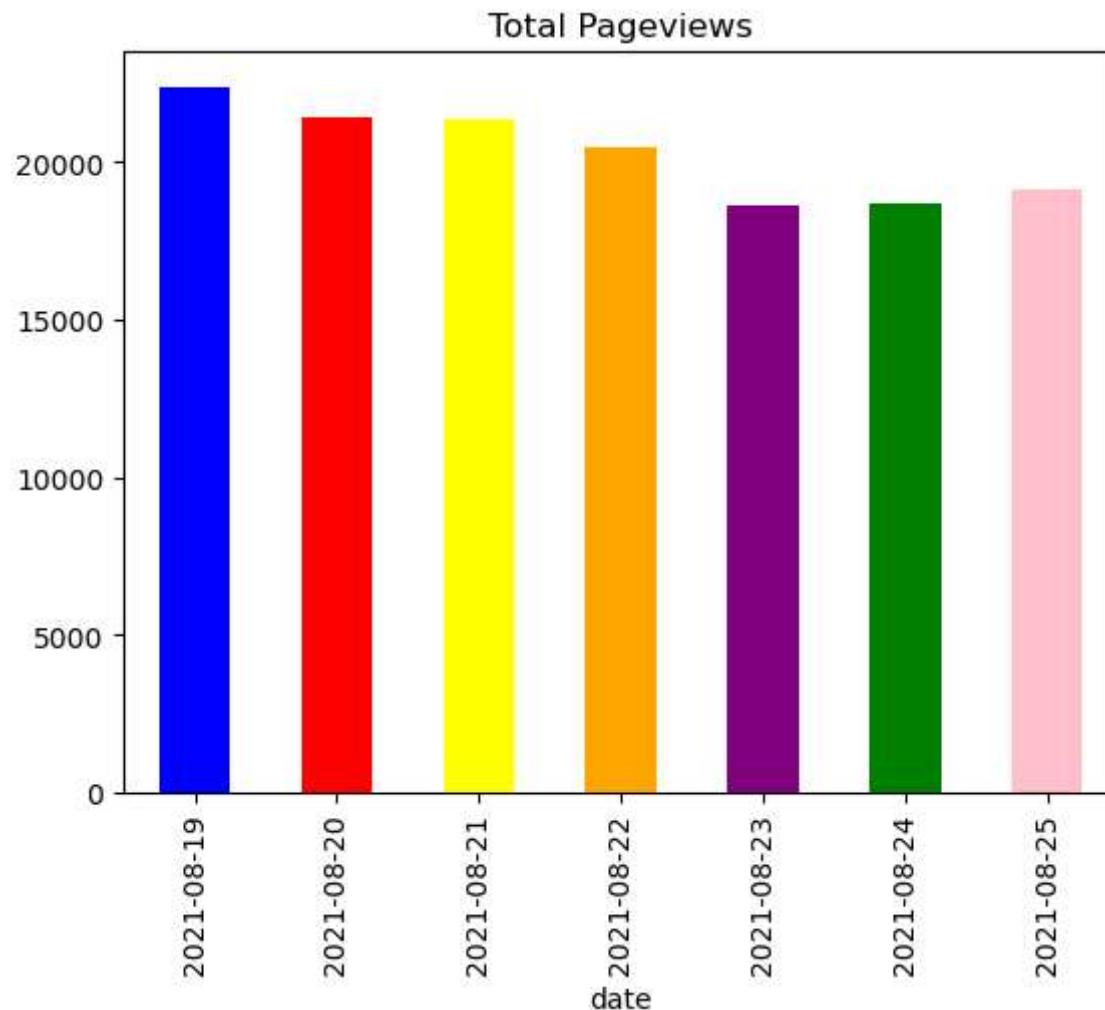
	event	date	country	city	artist	album	track	isrc	linkid
84043	pageview	2021-08-19	Saudi Arabia	Riyadh	Tesher	Jalebi Baby	Jalebi Baby	QZNWQ2070741	2d896d31-97b6-4869-967b-1c5fb9cd4bb8
84044	pageview	2021-08-19	United States	Absecon	KA\$HDAMI	epiphany	Reparations!	QZFYY2115255	9c61dba1-9369-4ee4-a215-1d34581cd811
84045	pageview	2021-08-19	Saudi Arabia	Riyadh	Tesher	Jalebi Baby	Jalebi Baby	QZNWQ2070741	2d896d31-97b6-4869-967b-1c5fb9cd4bb8
84046	pageview	2021-08-19	Ireland	Mullingar	Cardi B	Up	Up	USAT22100061	9d847c0e-6824-438f-9dc1-d76fca6dc87c
84047	pageview	2021-08-19	United Kingdom	Northampton	Kayla Nicole, Taylor Girlz, Flo Milli	Bundles 2 (feat. Flo Milli, Taylor Girlz)	Bundles 2 (feat. Flo Milli, Taylor Girlz)	QMCE32100359	fbf4b935-f961-4b13-b1d8-45ad47093559
...	...	...	...	...	...	...	...	...	...
226273	pageview	2021-08-24	Kuwait	Kuwait City	Sean Paul	The Trinity	Temperature	USAT20505520	04ba6b17-1055-4f0e-a75e-3cdb0d0d4106
226274	pageview	2021-08-24	India	Chennai	Miscél	when you left	when you left	QM42K1907890	2fc15e8b-83aa-40ee-9f9d-5d96c6d4356c
226275	pageview	2021-08-24	India	Jaipur	Trippie Redd, Lil Uzi Vert	Holy Smokes (feat. Lil Uzi Vert)	Holy Smokes	QZJ842001118	eedc8702-6bd2-499a-972f-260c35f72e0f
226276	pageview	2021-08-24	France	Unknown	Young Thug	Tick Tock	Tick Tock	USAT22104514	e0a76dc3-7ccd-405f-8ee8-a2c55c0e0ed2
226277	pageview	2021-08-24	Iraq	Duhok	Tesher	Jalebi Baby	Jalebi Baby	QZNWQ2070741	2d896d31-97b6-4869-967b-1c5fb9cd4bb8

142015 rows × 9 columns

```
In [5]: # QUESTION 1: How many total pageview events did the links in the provided dataset receive in the full period, how many  
pageview = traffic[traffic["event"] == "pageview"].groupby("date")["event"].count()
```

```
In [47]: colors= ['blue','red','yellow','orange','purple','green','pink']  
pageview.plot(kind="bar", x = "date", y ="views", color= colors)  
plt.title("Total Pageviews")
```

```
Out[47]: Text(0.5, 1.0, 'Total Pageviews')
```



```
In [7]: traffic_duplicated = traffic.duplicated()
```

```
In [8]: traffic_duplicated.value_counts()
```

```
Out[8]: False    122567  
True     103711  
Name: count, dtype: int64
```

```
In [9]: # I removed the duplicates so we can get more accurate reading
```

```
n_traffic = traffic.drop_duplicates()  
n_traffic
```

## Linkfire Project

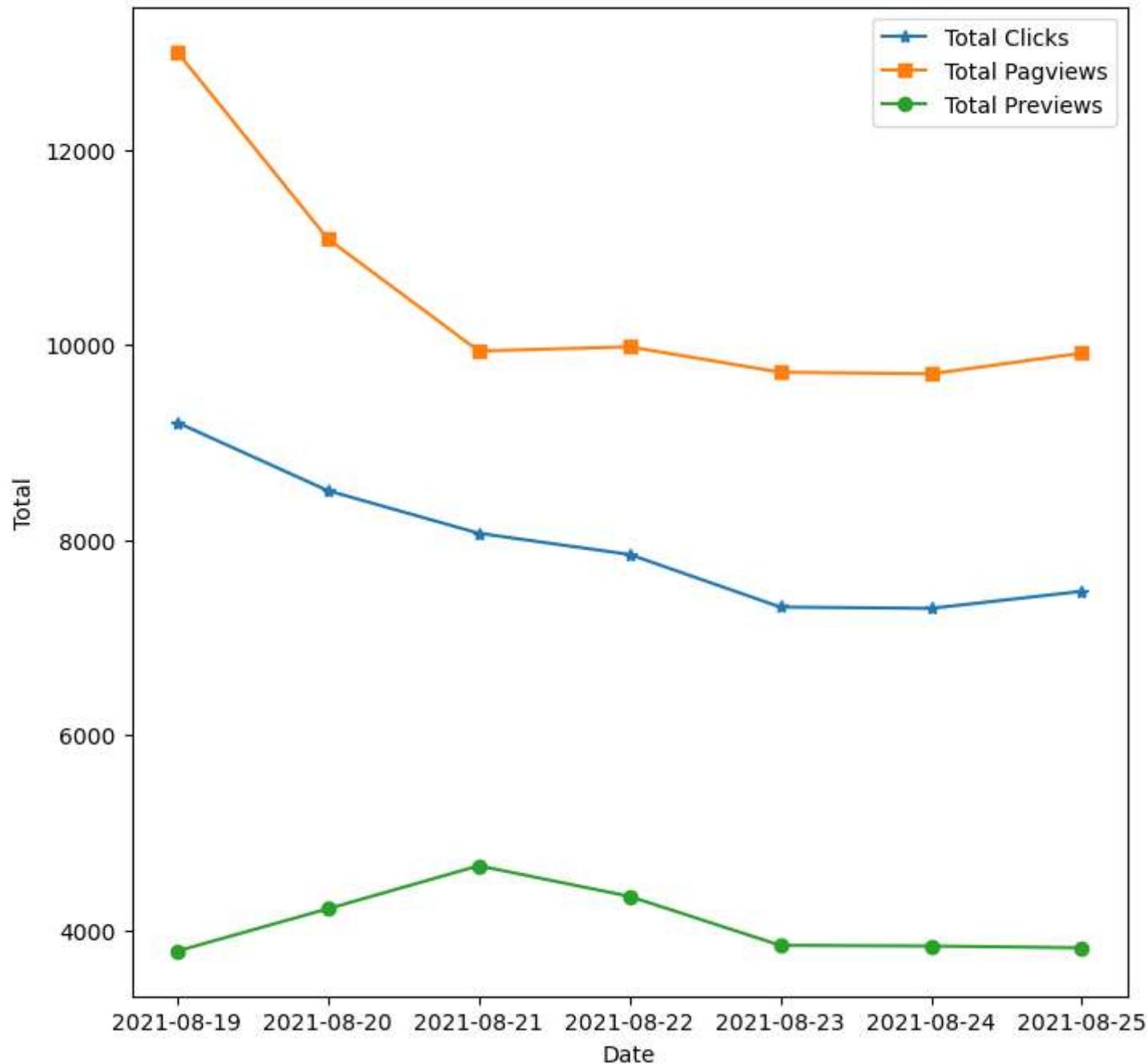
Out[9]:

	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
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
5	click	2021-08-21	United States	Los Angeles	KenTheMan	I'm Perfect	I'm Perfect	US39N2102090	190c7170-4044-4c97-9709-926917155b02
...	...	...	...	...	...	...	...	...	...
226263	pageview	2021-08-24	Turks and Caicos Islands	Providenciales	Kayla Nicole, Taylor Girlz, Flo Milli	Bundles 2 (feat. Flo Milli, Taylor Girlz)	Bundles 2 (feat. Flo Milli, Taylor Girlz)	QMCE32100359	fbf4b935-f961-4b13-b1d8-45ad47093559
226264	pageview	2021-08-24	Australia	Sydney	salem ilese	Mad at Disney	Mad at Disney	QMEZE2076989	61923241-cf99-4fe2-b9ec-6cf356565842
226269	pageview	2021-08-24	United Kingdom	London	Zeina	Talk Luv	Talk Luv	QZAKB2124527	4c402853-c65b-4ea5-ac65-04a9faf5c043
226272	pageview	2021-08-24	United States	Chicago	Pop Smoke, 21 Savage, 42 Dugg	Faith	Bout A Million (feat. 42 Dugg & 21 Savage)	USUM72112386	9ca8ca2d-3507-4287-8f66-a80314dc7169
226274	pageview	2021-08-24	India	Chennai	Miscél	when you left	when you left	QM42K1907890	2fc15e8b-83aa-40ee-9f9d-5d96c6d4356c

122567 rows × 9 columns

```
In [30]: # How many total pageview events did the links in the provided dataset receive in the full period, how many per day?  
total_pageview = n_traffic[n_traffic["event"] == "pageview"].groupby("date")["event"].count()  
  
In [36]: # QUESTION 2 What about the other recorded events?(clicks)  
total_clicks = traffic[traffic["event"] == "click"].groupby("date")["event"].count()  
  
In [43]: #What about the other recorded events?(preview)  
total_preview = traffic[traffic["event"] == "preview"].groupby("date")["event"].count()  
  
In [45]: plt.figure(figsize=(8,8))  
total_clicks.plot(kind="line", label="Total Clicks", marker="*")  
total_pageview.plot(kind="line", label="Total Pagviews", marker="s")  
total_preview.plot(kind="line", label="Total Previews", marker="o")  
plt.xlabel("Date")  
plt.ylabel("Total")  
plt.title("Total Recorded Events")  
plt.legend()  
plt.show()
```

## Total Recorded Events



```
In [13]: # QUESTION 3: Which countries did the clicks come from?  
clicks = traffic[traffic["event"] == "click"]  
country_click = clicks.groupby("country")["event"].nunique()  
country_click.sort_values(ascending=False)
```

```
Out[13]: country  
Albania      1  
Pakistan     1  
Nepal        1  
Netherlands  1  
New Caledonia 1  
..  
Gibraltar    1  
Greece       1  
Greenland    1  
Grenada      1  
Åland         1  
Name: event, Length: 202, dtype: int64
```

```
In [14]: # QUESTION 4: What was the overall click rate (clicks/pageviews)?  
# counting the number of times each link has been clicked on.  
hit_click = traffic[traffic["event"] == "click"].groupby("linkid")["event"].count().reset_index()  
hit_click
```

Out[14]:

	linkid	event
0	00126b32-0c35-507b-981c-02c80d2aa8e7	2
1	004b9724-abca-5481-b6e9-6148a7ca00a5	1
2	0063a982-41cd-5629-96d0-e1c4dd72ea11	2
3	006af6a0-1f0d-4b0c-93bf-756af9071c06	8
4	00759b81-3f04-4a61-b934-f8fb3185f4a0	3
...	...	...
2250	ffd8d5a7-91bc-48e1-a692-c26fca8a8ead	29
2251	fff38ca0-8043-50cd-a5f1-f65ebb7105c5	1
2252	fff84c0e-90a1-59d8-9997-adc909d50e16	1
2253	fffc17a7-f935-5d3e-bd3e-d761fd80d479	1
2254	ffffd0045-29de-522b-b5d8-35786363bf07	1

2255 rows × 2 columns

In [15]: `# renaming the column to 'clicks'  
hit_click = hit_click.rename(columns={"event": "clicks"})`In [16]: `hit_click.head()`

Out[16]:

	linkid	clicks
0	00126b32-0c35-507b-981c-02c80d2aa8e7	2
1	004b9724-abca-5481-b6e9-6148a7ca00a5	1
2	0063a982-41cd-5629-96d0-e1c4dd72ea11	2
3	006af6a0-1f0d-4b0c-93bf-756af9071c06	8
4	00759b81-3f04-4a61-b934-f8fb3185f4a0	3

In [17]: `# counting the number of views each Link has.  
hit_page = traffic[traffic["event"] == "pageview"].groupby("linkid")["event"].count().reset_index()  
hit_page`

Out[17]:

	linkid	event
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-d087bcdd3680	1
4	0034d6cf-3bd8-5ffe-aafc-b3959fc48608	1
...	...	...
3832	fff38ca0-8043-50cd-a5f1-f65ebb7105c5	1
3833	fff4e5f0-4ee5-5fe7-aa30-e870edaf6ed7	2
3834	fff84c0e-90a1-59d8-9997-adc909d50e16	1
3835	ffffc17a7-f935-5d3e-bd3e-d761fd80d479	2
3836	fffd0045-29de-522b-b5d8-35786363bf07	2

3837 rows × 2 columns

In [18]:  
# renaming the column to pageviews  
hit\_page = hit\_page.rename(columns={"event": "pageviews"})In [19]:  
# merging the columns clicks and pageviews creating a new dataframe  
merge = pd.merge(left=hit\_click, right=hit\_page, on="linkid", how="inner")

In [20]: merge

Out[20]:

		linkid	clicks	pageviews
0	00126b32-0c35-507b-981c-02c80d2aa8e7	2	2	
1	004b9724-abca-5481-b6e9-6148a7ca00a5	1	1	
2	0063a982-41cd-5629-96d0-e1c4dd72ea11	2	3	
3	006af6a0-1f0d-4b0c-93bf-756af9071c06	8	36	
4	00759b81-3f04-4a61-b934-f8fb3185f4a0	3	4	
...	...	...	...	...
2248	ffd8d5a7-91bc-48e1-a692-c26fca8a8ead	29	84	
2249	fff38ca0-8043-50cd-a5f1-f65ebb7105c5	1	1	
2250	fff84c0e-90a1-59d8-9997-adc909d50e16	1	1	
2251	fffc17a7-f935-5d3e-bd3e-d761fd80d479	1	2	
2252	ffffd0045-29de-522b-b5d8-35786363bf07	1	2	

2253 rows × 3 columns

```
In [21]: # calculating the clickrate by dividing the clicks and pageviews
merge["click_rate"] = merge["clicks"] / merge["pageviews"]
```

```
In [22]: merge
```

Out[22]:

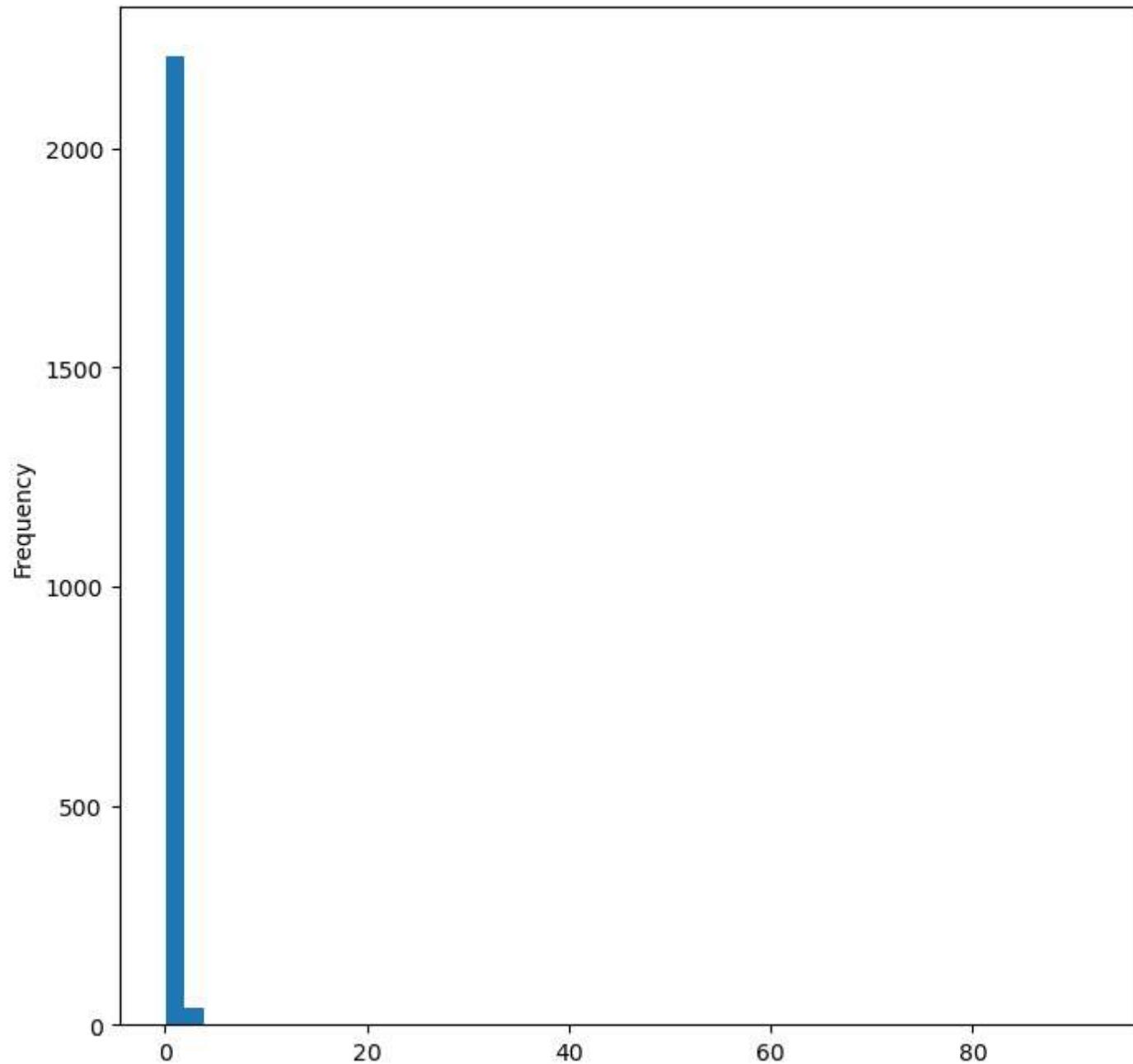
		linkid	clicks	pageviews	click_rate
0	00126b32-0c35-507b-981c-02c80d2aa8e7	2	2	1.000000	
1	004b9724-abca-5481-b6e9-6148a7ca00a5	1	1	1.000000	
2	0063a982-41cd-5629-96d0-e1c4dd72ea11	2	3	0.666667	
3	006af6a0-1f0d-4b0c-93bf-756af9071c06	8	36	0.222222	
4	00759b81-3f04-4a61-b934-f8fb3185f4a0	3	4	0.750000	
...	...	...	...	...	...
2248	ffd8d5a7-91bc-48e1-a692-c26fca8a8ead	29	84	0.345238	
2249	fff38ca0-8043-50cd-a5f1-f65ebb7105c5	1	1	1.000000	
2250	fff84c0e-90a1-59d8-9997-adc909d50e16	1	1	1.000000	
2251	ffffc17a7-f935-5d3e-bd3e-d761fd80d479	1	2	0.500000	
2252	ffffd0045-29de-522b-b5d8-35786363bf07	1	2	0.500000	

2253 rows × 4 columns

In [46]:

```
# QUESTION 5 How does the clickrate distribute across different links?
# I noticed that this histogram chart is not showing everything that I need.
merge["click_rate"].plot(kind="hist", bins =50, figsize=(8,8))
```

Out[46]: <Axes: ylabel='Frequency'>



In [24]: #Below is the much more improved chart

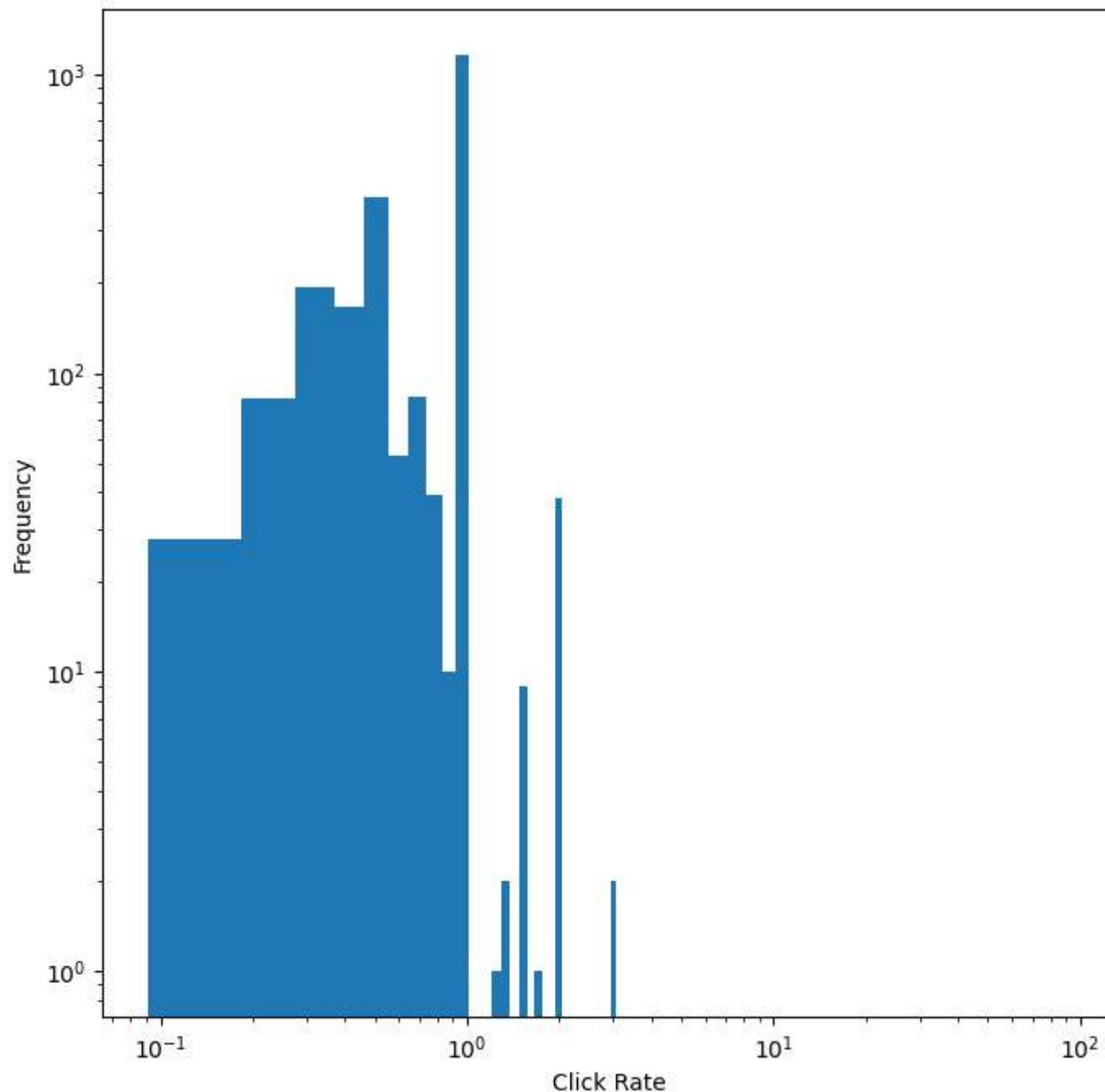
```
In [25]: # I decided to import matplotlib because I believe it easier to do so with this library
import matplotlib.pyplot as plt
```

```
In [26]: #Below is the improved histogram chart I implemented Log so the x and y axis can be on a Logarithmic scale.
fig, ax = plt.subplots(figsize=(8, 8))

# Plot the histogram with a Log scale on both axes
ax.hist(merge["click_rate"], bins=1000, log=True)

# Set the x-axis to a Logarithmic scale
ax.set_xscale('log')
ax.set_xlabel("Click Rate")
ax.set_ylabel("Frequency")
ax.set_title("Clickrate across different links")
plt.show()
```

## Clickrate across different links



In [27]: *# The histogram is highly skewed to the left meaning that the pages were clicked and viewed only once that is it.*