

In [1]: `import pandas as pd`

```
file_path = 'homepage_actions.csv'
df = pd.read_csv(file_path)
print(df.head())
```

	timestamp	id	group	action
0	2016-09-24 17:42:27.839496	804196	experiment	view
1	2016-09-24 19:19:03.542569	434745	experiment	view
2	2016-09-24 19:36:00.944135	507599	experiment	view
3	2016-09-24 19:59:02.646620	671993	control	view
4	2016-09-24 20:26:14.466886	536734	experiment	view

In [2]: `duplicate_ids = df[df.duplicated(subset='id')]`
`print("Duplicate IDs:")`
`print(duplicate_ids)`

Duplicate IDs:

	timestamp	id	group	action
8	2016-09-24 20:58:01.948663	349125	experiment	click
12	2016-09-24 21:06:27.553057	601714	experiment	click
15	2016-09-24 21:30:02.739756	487634	experiment	click
17	2016-09-24 23:01:12.108316	468601	experiment	click
21	2016-09-25 00:01:47.933853	555973	experiment	click
...
8162	2017-01-17 23:20:35.483601	451198	control	click
8164	2017-01-17 23:47:58.209653	252195	control	click
8167	2017-01-18 00:56:24.554729	344770	experiment	click
8180	2017-01-18 08:54:56.879682	615849	experiment	click
8182	2017-01-18 09:09:17.363917	795585	control	click

[1860 rows x 4 columns]

In [3]: `viewers_clicked = df[df['action'] == 'click']['id'].nunique()`
`print("Number of viewers who also clicked:", viewers_clicked)`

Number of viewers who also clicked: 1860

In [9]: `click_ids = df[df['action'] == 'click']['id']`
`view_ids = df[df['action'] == 'view']['id'].unique()`

`click_without_view = df[df['id'].isin(click_ids) & ~df['id'].isin(view_ids)]`
`print("Clicks without corresponding views:")`
`print(click_without_view)`

Clicks without corresponding views:

Empty DataFrame

Columns: [timestamp, id, group, action]

Index: []

```
In [6]: control_ids = df[df['group'] == 'control']['id']  
        experiment_ids = df[df['group'] == 'experiment']['id']  
        overlap_ids = control_ids[control_ids.isin(experiment_ids)]  
        print("Overlap between control and experiment groups:")  
        print(overlap_ids)
```

Overlap between control and experiment groups:
Series([], Name: id, dtype: int64)

Options include removing overlapping data, stratified sampling, or using statistical methods to control for the overlap.

In []: