04. Time difference between latest actions.

From the following table of user actions, write a query to return for each user the time elapsed between the last action and the second-to-last action, in ascending order by user ID.

```
!pip install polars
import pandas as pd
import numpy as  np
import polars as pl

data = {'User_id' : [1,1,2,2,3,3,4,1],
        'Actions'    : ['start','cancel','start',
                    'publish','start','cancel',
                    'start','publish'],
        'Action_date'   : ['12-FEB-20',
                         '13-FEB-20',
                         '11-FEB-20',
                         '14-FEB-20',
                         '15-FEB-20',
                         '15-FEB-20',
                         '18-FEB-20',
                         '19-FEB-20']
        }

pandas_users=pd.DataFrame(data)
polars_users=pl.DataFrame(data)
```

```
Requirement already satisfied: polars in /usr/local/lib/python3.11/dist-packages
```

```
pandas_users['Action_date']=(pd.to_datetime(pandas_users['Action_date'],
                                            format="%d-%b-%y"
                                )
)
print(f'----PANDAS----\nusers table:\n{pandas_users}')
```

```
----PANDAS----
users table:
   User_id  Actions Action_date
0        1    start  2020-02-12
1        1   cancel  2020-02-13
2        2    start  2020-02-11
3        2  publish  2020-02-14
4        3    start  2020-02-15
5        3   cancel  2020-02-15
6        4    start  2020-02-18
7        1  publish  2020-02-19
```

```
pandas_df1=(pandas_users.sort_values(by=['User_id','Action_date'],
                                     ascending=[True,False]
                         )
                        .assign(Prev_action=lambda x:
```

```
                                                x.groupby('User_id')['Action_date'].shift(-1)
                          )[['User_id','Action_date','Prev_action']]
                         .groupby('User_id',as_index=False).first()
 )
 print(f'What are the two last actions:\n{pandas_df1}')
```

```
What are the two last actions:
   User_id Action_date Prev_action
0        1  2020-02-19  2020-02-13
1        2  2020-02-14  2020-02-11
2        3  2020-02-15  2020-02-15
3        4  2020-02-18         NaT
```

```
ndas_duration=(pandas_users.sort_values(by=['User_id','Action_date'],
                                        ascending=[True,False]
                            )
                           .assign(Prev_action=lambda x:
                                               x.groupby('User_id')['Action_date']
                                                .shift(-1)
                            )[['User_id','Action_date','Prev_action']]
                           .groupby('User_id',as_index=False).first()
                           .assign(Elapsed_time=lambda x:
                                                (x['Action_date']-x['Prev_action']).dt.days
                            )[['User_id','Elapsed_time']]

int(f'Returning time elapsed using Pandas:')
ndas_duration
```

```
Returning time elapsed using Pandas:
```

	User_id	Elapsed_time
0	1	6.0
1	2	3.0
2	3	0.0
3	4	NaN

```
 polars_users=(polars_users.with_columns(pl.col('Action_date')
                                         .str
                                         .strptime(pl.Date,
                                                   format="%d-%b-%y"
                                          )
                            )
 )
 print(f'----POLARS----\nusers table:\n{polars_users}')
```

```
----POLARS----
users table:
shape: (8, 3)
┌─────────┬─────────┬─────────────┐
│ User_id ┆ Actions ┆ Action_date │
│ ---     ┆ ---     ┆ ---         │
│ i64     ┆ str     ┆ date        │
```

```
| 1       | start   | 2020-02-12 |
| 1       | cancel  | 2020-02-13 |
| 2       | start   | 2020-02-11 |
| 2       | publish | 2020-02-14 |
| 3       | start   | 2020-02-15 |
| 3       | cancel  | 2020-02-15 |
| 4       | start   | 2020-02-18 |
| 1       | publish | 2020-02-19 |
```

```
polars_df1=(polars_users.sort(by=['User_id','Action_date']
                              ,descending=[False,True]
                         )
                        .with_columns(Prev_action=pl.col('Action_date')
                                                   .shift(-1)
                                                   .over(partition_by='User_id')
                         )
                        .group_by('User_id')
                        .first()
                        .select(pl.col('User_id')
                                ,pl.col('Action_date')
                                ,pl.col('Prev_action')
                         )
)
print(f'What are the two last actions:\n{polars_df1}')
```

```
What are the two last actions:
shape: (4, 3)
| User_id | Action_date | Prev_action |
| ---     | ---         | ---         |
| i64     | date        | date        |
| 1       | 2020-02-19  | 2020-02-13  |
| 2       | 2020-02-14  | 2020-02-11  |
| 3       | 2020-02-15  | 2020-02-15  |
| 4       | 2020-02-18  | null        |
```

```
polars_durations=(polars_users.sort(by=['User_id','Action_date']
                                    ,descending=[False,True]
                               )
                              .with_columns(Elapsed_time=pl.col('Action_date')
                                                         .diff(-1)
                                                         .dt
                                                         .total_days()
                                                         .over(partition_by='User_id')
                               )
                              .group_by('User_id')
                              .first()
                              .select(pl.col('User_id'),
                                      pl.col('Elapsed_time')
                               )
)
```

```
print(f'Returning time elapsed with Polars:')
polars_durations
```

```
Returning time elapsed with Polars:
shape: (4, 2)
```

User_id	Elapsed_time
i64	i64
1	6
2	3
3	0
4	null