

✓ 10. Project Aggregation

The projects table contains three columns: task_id, start_date, and end_date. The difference between end_date and start_date is 1 day for each row in the table. If task end dates are consecutive they are part of the same project. Projects do not overlap. Write a query to return the start and end dates of each project, and the number of days it took to complete. Order by ascending project duration, and ascending start date in the case of a tie.

```
import pandas as pd
import numpy as np

data = {'task_id'      : [1,2,3,4,5,6,7],
        'start_date'   : ['01-OCT-20',
                           '02-OCT-20',
                           '03-OCT-20',
                           '13-OCT-20',
                           '14-OCT-20',
                           '28-OCT-20',
                           '30-OCT-20'],
        'end_date'     : ['02-OCT-20',
                           '03-OCT-20',
                           '04-OCT-20',
                           '14-OCT-20',
                           '15-OCT-20',
                           '29-OCT-20',
                           '31-OCT-20']}


projects = pd.DataFrame(data)



projects['start_date']=(pd.to_datetime(projects['start_date']
                                     ,format="%d-%b-%y"
                                     )
)
projects['end_date']=(pd.to_datetime(projects['end_date']
                                     ,format="%d-%b-%y"
                                     )
)
print(projects)
```

```
➡ task_id start_date end_date
0      1 2020-10-01 2020-10-02
1      2 2020-10-02 2020-10-03
2      3 2020-10-03 2020-10-04
3      4 2020-10-13 2020-10-14
4      5 2020-10-14 2020-10-15
5      6 2020-10-28 2020-10-29
6      7 2020-10-30 2020-10-31
```



```
        )
    )
    .assign(project_duration=lambda x:
            (x['project_end']-x['project_start']).d
    )
)
timeline_projects.sort_values(['project_duration','project_start'])
```



	project_id	project_start	project_end	project_duration	
2	3	2020-10-28	2020-10-29	1	
3	4	2020-10-30	2020-10-31	1	
1	2	2020-10-13	2020-10-15	2	
0	1	2020-10-01	2020-10-04	3	