

# Task

May 16, 2022

```
[1]: import pandas as pd
import numpy as np
import warnings
warnings.filterwarnings('ignore')
```

## 0.0.1 1. Read csv file

```
[2]: data = pd.read_csv("/home/user/INTERVIEWS/Python/input.csv")
data
```

```
[2]:      Id      dob first_name \
0    10.0  21-08-1992      liam
1     6.0   1-8-1999    olivia
2     2.0  22-03-1976     noah
3     7.0  10-6-2002    oliver
4     9.0  11-11-2020    Elijah
5     1.0  12-12-1996      NaN
6     4.0  30-07-2021    William
7     6.0   1-8-1999    olivia
8     NaN      NaN      NaN
9     NaN      NaN      NaN
10    NaN      NaN      NaN
11    NaN      NaN      NaN
12    NaN      NaN      NaN
13    NaN      NaN      NaN
14    NaN      NaN      NaN
15    NaN      NaN      NaN
16    NaN      NaN      NaN
17    NaN      NaN      NaN
18    4.0    William      Mia
```

```
      url      last_name  Unnamed: 5
0  https://im.rediff.com/money/2020/nov/15car-saf...      Emma      NaN
1  https://i0.wp.com/gomechanic.in/blog/wp-conten...  charlotte      NaN
2  https://i0.wp.com/gomechanic.in/blog/wp-conten...    Amelia      NaN
3      NaN      NaN      NaN
```

4		NaN	sophia	NaN
5	https://1000logos.net/wp-content/uploads/2016/...		isabella	NaN
6		NaN	mia	NaN
7	https://i0.wp.com/gomechanic.in/blog/wp-conten...		charlotte	NaN
8		NaN	NaN	NaN
9		NaN	NaN	NaN
10		NaN	NaN	NaN
11		NaN	NaN	NaN
12		NaN	NaN	NaN
13		NaN	NaN	NaN
14		NaN	NaN	NaN
15		NaN	NaN	NaN
16		NaN	NaN	NaN
17		NaN	NaN	NaN
18		William	Mia	30-07-2021

**0.0.2 2. Print total number of rows CSV files in th**

**0.0.3 Answer: 19**

```
[3]: data.shape[0]
```

```
[3]: 19
```

**0.0.4 3. No Duplicate records need to be processed**

```
[4]: data=data.drop_duplicates()
```

**0.0.5 4. Print how many rows exist excluding the duplicate records.**

**0.0.6 Answer: 10**

```
[5]: data.shape[0]
```

```
[5]: 10
```

**0.0.7 5. The first\_name and last\_name should start with a capital letter.**

```
[6]: data['first_name']=data['first_name'].str.title()
data['last_name']=data['last_name'].str.title()

data
```

```
[6]:      Id      dob first_name \
0    10.0  21-08-1992      Liam
1     6.0   1-8-1999     Olivia
2     2.0  22-03-1976      Noah
3     7.0  10-6-2002     Oliver
4     9.0  11-11-2020     Elijah
5     1.0  12-12-1996      NaN
6     4.0  30-07-2021     William
7     6.0   1-8-1999     Olivia
8     NaN      NaN      NaN
18    4.0      William      Mia
```

```

                                url      last_name  Unnamed: 5
0  https://im.rediff.com/money/2020/nov/15car-saf...      Emma      NaN
1  https://i0.wp.com/gomechanic.in/blog/wp-conten...  Charlotte      NaN
2  https://i0.wp.com/gomechanic.in/blog/wp-conten...    Amelia      NaN
3                                NaN      NaN      NaN
4                                NaN      Sophia      NaN
5  https://1000logos.net/wp-content/uploads/2016/...  Isabella      NaN
6                                NaN      Mia      NaN
7  https://i0.wp.com/gomechanic.in/blog/wp-conten...  Charlotte      NaN
8                                NaN      NaN      NaN
18                                William Mia  30-07-2021      0.0
```

**0.0.8 6.** The `first_name` and `last_name` columns in the CSV file have to be concatenated to form the full name, which must be fully capitalized.

```
[7]: data['full_name']=data['first_name'] +" "+ data['last_name']
      data['full_name']
```

```
[7]: 0      Liam Emma
1     Olivia Charlotte
2      Noah Amelia
3           NaN
4     Elijah Sophia
5           NaN
6     William Mia
7     Olivia Charlotte
8           NaN
18    Mia 30-07-2021
Name: full_name, dtype: object
```

### 0.0.9 9. Sort the row in ascending order according to the column “Id”.

```
[8]: data.sort_values("Id")
data
```

```
[8]:      Id      dob first_name \
0  10.0  21-08-1992      Liam
1   6.0   1-8-1999    Olivia
2   2.0  22-03-1976     Noah
3   7.0  10-6-2002    Oliver
4   9.0  11-11-2020    Elijah
5   1.0  12-12-1996      NaN
6   4.0  30-07-2021    William
7   6.0   1-8-1999    Olivia
8   NaN      NaN      NaN
18  4.0      William      Mia

      url      last_name  Unnamed: 5 \
0  https://im.rediff.com/money/2020/nov/15car-saf...      Emma      NaN
1  https://i0.wp.com/gomechanic.in/blog/wp-conten...  Charlotte      NaN
2  https://i0.wp.com/gomechanic.in/blog/wp-conten...    Amelia      NaN
3                        NaN      NaN      NaN
4                        NaN    Sophia      NaN
5  https://1000logos.net/wp-content/uploads/2016/...  Isabella      NaN
6                        NaN      Mia      NaN
7  https://i0.wp.com/gomechanic.in/blog/wp-conten...  Charlotte      NaN
8                        NaN      NaN      NaN
18                        William Mia  30-07-2021      0.0

      full_name
0      Liam Emma
1  Olivia Charlotte
2      Noah Amelia
3           NaN
4    Elijah Sophia
5           NaN
6    William Mia
7  Olivia Charlotte
8           NaN
18    Mia 30-07-2021
```

### 0.0.10 10. Then, save all of the data to a CSV file named “assignment1.csv”.

```
[9]: data.to_csv('/home/user/INTERVIEWS/assignment1.csv', index=False)
```