

```
In [2]: #Importation et affichage
pwd

Out[2]: '/Users/Yaha/Desktop/CAPSTONE 2/OUTPUT'

In [3]: import pandas as pd

In [4]: df_employees = pd.read_csv('/Users/Yaha/Desktop/CAPSTONE 2/Dataset/Employees/employees_table.csv', sep = ';')

In [5]: df_employees

Out[5]:
```

	employee_id	employee_first_name	employee_last_name	gender	date_of_birth	job	joining_date	day_tour_trained	night_tour_trained	versailles_train
0	34-0035381	Tom	Burden	m	24/06/1987	Office Manager	24/05/2017	y	y	
1	50-2207707	Alice	Richards	f	15/05/1981	Office Manager	17/06/2017	y	y	
2	12-3003928	Shaun	McLaughlin	m	01/02/1990	Senior Guide	22/07/2019	y	y	
3	13-0214740	Tim	Chang	m	24/01/1995	Senior Guide	25/09/2019	y	y	
4	82-4746874	Eric	Winterbourne	m	01/05/1996	Senior Guide	04/02/2020	y	y	
5	22-0191141	Helen	Smith	f	13/12/1994	Senior Guide	12/02/2020	y	y	
6	53-8952948	Tarik	Patel	m	08/01/2000	Guide	01/04/2020	y	y	
7	75-0269501	Luke	Thomas	m	11/07/1999	Guide	30/08/2020	y	y	
8	61-9022510	Annie	Spayne	f	30/07/1995	Guide	02/10/2020	y	y	
9	11-7101553	Michelle	Turner	f	17/02/1987	Guide	21/10/2021	y	y	
10	60-3815509	Emily	Ngo	f	15/08/1997	Guide	29/11/2021	y	y	
11	64-8956770	Laura	Tobin	f	30/01/1997	Guide	05/01/2022	y	y	
12	55-1433827	Ben	Jonsen	m	02/12/1981	Guide	18/02/2022	y	y	
13	11-9351390	Peter	Cornford	m	06/02/2002	Guide	01/04/2021	y	y	
14	60-8608095	Katherine	Chiswell	f	08/07/1996	Guide	13/05/2022	y	y	
15	52-4576356	Nolan	Pendergast	m	22/11/1994	Guide	12/05/2022	y	y	
16	29-9644211	Andrew	Nguyen	m	07/11/1990	Guide	16/05/2022	y	y	
17	39-6865023	Sally	Kittman	f	20/11/1990	Guide	16/05/2022	y	y	
18	51-5531588	Frances	Segrave	f	17/05/1989	Guide	31/05/2022	y	y	
19	58-1868935	Nicola	Georgi	m	08/05/1990	Guide	31/05/2022	y	y	
20	26-3559661	Aline	Robert	f	17/09/1991	Guide	01/06/2022	y	y	
21	21-6550197	Liliane	Dubourblanc	f	30/11/1999	Guide	01/06/2022	y	y	
22	01-3339282	Henry	Cumberdale	m	18/09/1991	Guide	01/06/2022	y	y	

```
In [7]: #Informations sur la dataframe
df_employees.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 23 entries, 0 to 22
Data columns (total 11 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   employee_id                          23 non-null    object
1   employee_first_name                  23 non-null    object
2   employee_last_name                   23 non-null    object
3   gender                               23 non-null    object
4   date_of_birth                        23 non-null    object
5   job                                  23 non-null    object
6   joining_date                         23 non-null    object
7   day_tour_trained                     23 non-null    object
8   night_tour_trained                  23 non-null    object
9   versailles_trained                  23 non-null    object
10  seniority (has already worked here)  23 non-null    object
dtypes: object(11)
memory usage: 2.1+ KB
```

```
In [8]: #Nombre de doublons
print(df_employees.duplicated().sum())

0
```

```
In [9]: #Valeurs manquantes
df_employees.isnull().sum()

Out[9]:
employee_id                0
employee_first_name        0
employee_last_name         0
gender                     0
date_of_birth              0
job                         0
joining_date               0
day_tour_trained           0
night_tour_trained         0
versailles_trained         0
seniority (has already worked here)  0
dtype: int64
```

```
In [11]: #Assemblage 'first_name' et 'last_name' sur une nouvelle colonne appelée 'name'
df_employees['name'] = df_employees['employee_first_name'].str.cat(df_employees['employee_last_name'],sep=" ")
```

```
In [12]: df_employees

Out[12]:
```

	employee_id	employee_first_name	employee_last_name	gender	date_of_birth	job	joining_date	day_tour_trained	night_tour_trained	versailles_train
0	34-0035381	Tom	Burden	m	24/06/1987	Office Manager	24/05/2017	y	y	
1	50-2207707	Alice	Richards	f	15/05/1981	Office Manager	17/06/2017	y	y	
2	12-3003928	Shaun	McLaughlin	m	01/02/1990	Senior Guide	22/07/2019	y	y	
3	13-0214740	Tim	Chang	m	24/01/1995	Senior Guide	25/09/2019	y	y	
4	82-4746874	Eric	Winterbourne	m	01/05/1996	Senior Guide	04/02/2020	y	y	
5	22-0191141	Helen	Smith	f	13/12/1994	Senior Guide	12/02/2020	y	y	
6	53-8952948	Tarik	Patel	m	08/01/2000	Guide	01/04/2020	y	y	
7	75-0269501	Luke	Thomas	m	11/07/1999	Guide	30/08/2020	y	y	
8	61-9022510	Annie	Spayne	f	30/07/1995	Guide	02/10/2020	y	y	
9	11-7101553	Michelle	Turner	f	17/02/1987	Guide	21/10/2021	y	y	
10	60-3815509	Emily	Ngo	f	15/08/1997	Guide	29/11/2021	y	y	
11	64-8956770	Laura	Tobin	f	30/01/1997	Guide	05/01/2022	y	y	
12	55-1433827	Ben	Jonsen	m	02/12/1981	Guide	18/02/2022	y	y	
13	11-9351390	Peter	Cornford	m	06/02/2002	Guide	01/04/2021	y	y	
14	60-8608095	Katherine	Chiswell	f	08/07/1996	Guide	13/05/2022	y	y	
15	52-4576356	Nolan	Pendergast	m	22/11/1994	Guide	12/05/2022	y	y	
16	29-9644211	Andrew	Nguyen	m	07/11/1990	Guide	16/05/2022	y	y	
17	39-6865023	Sally	Kittman	f	20/11/1990	Guide	16/05/2022	y	y	
18	51-5531588	Frances	Segrave	f	17/05/1989	Guide	31/05/2022	y	y	
19	58-1868935	Nicola	Georgi	m	08/05/1990	Guide	31/05/2022	y	y	
20	26-3559661	Aline	Robert	f	17/09/1991	Guide	01/06/2022	y	y	
21	21-6550197	Liliane	Dubourblanc	f	30/11/1999	Guide	01/06/2022	y	y	
22	01-3339282	Henry	Cumberdale	m	18/09/1991	Guide	01/06/2022	y	y	

```
In [13]: #Suppression des colonnes 'employee_first_name' et 'employee_last_name'
```

```
In [14]: df_employees=df_employees.drop(['employee_first_name'], axis=1)
```

```
In [15]: df_employees=df_employees.drop(['employee_last_name'], axis=1)
```

```
In [16]: df_employees

Out[16]:
```

	employee_id	gender	date_of_birth	job	joining_date	day_tour_trained	night_tour_trained	versailles_trained	seniority (has already worked here)	name
0	34-0035381	m	24/06/1987	Office Manager	24/05/2017	y	y	y	y	Tom Burden
1	50-2207707	f	15/05/1981	Office Manager	17/06/2017	y	y	y	y	Alice Richards
2	12-3003928	m	01/02/1990	Senior Guide	22/07/2019	y	y	y	y	Shaun McLaughlin
3	13-0214740	m	24/01/1995	Senior Guide	25/09/2019	y	y	y	y	Tim Chang
4	82-4746874	m	01/05/1996	Senior Guide	04/02/2020	y	y	y	y	Eric Winterbourne
5	22-0191141	f	13/12/1994	Senior Guide	12/02/2020	y	y	y	y	Helen Smith
6	53-8952948	m	08/01/2000	Guide	01/04/2020	y	y	y	y	Tarik Patel
7	75-0269501	m	11/07/1999	Guide	30/08/2020	y	y	y	y	Luke Thomas
8	61-9022510	f	30/07/1995	Guide	02/10/2020	y	y	y	n	Annie Spayne
9	11-7101553	f	17/02/1987	Guide	21/10/2021	y	y	y	n	Michelle Turner
10	60-3815509	f	15/08/1997	Guide	29/11/2021	y	y	y	n	Emily Ngo
11	64-8956770	f	30/01/1997	Guide	05/01/2022	y	y	n	n	Laura Tobin
12	55-1433827	m	02/12/1981	Guide	18/02/2022	y	y	n	n	Ben Jonsen
13	11-9351390	m	06/02/2002	Guide	01/04/2021	y	y	n	n	Peter Cornford
14	60-8608095	f	08/07/1996	Guide	13/05/2022	y	y	n	n	Katherine Chiswell
15	52-4576356	m	22/11/1994	Guide	12/05/2022	y	y	n	n	Nolan Pendergast
16	29-9644211	m	07/11/1990	Guide	16/05/2022	y	y	n	n	Andrew Nguyen
17	39-6865023	f	20/11/1990	Guide	16/05/2022	y	y	n	n	Sally Kittman
18	51-5531588	f	17/05/1989	Guide	31/05/2022	y	y	n	n	Frances Segrave
19	58-1868935	m	08/05/1990	Guide	31/05/2022	y	y	n	n	Nicola Georgi
20	26-3559661	f	17/09/1991	Guide	01/06/2022	y	y	n	n	Aline Robert
21	21-6550197	f	30/11/1999	Guide	01/06/2022	y	y	n	n	Liliane Dubourblanc
22	01-3339282	m	18/09/1991	Guide	01/06/2022	y	y	n	n	Henry Cumberdale

```
In [18]: #Uniformisation des formats 'date_of_birth' et 'joining_date'
df_employees['date_of_birth'] = pd.to_datetime(df_employees['date_of_birth'])
df_employees['joining_date'] = pd.to_datetime(df_employees['joining_date'])

/var/folders/7c/62r88x.x4s8ktt18cg1cph0000gp/T/ipykernel_4421/3653773561.py:1: UserWarning: Parsing '24/06/1987' in
DD/MM/YYYY format. Provide format or specify infer_datetime_format=True for consistent parsing.
df_employees['date_of_birth'] = pd.to_datetime(df_employees['date_of_birth'])
/var/folders/7c/62r88x.x4s8ktt18cg1cph0000gp/T/ipykernel_4421/3653773561.py:1: UserWarning: Parsing '15/05/1981' in
DD/MM/YYYY format. Provide format or specify infer_datetime_format=True for consistent parsing.
df_employees['date_of_birth'] = pd.to_datetime(df_employees['date_of_birth'])
/var/folders/7c/62r88x.x4s8ktt18cg1cph0000gp/T/ipykernel_4421/3653773561.py:1: UserWarning: Parsing '24/01/1995' in
DD/MM/YYYY format. Provide format or specify infer_datetime_format=True for consistent parsing.
df_employees['date_of_birth'] = pd.to_datetime(df_employees['date_of_birth'])
/var/folders/7c/62r88x.x4s8ktt18cg1cph0000gp/T/ipykernel_4421/3653773561.py:1: UserWarning: Parsing '13/12/1994' in
DD/MM/YYYY format. Provide format or specify infer_datetime_format=True for consistent parsing.
df_employees['date_of_birth'] = pd.to_datetime(df_employees['date_of_birth'])
/var/folders/7c/62r88x.x4s8ktt18cg1cph0000gp/T/ipykernel_4421/3653773561.py:1: UserWarning: Parsing '30/07/1995' in
DD/MM/YYYY format. Provide format or specify infer_datetime_format=True for consistent parsing.
df_employees['date_of_birth'] = pd.to_datetime(df_employees['date_of_birth'])
/var/folders/7c/62r88x.x4s8ktt18cg1cph0000gp/T/ipykernel_4421/3653773561.py:1: UserWarning: Parsing '17/02/1987' in
DD/MM/YYYY format. Provide format or specify infer_datetime_format=True for consistent parsing.
df_employees['date_of_birth'] = pd.to_datetime(df_employees['date_of_birth'])
/var/folders/7c/62r88x.x4s8ktt18cg1cph0000gp/T/ipykernel_4421/3653773561.py:1: UserWarning: Parsing '15/08/1997' in
DD/MM/YYYY format. Provide format or specify infer_datetime_format=True for consistent parsing.
df_employees['date_of_birth'] = pd.to_datetime(df_employees['date_of_birth'])
/var/folders/7c/62r88x.x4s8ktt18cg1cph0000gp/T/ipykernel_4421/3653773561.py:1: UserWarning: Parsing '30/01/1997' in
DD/MM/YYYY format. Provide format or specify infer_datetime_format=True for consistent parsing.
df_employees['date_of_birth'] = pd.to_datetime(df_employees['date_of_birth'])
/var/folders/7c/62r88x.x4s8ktt18cg1cph0000gp/T/ipykernel_4421/3653773561.py:1: UserWarning: Parsing '22/11/1994' in
DD/MM/YYYY format. Provide format or specify infer_datetime_format=True for consistent parsing.
df_employees['date_of_birth'] = pd.to_datetime(df_employees['date_of_birth'])
/var/folders/7c/62r88x.x4s8ktt18cg1cph0000gp/T/ipykernel_4421/3653773561.py:1: UserWarning: Parsing '20/11/1990' in
DD/MM/YYYY format. Provide format or specify infer_datetime_format=True for consistent parsing.
df_employees['date_of_birth'] = pd.to_datetime(df_employees['date_of_birth'])
/var/folders/7c/62r88x.x4s8ktt18cg1cph0000gp/T/ipykernel_4421/3653773561.py:1: UserWarning: Parsing '17/05/1989' in
DD/MM/YYYY format. Provide format or specify infer_datetime_format=True for consistent parsing.
df_employees['date_of_birth'] = pd.to_datetime(df_employees['date_of_birth'])
/var/folders/7c/62r88x.x4s8ktt18cg1cph0000gp/T/ipykernel_4421/3653773561.py:1: UserWarning: Parsing '17/09/1991' in
DD/MM/YYYY format. Provide format or specify infer_datetime_format=True for consistent parsing.
df_employees['date_of_birth'] = pd.to_datetime(df_employees['date_of_birth'])
/var/folders/7c/62r88x.x4s8ktt18cg1cph0000gp/T/ipykernel_4421/3653773561.py:1: UserWarning: Parsing '30/11/1999' in
DD/MM/YYYY format. Provide format or specify infer_datetime_format=True for consistent parsing.
df_employees['date_of_birth'] = pd.to_datetime(df_employees['date_of_birth'])
/var/folders/7c/62r88x.x4s8ktt18cg1cph0000gp/T/ipykernel_4421/3653773561.py:1: UserWarning: Parsing '18/09/1991' in
DD/MM/YYYY format. Provide format or specify infer_datetime_format=True for consistent parsing.
df_employees['date_of_birth'] = pd.to_datetime(df_employees['date_of_birth'])
/var/folders/7c/62r88x.x4s8ktt18cg1cph0000gp/T/ipykernel_4421/3653773561.py:2: UserWarning: Parsing '24/05/2017' in
DD/MM/YYYY format. Provide format or specify infer_datetime_format=True for consistent parsing.
df_employees['joining_date'] = pd.to_datetime(df_employees['joining_date'])
/var/folders/7c/62r88x.x4s8ktt18cg1cph0000gp/T/ipykernel_4421/3653773561.py:2: UserWarning: Parsing '17/06/2017' in
DD/MM/YYYY format. Provide format or specify infer_datetime_format=True for consistent parsing.
df_employees['joining_date'] = pd.to_datetime(df_employees['joining_date'])
/var/folders/7c/62r88x.x4s8ktt18cg1cph0000gp/T/ipykernel_4421/3653773561.py:2: UserWarning: Parsing '22/07/2019' in
DD/MM/YYYY format. Provide format or specify infer_datetime_format=True for consistent parsing.
df_employees['joining_date'] = pd.to_datetime(df_employees['joining_date'])
/var/folders/7c/62r88x.x4s8ktt18cg1cph0000gp/T/ipykernel_4421/3653773561.py:2: UserWarning: Parsing '30/08/2020' in
DD/MM/YYYY format. Provide format or specify infer_datetime_format=True for consistent parsing.
df_employees['joining_date'] = pd.to_datetime(df_employees['joining_date'])
/var/folders/7c/62r88x.x4s8ktt18cg1cph0000gp/T/ipykernel_4421/3653773561.py:2: UserWarning: Parsing '21/10/2021' in
DD/MM/YYYY format. Provide format or specify infer_datetime_format=True for consistent parsing.
df_employees['joining_date'] = pd.to_datetime(df_employees['joining_date'])
/var/folders/7c/62r88x.x4s8ktt18cg1cph0000gp/T/ipykernel_4421/3653773561.py:2: UserWarning: Parsing '29/11/2021' in
DD/MM/YYYY format. Provide format or specify infer_datetime_format=True for consistent parsing.
df_employees['joining_date'] = pd.to_datetime(df_employees['joining_date'])
/var/folders/7c/62r88x.x4s8ktt18cg1cph0000gp/T/ipykernel_4421/3653773561.py:2: UserWarning: Parsing '05/01/2022' in
DD/MM/YYYY format. Provide format or specify infer_datetime_format=True for consistent parsing.
df_employees['joining_date'] = pd.to_datetime(df_employees['joining_date'])
/var/folders/7c/62r88x.x4s8ktt18cg1cph0000gp/T/ipykernel_4421/3653773561.py:2: UserWarning: Parsing '18/02/2022' in
DD/MM/YYYY format. Provide format or specify infer_datetime_format=True for consistent parsing.
df_employees['joining_date'] = pd.to_datetime(df_employees['joining_date'])
/var/folders/7c/62r88x.x4s8ktt18cg1cph0000gp/T/ipykernel_4421/3653773561.py:2: UserWarning: Parsing '13/05/2022' in
DD/MM/YYYY format. Provide format or specify infer_datetime_format=True for consistent parsing.
df_employees['joining_date'] = pd.to_datetime(df_employees['joining_date'])
/var/folders/7c/62r88x.x4s8ktt18cg1cph0000gp/T/ipykernel_4421/3653773561.py:2: UserWarning: Parsing '16/05/2022' in
DD/MM/YYYY format. Provide format or specify infer_datetime_format=True for consistent parsing.
df_employees['joining_date'] = pd.to_datetime(df_employees['joining_date'])
/var/folders/7c/62r88x.x4s8ktt18cg1cph0000gp/T/ipykernel_4421/3653773561.py:2: UserWarning: Parsing '31/05/2022' in
DD/MM/YYYY format. Provide format or specify infer_datetime_format=True for consistent parsing.
df_employees['joining_date'] = pd.to_datetime(df_employees['joining_date'])
```

```
In [19]: #Tri par ordre croissant de la colonne 'joining_date', pour voir les employes les plus anciens aux plus récents
```

```
In [20]: df_employees = df_employees.sort_values(by='joining_date', ascending=True)
```

```
In [21]: df_employees

Out[21]:
```

	employee_id	gender	date_of_birth	job	joining_date	day_tour_trained	night_tour_trained	versailles_trained	seniority (has already worked here)	name
0	34-0035381	m	1987-06-24	Office Manager	2017-05-24	y	y	y	y	Tom Burden
1	50-2207707	f	1981-05-15	Office Manager	2017-06-17	y	y	y	y	Alice Richards
2	12-3003928	m	1990-01-02	Senior Guide	2019-07-22	y	y	y	y	Shaun McLaughlin
3	13-0214740	m	1995-01-24	Senior Guide	2019-09-25	y	y	y	y	Tim Chang
6	53-8952948	m	2000-08-01	Guide	2020-01-04	y	y	y	y	Tarik Patel
8	61-9022510	f	1995-07-30	Guide	2020-02-10	y	y	y	n	Annie Spayne
4	82-4746874	m	1996-01-05	Senior Guide	2020-04-02	y	y	y	y	Eric Winterbourne
7	75-0269501	m	1999-11-07	Guide	2020-08-30	y	y	y	y	Luke Thomas
5	22-0191141	f	1994-12-13	Senior Guide	2020-12-02	y	y	y	y	Helen Smith
13	11-9351390	m	2002-06-02	Guide	2021-01-04	y	y	n	n	Peter Cornford
9	11-7101553	f	1987-02-17	Guide	2021-10-21	y	y	y	n	Michelle Turner
10	60-3815509	f	1997-08-15	Guide	2021-11-29	y	y	y	n	Emily Ngo
20	26-3559661	f	1991-09-17	Guide	2022-01-06	y	y	n	n	Aline Robert
22	01-3339282	m	1991-09-18	Guide	2022-01-06	y	y	n	n	Henry Cumberdale
21	21-6550197	f	1999-11-30	Guide	2022-01-06	y	y	n	n	Liliane Dubourblanc
12	55-1433827	m	1981-02-12	Guide	2022-02-18	y	y	n	n	Ben Jorssen
11	64-9856770	f	1997-01-30	Guide	2022-05-01	y	y	n	n	Laura Tobin
14	60-8080895	f	1986-08-07	Guide	2022-05-13	y	y	n	n	Katherine Chowell
16	29-9644211	m	1990-07-11	Guide	2022-05-16	y	y	n	n	Andrew Nguyen
17	39-6865023	f	1990-11-20	Guide	2022-05-16	y	y	n	n	Sally Kittman
18	51-5531588	f	1989-05-17	Guide	2022-05-31	y	y	n	n	Francis Segrave
19	58-1868935	m	1980-08-05	Guide	2022-05-31	y	y	n	n	Nicola Genova