

```
In [ ]: #Importation et affichage

In [1]: pwd

Out[1]: '/Users/Yaha/Desktop/CAPSTONE 2/OUTPUT'

In [2]: import pandas as pd

In [3]: df_reviews = pd.read_csv('/Users/Yaha/Desktop/CAPSTONE 2/Dataset/Reviews/tour_reviews.csv', sep = ';')

In [4]: df_reviews

Out[4]:
```

	date	rating	comment
0	23/05/2022	3	Bikes a nice way to complement a walking tour ...
1	23/05/2022	5	Great time, good jokes and Tim was great too
2	23/05/2022	4	Loved our green bikes tours! We did the Night ...
3	23/05/2022	4	NaN
4	24/05/2022	3	When I arrived in Paris, I took part in the Ni...
...	...	...	...
151	25/06/2022	1	Really disappointed with Green Bike tours. We ...
152	25/06/2022	5	Impeccable!
153	25/06/2022	2	NaN
154	25/06/2022	1	Another bike tour, selling tickets at highly l...
155	25/06/2022	5	NaN

156 rows × 3 columns

```
In [13]: #Informations sur la dataframe

In [7]: df_reviews.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 156 entries, 0 to 155
Data columns (total 3 columns):
#   Column   Non-Null Count  Dtype
---  ---
0    date     156 non-null    object
1    rating   156 non-null    int64
2    comment  158 non-null    object
dtypes: int64(1), object(2)
memory usage: 3.8+ KB

In [ ]: #Nombre de doublons par colonne

In [15]: print(df_reviews['date'].duplicated().sum())

122

In [16]: print(df_reviews['rating'].duplicated().sum())

151

In [17]: print(df_reviews['comment'].duplicated().sum())

18

In [ ]: #On ne supprime pas les doublons parce qu'ils concernent les dates et les notes attribuées par les clients.

In [8]: #Détection des valeurs manquantes
df_reviews.isnull().sum()

Out[8]:
date      0
rating    0
comment   18
dtype: int64

In [103]: df_reviews[df_reviews['comment'].isnull()]

Out[103]:
```

	date	rating	comment
3	23/05/2022	4	NaN
7	25/05/2022	1	NaN
29	29/05/2022	4	NaN
42	31/05/2022	4	NaN
51	02/06/2022	5	NaN
53	02/06/2022	4	NaN
55	02/06/2022	4	NaN
64	06/06/2022	4	NaN
95	13/06/2022	4	NaN
108	17/06/2022	3	NaN
111	17/06/2022	4	NaN
115	19/06/2022	5	NaN
122	20/06/2022	4	NaN
133	22/06/2022	5	NaN
135	22/06/2022	4	NaN
138	23/06/2022	2	NaN
153	25/06/2022	2	NaN
155	25/06/2022	5	NaN

```
In [ ]: #On ne supprime pas les valeurs manquantes dans les commentaires sinon ça supprimerait aussi les notes des clients.

In [104]: #Uniformisation du format date

In [105]: df_reviews['date'] = pd.to_datetime(df_reviews['date'])

/Users/hakim/opt/anaconda3/lib/python3.9/site-packages/pandas/core/tools/datetimes.py:1047: UserWarning: Parsing '23/05/2022' in DD/MM/YYYY format. Provide format or specify infer_datetime_format=True for consistent parsing.
  cache_array = _maybe_cache(arg, format, cache, convert_listlike)
/Users/hakim/opt/anaconda3/lib/python3.9/site-packages/pandas/core/tools/datetimes.py:1047: UserWarning: Parsing '24/05/2022' in DD/MM/YYYY format. Provide format or specify infer_datetime_format=True for consistent parsing.
  cache_array = _maybe_cache(arg, format, cache, convert_listlike)
/Users/hakim/opt/anaconda3/lib/python3.9/site-packages/pandas/core/tools/datetimes.py:1047: UserWarning: Parsing '25/05/2022' in DD/MM/YYYY format. Provide format or specify infer_datetime_format=True for consistent parsing.
  cache_array = _maybe_cache(arg, format, cache, convert_listlike)
/Users/hakim/opt/anaconda3/lib/python3.9/site-packages/pandas/core/tools/datetimes.py:1047: UserWarning: Parsing '26/05/2022' in DD/MM/YYYY format. Provide format or specify infer_datetime_format=True for consistent parsing.
  cache_array = _maybe_cache(arg, format, cache, convert_listlike)
/Users/hakim/opt/anaconda3/lib/python3.9/site-packages/pandas/core/tools/datetimes.py:1047: UserWarning: Parsing '27/05/2022' in DD/MM/YYYY format. Provide format or specify infer_datetime_format=True for consistent parsing.
  cache_array = _maybe_cache(arg, format, cache, convert_listlike)
/Users/hakim/opt/anaconda3/lib/python3.9/site-packages/pandas/core/tools/datetimes.py:1047: UserWarning: Parsing '28/05/2022' in DD/MM/YYYY format. Provide format or specify infer_datetime_format=True for consistent parsing.
  cache_array = _maybe_cache(arg, format, cache, convert_listlike)
/Users/hakim/opt/anaconda3/lib/python3.9/site-packages/pandas/core/tools/datetimes.py:1047: UserWarning: Parsing '29/05/2022' in DD/MM/YYYY format. Provide format or specify infer_datetime_format=True for consistent parsing.
  cache_array = _maybe_cache(arg, format, cache, convert_listlike)
/Users/hakim/opt/anaconda3/lib/python3.9/site-packages/pandas/core/tools/datetimes.py:1047: UserWarning: Parsing '30/05/2022' in DD/MM/YYYY format. Provide format or specify infer_datetime_format=True for consistent parsing.
  cache_array = _maybe_cache(arg, format, cache, convert_listlike)
/Users/hakim/opt/anaconda3/lib/python3.9/site-packages/pandas/core/tools/datetimes.py:1047: UserWarning: Parsing '31/05/2022' in DD/MM/YYYY format. Provide format or specify infer_datetime_format=True for consistent parsing.
  cache_array = _maybe_cache(arg, format, cache, convert_listlike)
/Users/hakim/opt/anaconda3/lib/python3.9/site-packages/pandas/core/tools/datetimes.py:1047: UserWarning: Parsing '13/06/2022' in DD/MM/YYYY format. Provide format or specify infer_datetime_format=True for consistent parsing.
  cache_array = _maybe_cache(arg, format, cache, convert_listlike)
/Users/hakim/opt/anaconda3/lib/python3.9/site-packages/pandas/core/tools/datetimes.py:1047: UserWarning: Parsing '14/06/2022' in DD/MM/YYYY format. Provide format or specify infer_datetime_format=True for consistent parsing.
  cache_array = _maybe_cache(arg, format, cache, convert_listlike)
/Users/hakim/opt/anaconda3/lib/python3.9/site-packages/pandas/core/tools/datetimes.py:1047: UserWarning: Parsing '15/06/2022' in DD/MM/YYYY format. Provide format or specify infer_datetime_format=True for consistent parsing.
  cache_array = _maybe_cache(arg, format, cache, convert_listlike)
/Users/hakim/opt/anaconda3/lib/python3.9/site-packages/pandas/core/tools/datetimes.py:1047: UserWarning: Parsing '16/06/2022' in DD/MM/YYYY format. Provide format or specify infer_datetime_format=True for consistent parsing.
  cache_array = _maybe_cache(arg, format, cache, convert_listlike)
/Users/hakim/opt/anaconda3/lib/python3.9/site-packages/pandas/core/tools/datetimes.py:1047: UserWarning: Parsing '17/06/2022' in DD/MM/YYYY format. Provide format or specify infer_datetime_format=True for consistent parsing.
  cache_array = _maybe_cache(arg, format, cache, convert_listlike)
/Users/hakim/opt/anaconda3/lib/python3.9/site-packages/pandas/core/tools/datetimes.py:1047: UserWarning: Parsing '18/06/2022' in DD/MM/YYYY format. Provide format or specify infer_datetime_format=True for consistent parsing.
  cache_array = _maybe_cache(arg, format, cache, convert_listlike)
/Users/hakim/opt/anaconda3/lib/python3.9/site-packages/pandas/core/tools/datetimes.py:1047: UserWarning: Parsing '19/06/2022' in DD/MM/YYYY format. Provide format or specify infer_datetime_format=True for consistent parsing.
  cache_array = _maybe_cache(arg, format, cache, convert_listlike)
/Users/hakim/opt/anaconda3/lib/python3.9/site-packages/pandas/core/tools/datetimes.py:1047: UserWarning: Parsing '20/06/2022' in DD/MM/YYYY format. Provide format or specify infer_datetime_format=True for consistent parsing.
  cache_array = _maybe_cache(arg, format, cache, convert_listlike)
/Users/hakim/opt/anaconda3/lib/python3.9/site-packages/pandas/core/tools/datetimes.py:1047: UserWarning: Parsing '21/06/2022' in DD/MM/YYYY format. Provide format or specify infer_datetime_format=True for consistent parsing.
  cache_array = _maybe_cache(arg, format, cache, convert_listlike)
/Users/hakim/opt/anaconda3/lib/python3.9/site-packages/pandas/core/tools/datetimes.py:1047: UserWarning: Parsing '22/06/2022' in DD/MM/YYYY format. Provide format or specify infer_datetime_format=True for consistent parsing.
  cache_array = _maybe_cache(arg, format, cache, convert_listlike)
/Users/hakim/opt/anaconda3/lib/python3.9/site-packages/pandas/core/tools/datetimes.py:1047: UserWarning: Parsing '23/06/2022' in DD/MM/YYYY format. Provide format or specify infer_datetime_format=True for consistent parsing.
  cache_array = _maybe_cache(arg, format, cache, convert_listlike)
/Users/hakim/opt/anaconda3/lib/python3.9/site-packages/pandas/core/tools/datetimes.py:1047: UserWarning: Parsing '24/06/2022' in DD/MM/YYYY format. Provide format or specify infer_datetime_format=True for consistent parsing.
  cache_array = _maybe_cache(arg, format, cache, convert_listlike)
/Users/hakim/opt/anaconda3/lib/python3.9/site-packages/pandas/core/tools/datetimes.py:1047: UserWarning: Parsing '25/06/2022' in DD/MM/YYYY format. Provide format or specify infer_datetime_format=True for consistent parsing.
  cache_array = _maybe_cache(arg, format, cache, convert_listlike)

In [106]: df_reviews

Out[106]:
```

	date	rating	comment
0	2022-05-23	3	Bikes a nice way to complement a walking tour ...
1	2022-05-23	5	Great time, good jokes and Tim was great too
2	2022-05-23	4	Loved our green bikes tours! We did the Night ...
3	2022-05-23	4	NaN
4	2022-05-24	3	When I arrived in Paris, I took part in the Ni...
...	...	...	...
151	2022-06-25	1	Really disappointed with Green Bike tours. We ...
152	2022-06-25	5	Impeccable!
153	2022-06-25	2	NaN
154	2022-06-25	1	Another bike tour, selling tickets at highly i...
155	2022-06-25	5	NaN

156 rows × 3 columns

```
In [107]: #Tri de 'date' par ordre décroissant pour voir les commentaires les plus récents aux plus anciens
df_reviews = df_reviews.sort_values(by='date', ascending=False)

In [109]: df_reviews

Out[109]:
```

	date	rating	comment
91	2022-12-06	5	phenobenal treatment!
88	2022-12-06	2	We took the Paris night bike tour hoping to ge...
89	2022-12-06	4	Great service, very friendly people. Our tour ...
90	2022-12-06	5	amazing tour, we felt safe and an illuminated ...
87	2022-11-06	4	Tour was informative, fun and generally a grea...
...	...	...	...
50	2022-01-06	5	Very nice guides and the information given was...
49	2022-01-06	4	tour was fantastic - made more so by our guide...
48	2022-01-06	4	Great tour! I would do it again our tour guide...
47	2022-01-06	4	Great tour with a very knowledgeable guide.
45	2022-01-06	5	Lovely bicycle tour of Paris this afternoon wi...

156 rows × 3 columns

```
In [108]: #Output
df_reviews.to_csv('Output\\tour_reviews.csv', index=False)
```