In [1]:

```
#Importation et affichage
```

In [55]:

```
pwd
```

Out[55]:

```
'/Users/Yaha/Desktop/CAPSTONE 2/CONCAT'
```

In [56]:

```
import pandas as pd
```

In [74]:

```
df=pd.read_excel('/Users/Yaha/Desktop/CAPSTONE 2/Dataset/Office Schedule/Employee_Evening
_Schedule.xlsx', None)
```

In [75]:

```
#Changement des colonnes 'No.People' et 'duty' renommées en 'name'
```

In [76]:

```
df['13 06 2022'] = df['13 06 2022'].rename(columns ={'Name':'name', 'No. People':'duty'}
)
```

In [77]:

```
#Suppression des premières lignes vides sur la feuille nommée "05 06 2022" dans df
```

In [78]:

```
df['05 06 2022'] = pd.DataFrame(df['05 06 2022'].values[3:], columns=df['05 06 2022'].il
oc[2])
```

In [79]:

```
#Concaténation
```

In [80]:

```
list_fichier = list(df.keys())
```

In [81]:

```
df_new_Office_Schedule=pd.DataFrame()
```

In [82]:

```
list_fichier
```

Out[82]:

```
['23 05 2022',
 '24 05 2022',
 '25 05 2022',
 '26 05 2022',
 '27 05 2022',
 '28 05 2022',
 '29 05 2022',
 '30 05 2022',
 '31 05 2022',
 '01 06 2022',
 '02 06 2022',
```

```
 '03 06 2022',
 '04 06 2022',
 '05 06 2022',
 '06 06 2022',
 '07 06 2022',
 '08 06 2022',
 '09 06 2022',
 '10 06 2022',
 '11 06 2022',
 '12 06 2022',
 '13 06 2022',
 '14 06 2022',
 '15 06 2022',
 '16 06 2022',
 '17 06 2022',
 '18 06 2022',
 '19 06 2022']
```

In [83]:

```
for fichier in list_fichier:
    A=df[fichier]
    A['tour_date']=fichier
    df_new_Office_Schedule = pd.concat([df_new_Office_Schedule, A])
```

In [84]:

```
df_new_Office_Schedule[df_new_Office_Schedule.columns]
```

Out[84]:

	name	duty	start_time	end_time	tour_date	leaving_time
0	Tom Burden	Office Shift - M	16:00:00	19:30:00	23 05 2022	NaN
1	Tim Chang	Office Shift	17:00:00	19:30:00	23 05 2022	NaN
2	Ben Jonsen	Office Shift	17:30:00	19:30:00	23 05 2022	NaN
3	Laura Tobin	Guide 1	17:30:00	23:30:00	23 05 2022	NaN
4	Emily Ngo	Guide 2	17:30:00	23:30:00	23 05 2022	NaN
...	...	...	...	...	...	...
12	Nolan Pendergast	Guide 10	17:30:00	23:30:00	19 06 2022	NaN
13	David Chiswell	Guide 11	17:30:00	23:30:00	19 06 2022	NaN
14	Sally Kittman	Guide 12	17:30:00	23:30:00	19 06 2022	NaN
15	Andrew Brewer	Guide 13	17:30:00	23:30:00	19 06 2022	NaN
16	Nicola Georgi	Office Shift - M	22:00:00	23:30:00	19 06 2022	NaN

322 rows × 6 columns

In [93]:

```
#Détection des valeurs manquantes
df_new_Office_Schedule.isnull().sum()
```

In [94]:

```
#La colonne 'leaving_time' contient 321 valeurs vides.
#Après vérification, la seule donnée présente dans la colonne 'leaving_time' n'est pas pe
rtinente pour l'exploitation.
```

In [95]:

```
#On sélectionne et garde que les colonnes qui ne sont pas vides
```

In [97]:

```
df_new_Office_Schedule=df_new_Office_Schedule[df_new_Office_Schedule.columns[0:5]]
```

In [48]:

```
df_new_Office_Schedule
```

Out[48]:

	name	duty	start_time	end_time	tour_date
0	Tom Burden	Office Shift - M	16:00:00	19:30:00	23 05 2022
1	Tim Chang	Office Shift	17:00:00	19:30:00	23 05 2022
2	Ben Jonsen	Office Shift	17:30:00	19:30:00	23 05 2022
3	Laura Tobin	Guide 1	17:30:00	23:30:00	23 05 2022
4	Emily Ngo	Guide 2	17:30:00	23:30:00	23 05 2022
...	...	...	...	...	...
12	Nolan Pendergast	Guide 10	17:30:00	23:30:00	19 06 2022
13	David Chiswell	Guide 11	17:30:00	23:30:00	19 06 2022
14	Sally Kittman	Guide 12	17:30:00	23:30:00	19 06 2022
15	Andrew Brewer	Guide 13	17:30:00	23:30:00	19 06 2022
16	Nicola Georgi	Office Shift - M	22:00:00	23:30:00	19 06 2022

322 rows × 5 columns

In [111]:

```
#Uniformisation du format 'tour_date'
```

In [98]:

```
#df_new_Office_Schedule['tour_date'] = pd.to_datetime(df_new_Office_Schedule['tour_date']
)
```

In [99]:

```
#df_new_Office_Schedule
```

Out[99]:

	name	duty	start_time	end_time	tour_date
0	Tom Burden	Office Shift - M	16:00:00	19:30:00	23 05 2022
1	Tim Chang	Office Shift	17:00:00	19:30:00	23 05 2022
2	Ben Jonsen	Office Shift	17:30:00	19:30:00	23 05 2022
3	Laura Tobin	Guide 1	17:30:00	23:30:00	23 05 2022
4	Emily Ngo	Guide 2	17:30:00	23:30:00	23 05 2022
...	...	...	...	...	...
12	Nolan Pendergast	Guide 10	17:30:00	23:30:00	19 06 2022
13	David Chiswell	Guide 11	17:30:00	23:30:00	19 06 2022
14	Sally Kittman	Guide 12	17:30:00	23:30:00	19 06 2022
15	Andrew Brewer	Guide 13	17:30:00	23:30:00	19 06 2022
16	Nicola Georgi	Office Shift - M	22:00:00	23:30:00	19 06 2022

322 rows × 5 columns

322 rows × 5 columns

In [101]:

```
#Informations sur la dataframe
```

In [102]:

```
df_new_Office_Schedule.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 322 entries, 0 to 16
Data columns (total 5 columns):
 #   Column      Non-Null Count  Dtype 
---  ------      --------------  ----- 
 0   name        322 non-null    object
 1   duty        322 non-null    object
 2   start_time  322 non-null    object
 3   end_time    322 non-null    object
 4   tour_date   322 non-null    object
dtypes: object(5)
memory usage: 15.1+ KB
```

In [103]:

```
#Détection des doublons
```

In [104]:

```
df_new_Office_Schedule.duplicated().sum()
```

Out[104]:

```
0
```

In [105]:

```
#Détection des valeurs manquantes
```

In [106]:

```
df_new_Office_Schedule.isnull().sum()
```

Out[106]:

```
name          0
duty          0
start_time    0
end_time      0
tour_date     0
dtype: int64
```

In [107]:

```
#Output
df_new_Office_Schedule.to_csv('Output\\Employee_Evening_Schedule.csv',index=False)
```

In [ ]:

```
#Visualisation du code en ligne
https://cutt.ly/aVmUxcH
```