

Python Project Assignment using Pandas Library

The difficulty of this assignment has been tuned to provide more open ended goal rather than just giving you the task so that you can polish your problem solving skills along with logic building skills.

Best Of Luck!

Do not take any help from any AI tool. Try to complete it by yourself.

Importing the dataset

- Use the csv library to import the artworks dataset

```
In [27]: 1 import csv
          2 import pandas as pd
```

```
In [28]: 1 dataset = pd.read_csv('artworks.csv', encoding = 'utf-8')
```

```
In [29]: 1 dataset.head(3)
```

```
Out[29]:
```

	Title	Artist	Nationality	BeginDate	EndDate	Gender	Date	Department
0	Dress MacLeod from Tartan Sets	Sarah Charlesworth	(American)	-1947.0	-2013.0	(Female)	1986	Prints & Illustrated Books
1	Duplicate of plate from folio 11 verso (supple...	Pablo Palazuelo	(Spanish)	-1916.0	-2007.0	(Male)	1978	Prints & Illustrated Books
2	Tailpiece (page 55) from SAGESSE	Maurice Denis	(French)	-1870.0	-1943.0	(Male)	1889- 1911	Prints & Illustrated Books

```
In [30]: 1 dataset.columns
```

```
Out[30]: Index(['Title', 'Artist', 'Nationality', 'BeginDate', 'EndDate', 'Gender',  
               'Date', 'Department'],  
              dtype='object')
```

```
In [31]: 1 dataset.isnull().sum()
```

```
Out[31]: Title      1  
         Artist      0  
         Nationality  0  
         BeginDate   942  
         EndDate     6254  
         Gender      0  
         Date        0  
         Department  0  
         dtype: int64
```

	A	B	C	D	E	F	G	H	I	J	K	L
	Title	Artist	Nationality	BeginDate	EndDate	Gender	Date	Department				
1	Dress MacLeod from Tartan Sets	Sarah Charlesworth	(American)	-1947	-2013	(Female)	1986	Prints & Illustrated Books				
2	ie from folio 11 verso (supplementary suite, plate	Pablo Palazuelo	(Spanish)	-1916	-2007	(Male)	1978	Prints & Illustrated Books				
3	Tailpiece (page 55) from SAGESSE	Maurice Denis	(French)	-1870	-1943	(Male)	1889-1911	Prints & Illustrated Books				
4	(page 129) from LIVRET DE FOULASTRIES, À ÉJANOT	Aristide Maillol	(French)	-1861	-1944	(Male)	1927-1940	Prints & Illustrated Books				
5	97 rue du Bac	Eugène Atget	(French)	-1857	-1927	(Male)	1903	Photography				
6	Pictorial ornament (folio 11) from WOODCUTS	Antonio Frasconi	(American)	-1919	-2013	(Male)	1957	Prints & Illustrated Books				
7	Rue de l'Hôtel-de-Ville	Eugène Atget	(French)	-1857	-1927	(Male)	1924	Photography				
8	Los Angeles Airport	Garry Winogrand	(American)	-1928	-1984	(Male)	1978-1983	Photography				
9	Why Defy from Disasters of Peace	Diane Victor	South African	-1964		(Female)	2001	Prints & Illustrated Books				
10	STILL WATER AND FISH	David Brown Milne	(Canadian)	-1882	-1953	(Male)	-1941	Prints & Illustrated Books				
11	In-text plate (folio 11) from LA MÃTROMANIE	Jean Dubuffet	(French)	-1901	-1985	(Male)	1949-1950	Prints & Illustrated Books				
12	Black Bathroom	Jim Dine	(American)	-1935		(Male)	1963	Prints & Illustrated Books				
13	In-text plate (page 108) from LYSISTRATA	František Kupka	(Czech)	-1871	-1957	(Male)	1908-1911	Prints & Illustrated Books				
14	Transcendence. Designs for the Ballet	Franklin Chenault Watkins	(American)	-1894	-1972	(Male)	-1934	Drawings				
15	Untitled	Christopher Wool	(American)	-1955		(Male)	1997	Prints & Illustrated Books				
16	Isadora Duncan	Abraham Walkowitz	(American)	-1878	-1965	(Male)	1931-1933	Drawings				
17	L'Avenir de la Propriété	Pierre Alechinsky	(Belgian)	-1927		(Male)	1972	Prints & Illustrated Books				
18	Arundel Castle from Black Series I	Frank Stella	(American)	-1936		(Male)	1967	Prints & Illustrated Books				
19	gments from the Samuel Freeman House, Los Ang	Frank Lloyd Wright	(American)	-1867	-1959	(Male)	1923-1924	Architecture & Design				
20	Plate (folio 35) from ACORDE	Vicente Rojo	(Spanish)	-1932		(Male)	1979	Prints & Illustrated Books				
21	Wolf House, Gubin, Poland (Third floor plan)	Ludwig Mies van der Rohe	(American)	-1886	-1969	(Male)	1925-1927	Architecture & Design				
22	ennaia Arkhitektura (CA. Contemporary Architec	Varvara Stepanova	(Russian)	-1894	-1958	(Female)	1929	Prints & Illustrated Books				

Skill Test - 1

- Using python related functions and attributes, clean the Nationality and Gender Columns

```
In [32]: 1 dataset['Nationality'] = dataset['Nationality'].str.replace('(', '', re
```

```
In [33]: 1 dataset['Nationality']
```

```
Out[33]: 0      American
1      Spanish
2      French
3      French
4      French
...
16724   British
16725   Chilean
16726   American
16727
16728   American
Name: Nationality, Length: 16729, dtype: object
```

```
In [34]: 1 dataset['Gender'] = dataset['Gender'].str.replace('(', '', regex=False)
```

```
In [35]: 1 dataset['Gender']
```

```
Out[35]: 0      Female
          1      Male
          2      Male
          3      Male
          4      Male
          ...
        16724    Male
        16725    Male
        16726    Male
        16727
        16728    Male
Name: Gender, Length: 16729, dtype: object
```

Skill Test - 2

- Clean BeginDate and EndDate Columns

```
In [36]: 1 dataset['BeginDate'] = dataset['BeginDate'].astype(str).str.replace('-', ''
```

```
In [37]: 1 # pd.to_numeric = Converts the column to numeric values
          2 # error='coerce' = It tells pandas how to handle errors that occur duri
          3 dataset['BeginDate'] = pd.to_numeric(dataset['BeginDate'], errors='coer
          4 dataset['BeginDate'].head(20)
```

```
Out[37]: 0      1947
          1      1916
          2      1870
          3      1861
          4      1857
          5      1919
          6      1857
          7      1928
          8      1964
          9      1882
         10      1901
         11      1935
         12      1871
         13      1894
         14      1955
         15      1878
         16      1927
         17      1936
         18      1867
         19      1932
Name: BeginDate, dtype: Int64
```

```
In [38]: 1 dataset['EndDate'] = dataset['EndDate'].astype(str).str.replace('-', '')
```

```
In [39]: 1 dataset['EndDate'] = pd.to_numeric(dataset['EndDate'], errors='coerce')
         2 dataset['EndDate'].head(20)
```

```
Out[39]: 0      2013
         1      2007
         2      1943
         3      1944
         4      1927
         5      2013
         6      1927
         7      1984
         8      <NA>
         9      1953
        10      1985
        11      <NA>
        12      1957
        13      1972
        14      <NA>
        15      1965
        16      <NA>
        17      <NA>
        18      1959
        19      <NA>
        Name: EndDate, dtype: Int64
```

Skill Test - 3

- The BeginDate and EndDate are BirthDate and Day of Death for each Artist
- Find out the age of each artist and add it into the dataset

```
In [40]: 1 dataset['Age'] = dataset['EndDate'] - dataset['BeginDate']
```

```
In [41]: 1 dataset['Age'] = dataset['Age'].astype(str).str.replace('-', '')
```

```
In [42]: 1 dataset['Age']
```

```
Out[42]: 0      66
         1      91
         2      73
         3      83
         4      70
         ...
        16724    88
        16725    <NA>
        16726    85
        16727    <NA>
        16728    <NA>
        Name: Age, Length: 16729, dtype: object
```

Skill Test - 4

- Clean the Date Column excluding the hyphen (-) in between as it specifies a range of date

Hint: Make sure to find each possible character in the column which makes the date

```
In [43]: 1 dataset['Date'] = (dataset['Date'].str.replace(' ', '', regex=False).str
```

```
In [44]: 1 dataset['Date'].head(20)
```

```
Out[44]: 0          1986
1          1978
2      1889-1911
3      1927-1940
4          1903
5          1957
6          1924
7      1978-1983
8          2001
9          -1941
10     1949-1950
11          1963
12     1908-1911
13          -1934
14          1997
15     1931-1933
16          1972
17          1967
18     1923-1924
19          1979
Name: Date, dtype: object
```

startswith()

- It returns True if the string starts with the specified value, otherwise False
- Syntax = string.startswith(value, start, end)

```
In [45]: 1 '-1947'.startswith('-')
```

```
Out[45]: True
```

The above method can be used as per your ease in any skill test, if you want to. Using it doesn't award you any special remarks.

Skill Test - 5

- After cleaning the Date Column, split it into two columns IF the date has a starting and ending range else don't split.
- For each date being split, calculate the average date and store it in a new column.

```
In [46]: 1 def clean_date(value):
2         if value.startswith('-'):
3             return value.replace('-', '')
4         return value
```

```
In [47]: 1 dataset['Date'] = dataset['Date'].apply(clean_date)
```

```
In [48]: 1 dataset['Date'].head(20)
```

```
Out[48]: 0      1986
1      1978
2    1889-1911
3    1927-1940
4      1903
5      1957
6      1924
7    1978-1983
8      2001
9      1941
10   1949-1950
11      1963
12   1908-1911
13      1934
14      1997
15   1931-1933
16      1972
17      1967
18   1923-1924
19      1979
Name: Date, dtype: object
```

```
In [49]: 1 def avg_date(date):
2         avg_values = []
3         for i in date:
4             if '-' in i:
5                 value = i.split('-')
6                 F = int(value[0])
7                 S = int(value[1])
8                 value = round((F + S) / 2)
9                 avg_values.append(value)
10            else:
11                avg_values.append(i)
12            return avg_values
```

```
In [50]: 1 dataset['Avg_date'] = avg_date(dataset['Date'])
```

```
In [51]: 1 dataset['Avg_date'].head(20)
```

```
Out[51]: 0      1986  
1      1978  
2      1900  
3      1934  
4      1903  
5      1957  
6      1924  
7      1980  
8      2001  
9      1941  
10     1950  
11     1963  
12     1910  
13     1934  
14     1997  
15     1932  
16     1972  
17     1967  
18     1924  
19     1979
```

Name: Avg_date, dtype: object

Storing the updated dataset

- Use pandas to save the updated dataset in csv format.
- Name the file as 'Artworks_updated.csv'.
- You can look at previous videos and files to find the code for it.

In [52]:

1	<code>dataset.head(20)</code>
---	-------------------------------

Out[52]:

	Title	Artist	Nationality	BeginDate	EndDate	Gender	Date	Department
0	Dress MacLeod from Tartan Sets	Sarah Charlesworth	American	1947	2013	Female	1986	Prints & Illustrated Books
1	Duplicate of plate from folio 11 verso (supple...	Pablo Palazuelo	Spanish	1916	2007	Male	1978	Prints & Illustrated Books
2	Tailpiece (page 55) from SAGESSE	Maurice Denis	French	1870	1943	Male	1889-1911	Prints & Illustrated Books
3	Headpiece (page 129) from LIVRET DE FOLASTRIES...	Aristide Maillol	French	1861	1944	Male	1927-1940	Prints & Illustrated Books
4	97 rue du Bac	Eugène Atget	French	1857	1927	Male	1903	Photography
5	Pictorial ornament (folio 11) from WOODCUTS	Antonio Frasconi	American	1919	2013	Male	1957	Prints & Illustrated Books
6	Rue de l'Hôtel-de-Ville	Eugène Atget	French	1857	1927	Male	1924	Photography
7	Los Angeles Airport	Garry Winogrand	American	1928	1984	Male	1978-1983	Photography
8	Why Defy from Disasters of Peace	Diane Victor	South African	1964	<NA>	Female	2001	Prints & Illustrated Books
9	STILL WATER AND FISH	David Brown Milne	Canadian	1882	1953	Male	1941	Prints & Illustrated Books
10	In-text plate (folio 11) from LA MÉTROMANIE	Jean Dubuffet	French	1901	1985	Male	1949-1950	Prints & Illustrated Books
11	Black Bathroom	Jim Dine	American	1935	<NA>	Male	1963	Prints & Illustrated Books
12	In-text plate (page 108) from LYSISTRATA	František Kupka	Czech	1871	1957	Male	1908-1911	Prints & Illustrated Books
13	Transcendence. Designs for the Ballet	Franklin Chenault Watkins	American	1894	1972	Male	1934	Drawings
14	Untitled	Christopher Wool	American	1955	<NA>	Male	1997	Prints & Illustrated Books
15	Isadora Duncan	Abraham Walkowitz	American	1878	1965	Male	1931-1933	Drawings
16	L'Avenir de la Propriété	Pierre Alechinsky	Belgian	1927	<NA>	Male	1972	Prints & Illustrated Books
17	Arundel Castle from Black Series I	Frank Stella	American	1936	<NA>	Male	1967	Prints & Illustrated Books

	Title	Artist	Nationality	BeginDate	EndDate	Gender	Date	Department
18	Block fragments from the Samuel Freeman House,...	Frank Lloyd Wright	American	1867	1959	Male	1923-1924	Architecture & Design
19	Plate (folio 35) from ACORDE	Vicente Rojo	Spanish	1932	<NA>	Male	1979	Prints & Illustrated Books

In [53]:

```
1 dataset.to_csv( 'Artworks_updated.csv', index=False)
```