```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

new_data = pd.read_csv('vgsales.csv')
new_data
```

```
        Rank                                                  Name
Platform  \
0          1                                        Wii Sports
Wii
1          2                                 Super Mario Bros.
NES
2          3                                    Mario Kart Wii
Wii
3          4                                 Wii Sports Resort
Wii
4          5                           Pokemon Red/Pokemon Blue
GB
...      ...                                               ...       ..
.
16593  16596                Woody Woodpecker in Crazy Castle 5
GBA
16594  16597                      Men in Black II: Alien Escape
GC
16595  16598  SCORE International Baja 1000: The Official Game
PS2
16596  16599                                        Know How 2
DS
16597  16600                                  Spirits & Spells
GBA

          Year         Genre   Publisher  NA_Sales  EU_Sales  JP_Sales
\
0      2006.0        Sports    Nintendo     41.49     29.02      3.77

1      1985.0      Platform    Nintendo     29.08      3.58      6.81

2      2008.0        Racing    Nintendo     15.85     12.88      3.79

3      2009.0        Sports    Nintendo     15.75     11.01      3.28

4      1996.0  Role-Playing    Nintendo     11.27      8.89     10.22

...       ...           ...         ...       ...       ...       ...

16593  2002.0      Platform       Kemco      0.01      0.00      0.00

16594  2003.0       Shooter  Infogrames      0.01      0.00      0.00
```

```
16595  2008.0        Racing  Activision      0.00      0.00      0.00

16596  2010.0        Puzzle    7G//AMES      0.00      0.01      0.00

16597  2003.0      Platform     Wanadoo      0.01      0.00      0.00


       Other_Sales  Global_Sales
0             8.46         82.74
1             0.77         40.24
2             3.31         35.82
3             2.96         33.00
4             1.00         31.37
...            ...           ...
16593         0.00          0.01
16594         0.00          0.01
16595         0.00          0.01
16596         0.00          0.01
16597         0.00          0.01

[16598 rows x 11 columns]
```

```
new_data.head()
```

```
   Rank                      Name Platform    Year          Genre
Publisher  \
0     1                Wii Sports      Wii  2006.0         Sports
Nintendo
1     2         Super Mario Bros.      NES  1985.0       Platform
Nintendo
2     3            Mario Kart Wii      Wii  2008.0         Racing
Nintendo
3     4         Wii Sports Resort      Wii  2009.0         Sports
Nintendo
4     5  Pokemon Red/Pokemon Blue       GB  1996.0   Role-Playing
Nintendo

   NA_Sales  EU_Sales  JP_Sales  Other_Sales  Global_Sales
0     41.49     29.02      3.77         8.46         82.74
1     29.08      3.58      6.81         0.77         40.24
2     15.85     12.88      3.79         3.31         35.82
3     15.75     11.01      3.28         2.96         33.00
4     11.27      8.89     10.22         1.00         31.37
```

```
new_data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 16598 entries, 0 to 16597
Data columns (total 11 columns):
 #   Column        Non-Null Count  Dtype
---  ------        --------------  -----
```

```
 0   Rank          16598 non-null  int64
 1   Name          16598 non-null  object
 2   Platform      16598 non-null  object
 3   Year          16327 non-null  float64
 4   Genre         16598 non-null  object
 5   Publisher     16540 non-null  object
 6   NA_Sales      16598 non-null  float64
 7   EU_Sales      16598 non-null  float64
 8   JP_Sales      16598 non-null  float64
 9   Other_Sales   16598 non-null  float64
 10  Global_Sales  16598 non-null  float64
dtypes: float64(6), int64(1), object(4)
memory usage: 1.4+ MB
```

```
new_data.describe()
```

```
               Rank          Year      NA_Sales      EU_Sales  \
count  16598.000000  16327.000000  16598.000000  16598.000000
mean    8300.605254   2006.406443      0.264667      0.146652
std     4791.853933      5.828981      0.816683      0.505351
min        1.000000   1980.000000      0.000000      0.000000
25%     4151.250000   2003.000000      0.000000      0.000000
50%     8300.500000   2007.000000      0.080000      0.020000
75%    12449.750000   2010.000000      0.240000      0.110000
max    16600.000000   2020.000000     41.490000     29.020000

       JP_Sales  \
count  16598.000000
mean       0.077782
std        0.309291
min        0.000000
25%        0.000000
50%        0.000000
75%        0.040000
max       10.220000

        Other_Sales  Global_Sales
count  16598.000000  16598.000000
mean       0.048063      0.537441
std        0.188588      1.555028
min        0.000000      0.010000
25%        0.000000      0.060000
50%        0.010000      0.170000
75%        0.040000      0.470000
max       10.570000     82.740000
```

```
new_data.shape
```

```
(16598, 11)
```

```
new_data.columns
```

```
Index(['Rank', 'Name', 'Platform', 'Year', 'Genre', 'Publisher',
'NA_Sales',
       'EU_Sales', 'JP_Sales', 'Other_Sales', 'Global_Sales'],
      dtype='object')
```

```
new_data.isnull()
```

```
        Rank   Name  Platform   Year  Genre  Publisher  NA_Sales
EU_Sales  \
0      False  False     False  False  False      False     False
False
1      False  False     False  False  False      False     False
False
2      False  False     False  False  False      False     False
False
3      False  False     False  False  False      False     False
False
4      False  False     False  False  False      False     False
False
...      ...    ...       ...    ...    ...        ...       ...
...
16593  False  False     False  False  False      False     False
False
16594  False  False     False  False  False      False     False
False
16595  False  False     False  False  False      False     False
False
16596  False  False     False  False  False      False     False
False
16597  False  False     False  False  False      False     False
False

       JP_Sales  Other_Sales  Global_Sales
0         False        False         False
1         False        False         False
2         False        False         False
3         False        False         False
4         False        False         False
...         ...          ...           ...
16593     False        False         False
16594     False        False         False
16595     False        False         False
16596     False        False         False
16597     False        False         False

[16598 rows x 11 columns]
```

```
new_data.isnull().sum()
```

```
Rank              0
Name              0
Platform          0
Year            271
Genre             0
Publisher        58
NA_Sales          0
EU_Sales          0
JP_Sales          0
Other_Sales       0
Global_Sales      0
dtype: int64
```

```
A = new_data['Year'].fillna(new_data['Year'].mean())
A
```

```
0        2006.0
1        1985.0
2        2008.0
3        2009.0
4        1996.0
          ...
16593    2002.0
16594    2003.0
16595    2008.0
16596    2010.0
16597    2003.0
Name: Year, Length: 16598, dtype: float64
```

```
A.isnull()
```

```
0        False
1        False
2        False
3        False
4        False
         ...
16593    False
16594    False
16595    False
16596    False
16597    False
Name: Year, Length: 16598, dtype: bool
```

```
A.isnull().sum()
```

```
0
```

```
A.isnull().sum().sum()
```

```
0
```

```
B = new_data['Publisher'].fillna(new_data['Publisher'].mode()[0])
B
```

```
0          Nintendo
1          Nintendo
2          Nintendo
3          Nintendo
4          Nintendo
            ...
16593        Kemco
16594    Infogrames
16595    Activision
16596      7G//AMES
16597       Wanadoo
Name: Publisher, Length: 16598, dtype: object
```

```
B.isnull()
```

```
0        False
1        False
2        False
3        False
4        False
         ...
16593    False
16594    False
16595    False
16596    False
16597    False
Name: Publisher, Length: 16598, dtype: bool
```

```
B.isnull().sum()
```

```
0
```

```
B.isnull().sum().sum()
```

```
0
```

```
filled_data = pd.concat([A,B],axis=1)
filled_data
```

```
         Year   Publisher
0      2006.0    Nintendo
1      1985.0    Nintendo
2      2008.0    Nintendo
3      2009.0    Nintendo
4      1996.0    Nintendo
...       ...         ...
16593  2002.0       Kemco
16594  2003.0  Infogrames
```

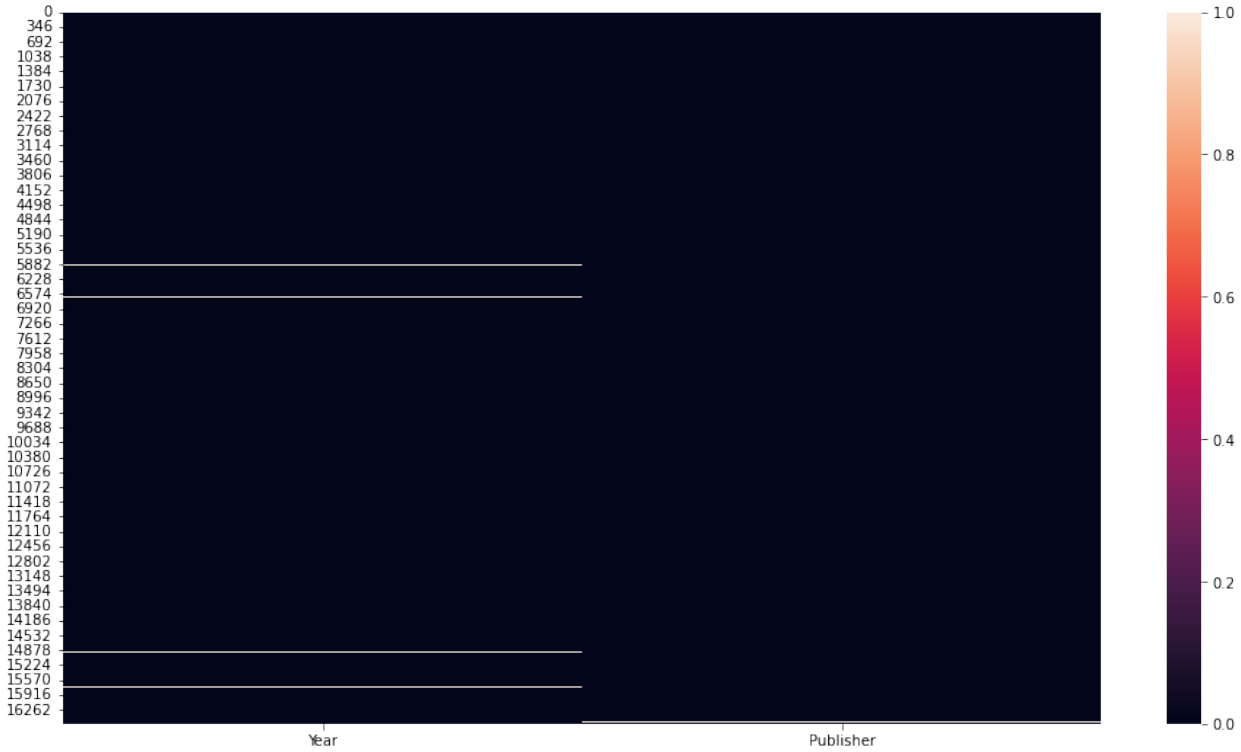
```
16595  2008.0  Activision
16596  2010.0    7G//AMES
16597  2003.0     Wanadoo

[16598 rows x 2 columns]
```

```
unfilled_data =
pd.concat((new_data['Year'],new_data['Publisher']),axis=1)
unfilled_data
```

```
          Year   Publisher
0      2006.0    Nintendo
1      1985.0    Nintendo
2      2008.0    Nintendo
3      2009.0    Nintendo
4      1996.0    Nintendo
...       ...         ...
16593  2002.0       Kemco
16594  2003.0  Infogrames
16595  2008.0  Activision
16596  2010.0    7G//AMES
16597  2003.0     Wanadoo

[16598 rows x 2 columns]
```

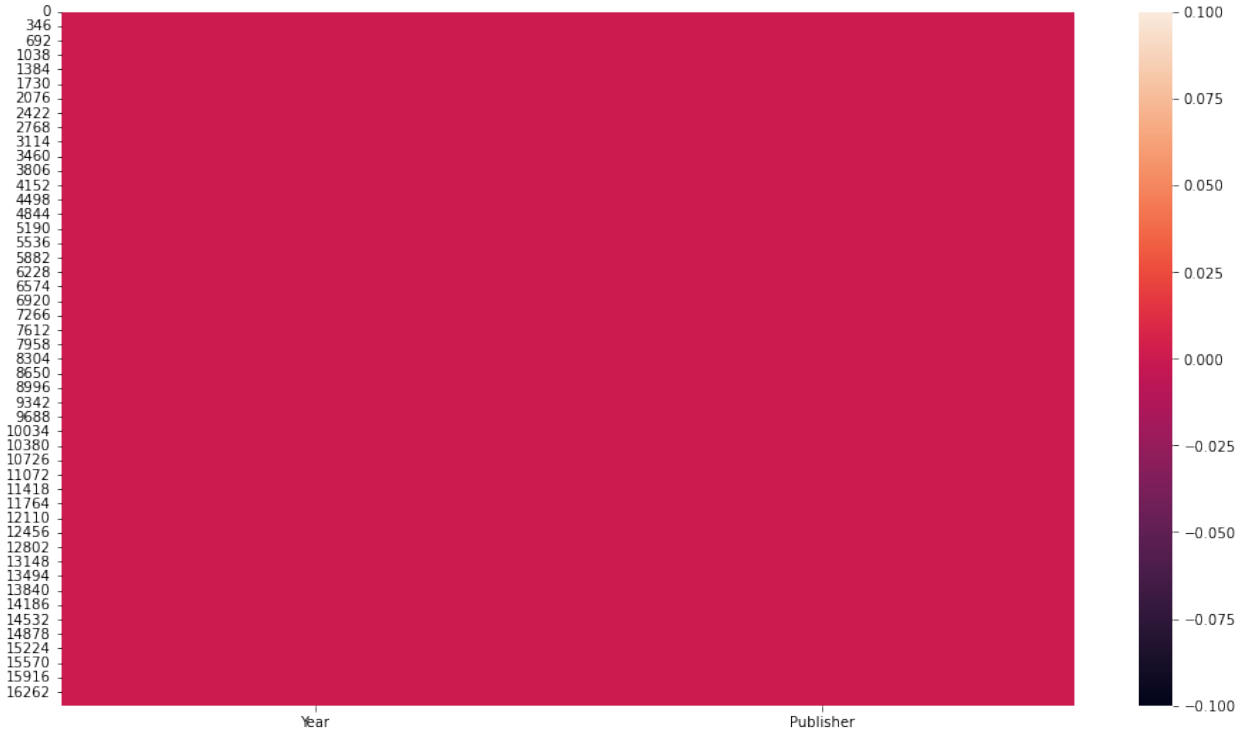
```
plt.figure(figsize=[16,9])
sns.heatmap(unfilled_data.isnull())
```

```
<AxesSubplot:>
```



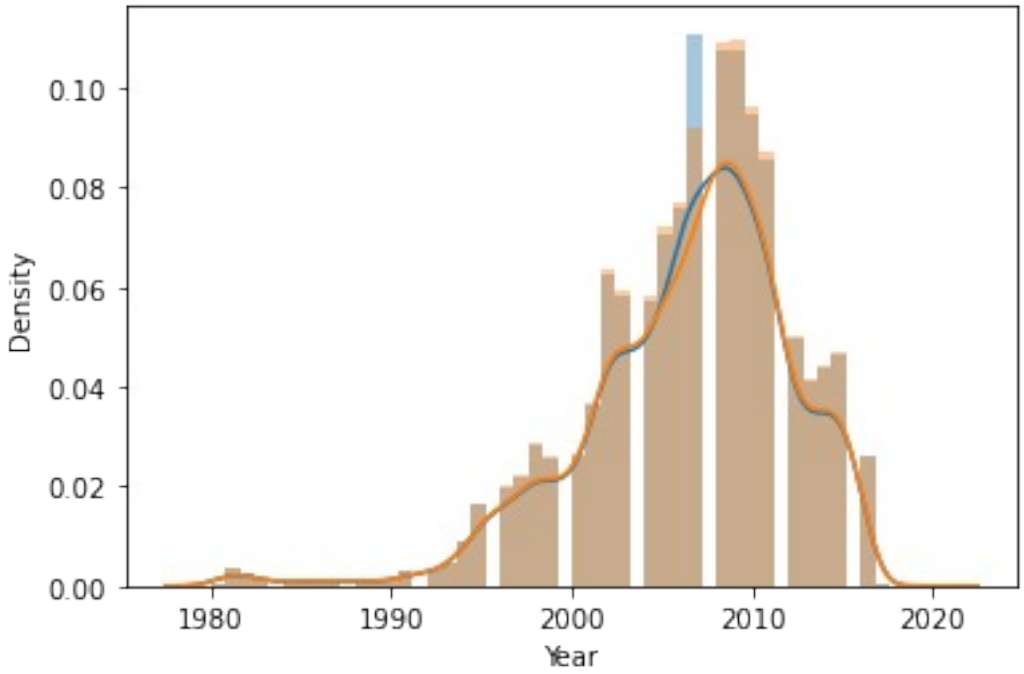

```
plt.figure(figsize=[16,9])
sns.heatmap(filled_data.isnull())
```

```
<AxesSubplot:>
```



```
sns.distplot(filled_data['Year'])
sns.distplot(new_data['Year'])
plt.show()
```

```
C:\ProgramData\Anaconda3\lib\site-packages\seaborn\
distributions.py:2619: FutureWarning: `distplot` is a deprecated
function and will be removed in a future version. Please adapt your
code to use either `displot` (a figure-level function with similar
flexibility) or `histplot` (an axes-level function for histograms).
  warnings.warn(msg, FutureWarning)
C:\ProgramData\Anaconda3\lib\site-packages\seaborn\
distributions.py:2619: FutureWarning: `distplot` is a deprecated
function and will be removed in a future version. Please adapt your
code to use either `displot` (a figure-level function with similar
flexibility) or `histplot` (an axes-level function for histograms).
  warnings.warn(msg, FutureWarning)
```




```
new_data['Publisher'].value_counts()
```

```
Electronic Arts                 1351
Activision                       975
Namco Bandai Games               932
Ubisoft                          921
Konami Digital Entertainment     832
                                 ...
Warp                               1
New                                1
Elite                              1
```

```
Evolution Games                     1
UIG Entertainment                   1
Name: Publisher, Length: 578, dtype: int64

new_data

        Rank                                                  Name
Platform  \
0          1                                        Wii Sports
Wii
1          2                                 Super Mario Bros.
NES
2          3                                    Mario Kart Wii
Wii
3          4                                 Wii Sports Resort
Wii
4          5                           Pokemon Red/Pokemon Blue
GB
...      ...                                                 ...      ..
.
16593  16596                Woody Woodpecker in Crazy Castle 5
GBA
16594  16597                     Men in Black II: Alien Escape
GC
16595  16598  SCORE International Baja 1000: The Official Game
PS2
16596  16599                                        Know How 2
DS
16597  16600                                  Spirits & Spells
GBA

         Year         Genre   Publisher  NA_Sales  EU_Sales  JP_Sales
\
0      2006.0        Sports    Nintendo     41.49     29.02      3.77

1      1985.0      Platform    Nintendo     29.08      3.58      6.81

2      2008.0        Racing    Nintendo     15.85     12.88      3.79

3      2009.0        Sports    Nintendo     15.75     11.01      3.28

4      1996.0  Role-Playing    Nintendo     11.27      8.89     10.22

...       ...           ...         ...       ...       ...       ...

16593  2002.0      Platform       Kemco      0.01      0.00      0.00

16594  2003.0       Shooter  Infogrames      0.01      0.00      0.00

16595  2008.0        Racing  Activision      0.00      0.00      0.00
```

```
16596  2010.0        Puzzle    7G//AMES      0.00      0.01      0.00

16597  2003.0      Platform     Wanadoo      0.01      0.00      0.00


       Other_Sales  Global_Sales
0             8.46         82.74
1             0.77         40.24
2             3.31         35.82
3             2.96         33.00
4             1.00         31.37
...            ...           ...
16593         0.00          0.01
16594         0.00          0.01
16595         0.00          0.01
16596         0.00          0.01
16597         0.00          0.01

[16598 rows x 11 columns]
```

```
new_data.columns
```

```
Index(['Rank', 'Name', 'Platform', 'Year', 'Genre', 'Publisher',
'NA_Sales',
       'EU_Sales', 'JP_Sales', 'Other_Sales', 'Global_Sales'],
      dtype='object')
```

```
reprised_data = new_data.drop(columns = ['Year','Publisher'])
reprised_data
```

```
        Rank                                                Name
Platform  \
0          1                                        Wii Sports
Wii
1          2                                 Super Mario Bros.
NES
2          3                                    Mario Kart Wii
Wii
3          4                                 Wii Sports Resort
Wii
4          5                           Pokemon Red/Pokemon Blue
GB
...      ...                                               ...      ..
.
16593  16596                Woody Woodpecker in Crazy Castle 5
GBA
16594  16597                     Men in Black II: Alien Escape
GC
16595  16598  SCORE International Baja 1000: The Official Game
PS2
16596  16599                                        Know How 2
```

```
DS
16597  16600                                  Spirits & Spells
GBA

               Genre  NA_Sales  EU_Sales  JP_Sales  Other_Sales
Global_Sales
0             Sports     41.49     29.02      3.77         8.46
82.74
1           Platform     29.08      3.58      6.81         0.77
40.24
2             Racing     15.85     12.88      3.79         3.31
35.82
3             Sports     15.75     11.01      3.28         2.96
33.00
4       Role-Playing     11.27      8.89     10.22         1.00
31.37
...              ...       ...       ...       ...          ...
...
16593       Platform      0.01      0.00      0.00         0.00
0.01
16594        Shooter      0.01      0.00      0.00         0.00
0.01
16595         Racing      0.00      0.00      0.00         0.00
0.01
16596         Puzzle      0.00      0.01      0.00         0.00
0.01
16597       Platform      0.01      0.00      0.00         0.00
0.01

[16598 rows x 9 columns]

reprised_data.isnull()

        Rank   Name  Platform  Genre  NA_Sales  EU_Sales  JP_Sales  \
0      False  False     False  False     False     False     False
1      False  False     False  False     False     False     False
2      False  False     False  False     False     False     False
3      False  False     False  False     False     False     False
4      False  False     False  False     False     False     False
...      ...    ...       ...    ...       ...       ...       ...
16593  False  False     False  False     False     False     False
16594  False  False     False  False     False     False     False
16595  False  False     False  False     False     False     False
16596  False  False     False  False     False     False     False
16597  False  False     False  False     False     False     False

       Other_Sales  Global_Sales
0            False         False
1            False         False
2            False         False
```

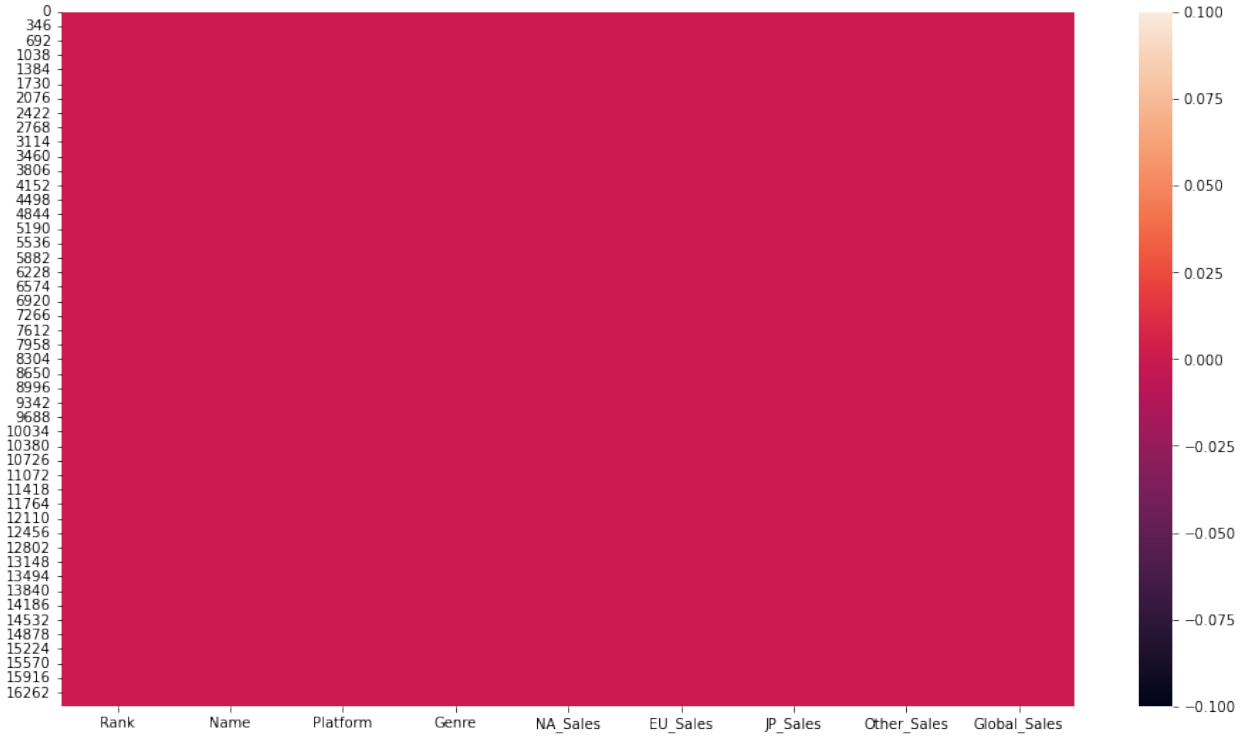
```
3             False         False
4             False         False
...             ...           ...
16593         False         False
16594         False         False
16595         False         False
16596         False         False
16597         False         False

[16598 rows x 9 columns]
```

```
reprised_data.isnull().sum()
```

```
Rank            0
Name            0
Platform        0
Genre           0
NA_Sales        0
EU_Sales        0
JP_Sales        0
Other_Sales     0
Global_Sales    0
dtype: int64
```

```
plt.figure(figsize=[16,9])
sns.heatmap(reprised_data.isnull())
```

```
<AxesSubplot:>
```



```
data_for_heatmap = new_data.select_dtypes(include =
['int64','float64'])
data_for_heatmap
```

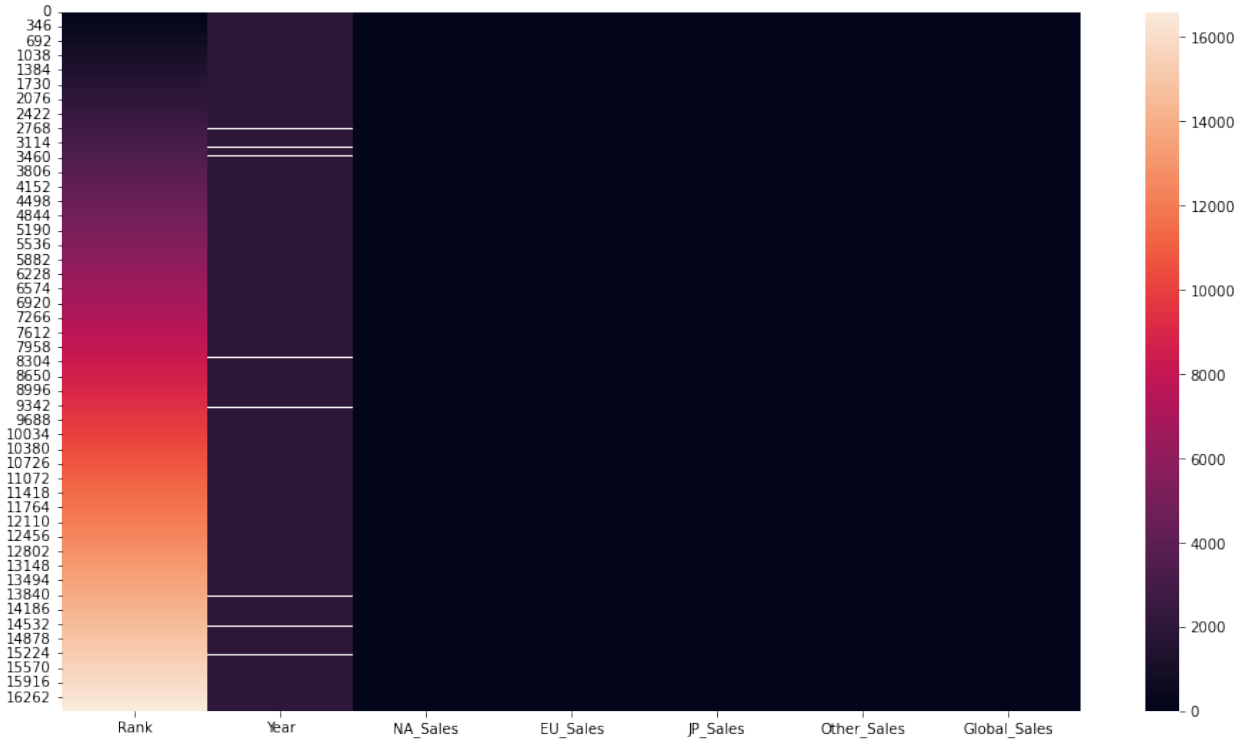
```
        Rank    Year  NA_Sales  EU_Sales  JP_Sales  Other_Sales
Global_Sales
0          1  2006.0     41.49     29.02      3.77         8.46
82.74
1          2  1985.0     29.08      3.58      6.81         0.77
40.24
2          3  2008.0     15.85     12.88      3.79         3.31
35.82
3          4  2009.0     15.75     11.01      3.28         2.96
33.00
4          5  1996.0     11.27      8.89     10.22         1.00
31.37
...      ...     ...       ...       ...       ...          ...
...
16593  16596  2002.0      0.01      0.00      0.00         0.00
0.01
16594  16597  2003.0      0.01      0.00      0.00         0.00
0.01
16595  16598  2008.0      0.00      0.00      0.00         0.00
0.01
16596  16599  2010.0      0.00      0.01      0.00         0.00
0.01
16597  16600  2003.0      0.01      0.00      0.00         0.00
0.01

[16598 rows x 7 columns]
```

```
data_for_heatmap.columns
```

```
Index(['Rank', 'Year', 'NA_Sales', 'EU_Sales', 'JP_Sales',
'Other_Sales',
       'Global_Sales'],
      dtype='object')
```

```
plt.figure(figsize=(16,9))
sns.heatmap(data_for_heatmap)
```

```
<AxesSubplot:>
```




```
data_for_heatmap.isnull()
```

	Rank	Year	NA_Sales	EU_Sales	JP_Sales	Other_Sales	Global_Sales
0	False	False	False	False	False	False	False
1	False	False	False	False	False	False	False
2	False	False	False	False	False	False	False
3	False	False	False	False	False	False	False
4	False	False	False	False	False	False	False
...	...	...	...	...	...	...	...
16593	False	False	False	False	False	False	False
16594	False	False	False	False	False	False	False
16595	False	False	False	False	False	False	False
16596	False	False	False	False	False	False	False
16597	False	False	False	False	False	False	False

```
[16598 rows x 7 columns]
```

```
data_for_heatmap.isnull().sum()
```

```
Rank              0
Year            271
NA_Sales          0
EU_Sales          0
JP_Sales          0
Other_Sales       0
Global_Sales      0
dtype: int64
```

```
new_data.columns
```

```
Index(['Rank', 'Name', 'Platform', 'Year', 'Genre', 'Publisher',
'NA_Sales',
       'EU_Sales', 'JP_Sales', 'Other_Sales', 'Global_Sales'],
      dtype='object')
```

```
reprised_dataforheatmap = pd.concat((filled_data['Year'],
new_data['Rank'], new_data['NA_Sales'], new_data['EU_Sales'],
                                     new_data['JP_Sales'],
new_data['Other_Sales'], new_data['Global_Sales']),axis=1)
reprised_dataforheatmap
```

```
         Year   Rank  NA_Sales  EU_Sales  JP_Sales  Other_Sales
Global_Sales
0      2006.0      1     41.49     29.02      3.77         8.46
82.74
1      1985.0      2     29.08      3.58      6.81         0.77
40.24
2      2008.0      3     15.85     12.88      3.79         3.31
35.82
3      2009.0      4     15.75     11.01      3.28         2.96
33.00
4      1996.0      5     11.27      8.89     10.22         1.00
31.37
...       ...    ...       ...       ...       ...          ...
...
16593  2002.0  16596      0.01      0.00      0.00         0.00
0.01
16594  2003.0  16597      0.01      0.00      0.00         0.00
0.01
16595  2008.0  16598      0.00      0.00      0.00         0.00
0.01
16596  2010.0  16599      0.00      0.01      0.00         0.00
0.01
16597  2003.0  16600      0.01      0.00      0.00         0.00
0.01
```

```
[16598 rows x 7 columns]
```

```
reprised_dataforheatmap.isnull()
```

```
        Year   Rank  NA_Sales  EU_Sales  JP_Sales  Other_Sales
Global_Sales
0      False  False     False     False     False        False
False
1      False  False     False     False     False        False
False
2      False  False     False     False     False        False
False
3      False  False     False     False     False        False
False
4      False  False     False     False     False        False
False
...      ...    ...       ...       ...       ...          ...
...
16593  False  False     False     False     False        False
False
16594  False  False     False     False     False        False
False
16595  False  False     False     False     False        False
False
16596  False  False     False     False     False        False
False
16597  False  False     False     False     False        False
False

[16598 rows x 7 columns]
```

```
reprised_dataforheatmap.isnull().sum()
```

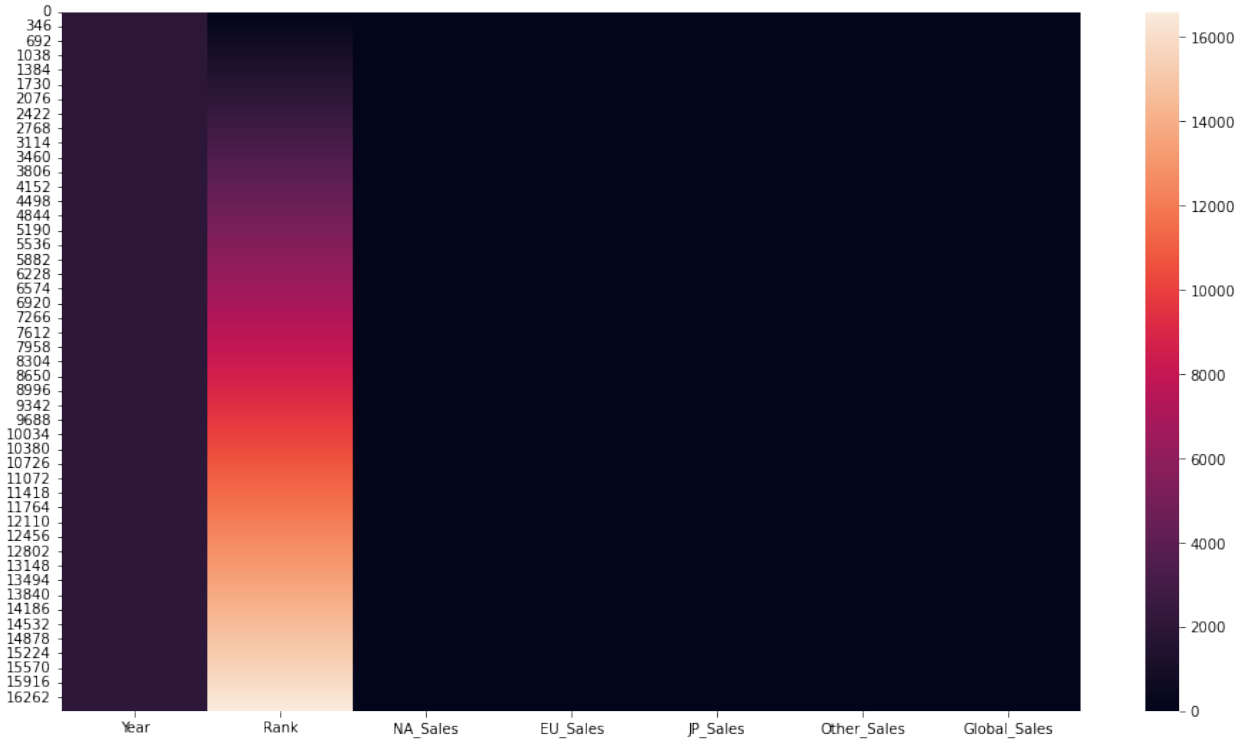
```
Year            0
Rank            0
NA_Sales        0
EU_Sales        0
JP_Sales        0
Other_Sales     0
Global_Sales    0
dtype: int64
```

```
reprised_dataforheatmap.isnull().sum().sum()
```

```
0
```

```
plt.figure(figsize=(16,9))
sns.heatmap(reprised_dataforheatmap)
```

```
<AxesSubplot:>
```




```
plt.figure(figsize=(16,9))
sns.heatmap(reprised_dataforheatmap.isnull())
```

```
<AxesSubplot:>
```

```
import pandas as pd
from sklearn.preprocessing import LabelEncoder

char_data = new_data.select_dtypes(include = ['object'])
char_data
```

```
                                                    Name Platform  \
0                                             Wii Sports      Wii
1                                      Super Mario Bros.      NES
2                                         Mario Kart Wii      Wii
3                                      Wii Sports Resort      Wii
4                           Pokemon Red/Pokemon Blue       GB
...                                                  ...      ...
16593                Woody Woodpecker in Crazy Castle 5      GBA
16594                     Men in Black II: Alien Escape       GC
16595  SCORE International Baja 1000: The Official Game      PS2
16596                                           Know How 2       DS
16597                                    Spirits & Spells      GBA

               Genre   Publisher
0             Sports    Nintendo
1           Platform    Nintendo
2             Racing    Nintendo
3             Sports    Nintendo
4      Role-Playing    Nintendo
...              ...         ...
16593       Platform       Kemco
16594        Shooter  Infogrames
16595         Racing  Activision
16596         Puzzle    7G//AMES
16597       Platform     Wanadoo

[16598 rows x 4 columns]
```

```
le = LabelEncoder()

le.fit_transform(char_data[['Name']])
```

```
C:\ProgramData\Anaconda3\lib\site-packages\sklearn\utils\
validation.py:63: DataConversionWarning: A column-vector y was passed
when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
  return f(*args, **kwargs)

array([11007,  9327,  5573, ...,  8144,  5014,  8967])
```

```
char_data['Name_L_enc'] = le.fit_transform(char_data[['Name']])
char_data
```

```
C:\Users\Ashish\AppData\Local\Temp/ipykernel_17260/2079822923.py:1:
SettingWithCopyWarning:
```

```
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation:
https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#
returning-a-view-versus-a-copy
  char_data['Name_L_enc'] = le.fit_transform(char_data[['Name']])
```

	Name	Platform	\
0	Wii Sports	Wii	
1	Super Mario Bros.	NES	
2	Mario Kart Wii	Wii	
3	Wii Sports Resort	Wii	
4	Pokemon Red/Pokemon Blue	GB	
...	...	...	
16593	Woody Woodpecker in Crazy Castle 5	GBA	
16594	Men in Black II: Alien Escape	GC	
16595	SCORE International Baja 1000: The Official Game	PS2	
16596	Know How 2	DS	
16597	Spirits & Spells	GBA	

	Genre	Publisher	Name_L_enc
0	Sports	Nintendo	11007
1	Platform	Nintendo	9327
2	Racing	Nintendo	5573
3	Sports	Nintendo	11009
4	Role-Playing	Nintendo	7346
...	...	...	...
16593	Platform	Kemco	11101
16594	Shooter	Infogrames	5796
16595	Racing	Activision	8144
16596	Puzzle	7G//AMES	5014
16597	Platform	Wanadoo	8967

```
[16598 rows x 5 columns]

le.fit_transform(char_data[['Platform']])

C:\ProgramData\Anaconda3\lib\site-packages\sklearn\utils\
validation.py:63: DataConversionWarning: A column-vector y was passed
when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
  return f(*args, **kwargs)

array([26, 11, 26, ..., 16,  4,  6])

char_data['Platform_L_enc'] =
le.fit_transform(char_data[['Platform']])
char_data
```

```
C:\Users\Ashish\AppData\Local\Temp/ipykernel_17260/2887531553.py:1:
SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation:
https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#
returning-a-view-versus-a-copy
  char_data['Platform_L_enc'] =
le.fit_transform(char_data[['Platform']])
```

	Name	Platform
0	Wii Sports	Wii
1	Super Mario Bros.	NES
2	Mario Kart Wii	Wii
3	Wii Sports Resort	Wii
4	Pokemon Red/Pokemon Blue	GB
...	...	...
16593	Woody Woodpecker in Crazy Castle 5	GBA
16594	Men in Black II: Alien Escape	GC
16595	SCORE International Baja 1000: The Official Game	PS2
16596	Know How 2	DS
16597	Spirits & Spells	GBA

	Genre	Publisher	Name_L_enc	Platform_L_enc
0	Sports	Nintendo	11007	26
1	Platform	Nintendo	9327	11
2	Racing	Nintendo	5573	26
3	Sports	Nintendo	11009	26
4	Role-Playing	Nintendo	7346	5
...	...	...	...	...
16593	Platform	Kemco	11101	6
16594	Shooter	Infogrames	5796	7
16595	Racing	Activision	8144	16
16596	Puzzle	7G//AMES	5014	4
16597	Platform	Wanadoo	8967	6

```
[16598 rows x 6 columns]
```

```
le.fit_transform(char_data[['Genre']])
```

```
C:\ProgramData\Anaconda3\lib\site-packages\sklearn\utils\
validation.py:63: DataConversionWarning: A column-vector y was passed
when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
  return f(*args, **kwargs)
```

```
array([10,  4,  6, ...,  6,  5,  4])
```

```
char_data['Genre_L_enc'] = le.fit_transform(char_data[['Genre']])
char_data
```

```
C:\Users\Ashish\AppData\Local\Temp/ipykernel_17260/3108553845.py:1:
SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation:
https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#
returning-a-view-versus-a-copy
  char_data['Genre_L_enc'] = le.fit_transform(char_data[['Genre']])

                                                  Name Platform  \
0                                           Wii Sports      Wii
1                                    Super Mario Bros.      NES
2                                       Mario Kart Wii      Wii
3                                    Wii Sports Resort      Wii
4                              Pokemon Red/Pokemon Blue       GB
...                                                 ...      ...
16593                Woody Woodpecker in Crazy Castle 5      GBA
16594                     Men in Black II: Alien Escape       GC
16595  SCORE International Baja 1000: The Official Game      PS2
16596                                       Know How 2       DS
16597                                 Spirits & Spells      GBA

               Genre   Publisher  Name_L_enc  Platform_L_enc
Genre_L_enc
0             Sports    Nintendo       11007              26
10
1           Platform    Nintendo        9327              11
4
2             Racing    Nintendo        5573              26
6
3             Sports    Nintendo       11009              26
10
4       Role-Playing    Nintendo        7346               5
7
...              ...         ...         ...             ...          .
..
16593       Platform       Kemco       11101               6
4
16594        Shooter  Infogrames        5796               7
8
16595         Racing  Activision        8144              16
6
16596         Puzzle    7G//AMES        5014               4
5
16597       Platform     Wanadoo        8967               6
4

[16598 rows x 7 columns]
```

```
le.fit_transform(char_data[['Publisher']])
```

```
C:\ProgramData\Anaconda3\lib\site-packages\sklearn\utils\
validation.py:63: DataConversionWarning: A column-vector y was passed
when a 1d array was expected. Please change the shape of y to
(n_samples, ), for example using ravel().
  return f(*args, **kwargs)
```

```
array([359, 359, 359, ...,  21,   8, 546])
```

```
char_data['Publisher_L_enc'] =
le.fit_transform(char_data[['Publisher']])
char_data
```

```
C:\Users\Ashish\AppData\Local\Temp/ipykernel_17260/3201290837.py:1:
SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation:
https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#
returning-a-view-versus-a-copy
  char_data['Publisher_L_enc'] =
le.fit_transform(char_data[['Publisher']])
```

```
                                                    Name Platform  \
0                                             Wii Sports      Wii
1                                      Super Mario Bros.      NES
2                                         Mario Kart Wii      Wii
3                                      Wii Sports Resort      Wii
4                               Pokemon Red/Pokemon Blue       GB
...                                                  ...      ...
16593                Woody Woodpecker in Crazy Castle 5      GBA
16594                     Men in Black II: Alien Escape       GC
16595  SCORE International Baja 1000: The Official Game      PS2
16596                                       Know How 2       DS
16597                                Spirits & Spells      GBA

               Genre   Publisher  Name_L_enc  Platform_L_enc
Genre_L_enc  \
0             Sports    Nintendo       11007              26
10
1           Platform    Nintendo        9327              11
4
2             Racing    Nintendo        5573              26
6
3             Sports    Nintendo       11009              26
10
4       Role-Playing    Nintendo        7346               5
7
...              ...         ...         ...             ...          .
```

```
..
16593      Platform       Kemco       11101                6
4
16594       Shooter  Infogrames        5796                7
8
16595        Racing  Activision        8144               16
6
16596        Puzzle    7G//AMES        5014                4
5
16597      Platform     Wanadoo        8967                6
4

       Publisher_L_enc
0                  359
1                  359
2                  359
3                  359
4                  359
...                ...
16593              269
16594              241
16595               21
16596                8
16597              546

[16598 rows x 8 columns]

char_data.isnull()

        Name  Platform  Genre  Publisher  Name_L_enc
Platform_L_enc  \
0      False     False  False      False       False           False

1      False     False  False      False       False           False

2      False     False  False      False       False           False

3      False     False  False      False       False           False

4      False     False  False      False       False           False

...      ...       ...    ...        ...         ...             ...

16593  False     False  False      False       False           False

16594  False     False  False      False       False           False

16595  False     False  False      False       False           False

16596  False     False  False      False       False           False
```

```
16597  False     False  False      False       False           False


       Genre_L_enc  Publisher_L_enc
0            False            False
1            False            False
2            False            False
3            False            False
4            False            False
...            ...              ...
16593        False            False
16594        False            False
16595        False            False
16596        False            False
16597        False            False

[16598 rows x 8 columns]
```

```
char_data.isnull().sum()
```

```
Name                 0
Platform             0
Genre                0
Publisher           58
Name_L_enc           0
Platform_L_enc       0
Genre_L_enc          0
Publisher_L_enc      0
dtype: int64
```

```
char_data.isnull().sum().sum()
```

```
58
```

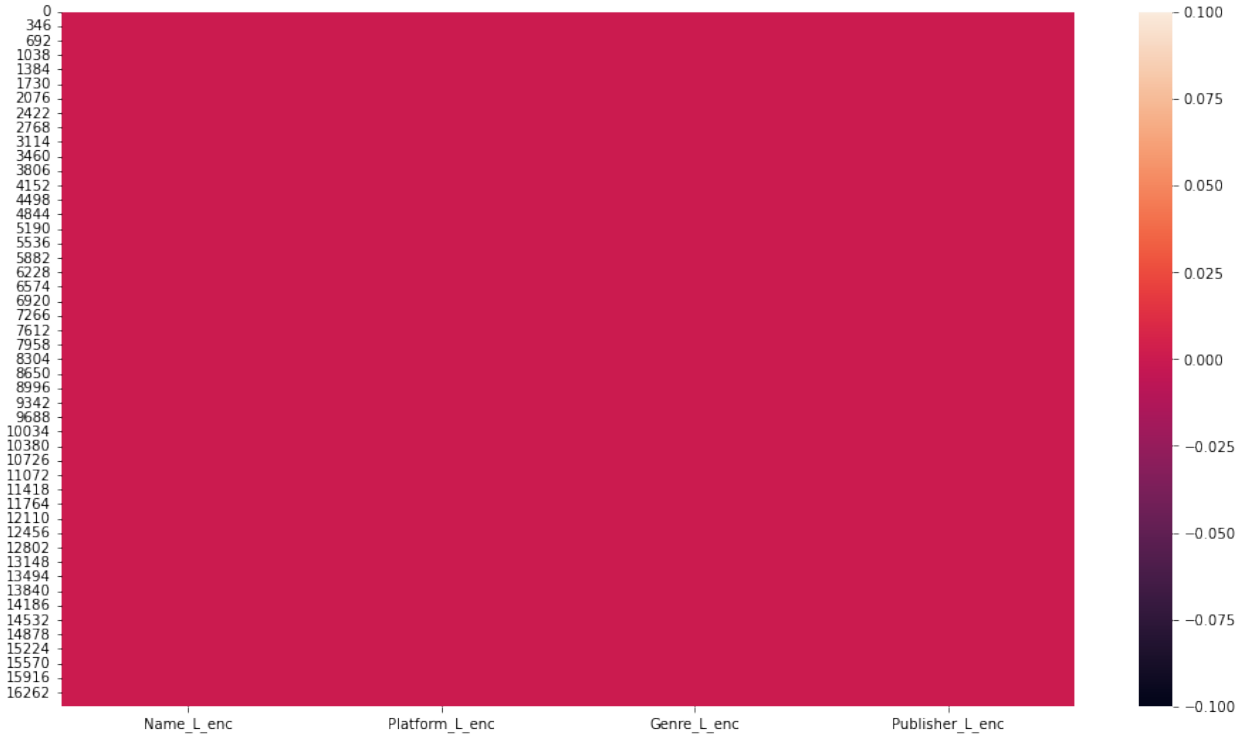
```
char_data_to_numerical = pd.concat((char_data['Name_L_enc'],
char_data['Platform_L_enc'], char_data['Genre_L_enc'],
                                    char_data['Publisher_L_enc']),
axis=1)
char_data_to_numerical
```

```
       Name_L_enc  Platform_L_enc  Genre_L_enc  Publisher_L_enc
0           11007              26           10              359
1            9327              11            4              359
2            5573              26            6              359
3           11009              26           10              359
4            7346               5            7              359
...           ...             ...          ...              ...
16593       11101               6            4              269
16594        5796               7            8              241
16595        8144              16            6               21
16596        5014               4            5                8
```

```
16597        8967                  6              4                  546

[16598 rows x 4 columns]
```

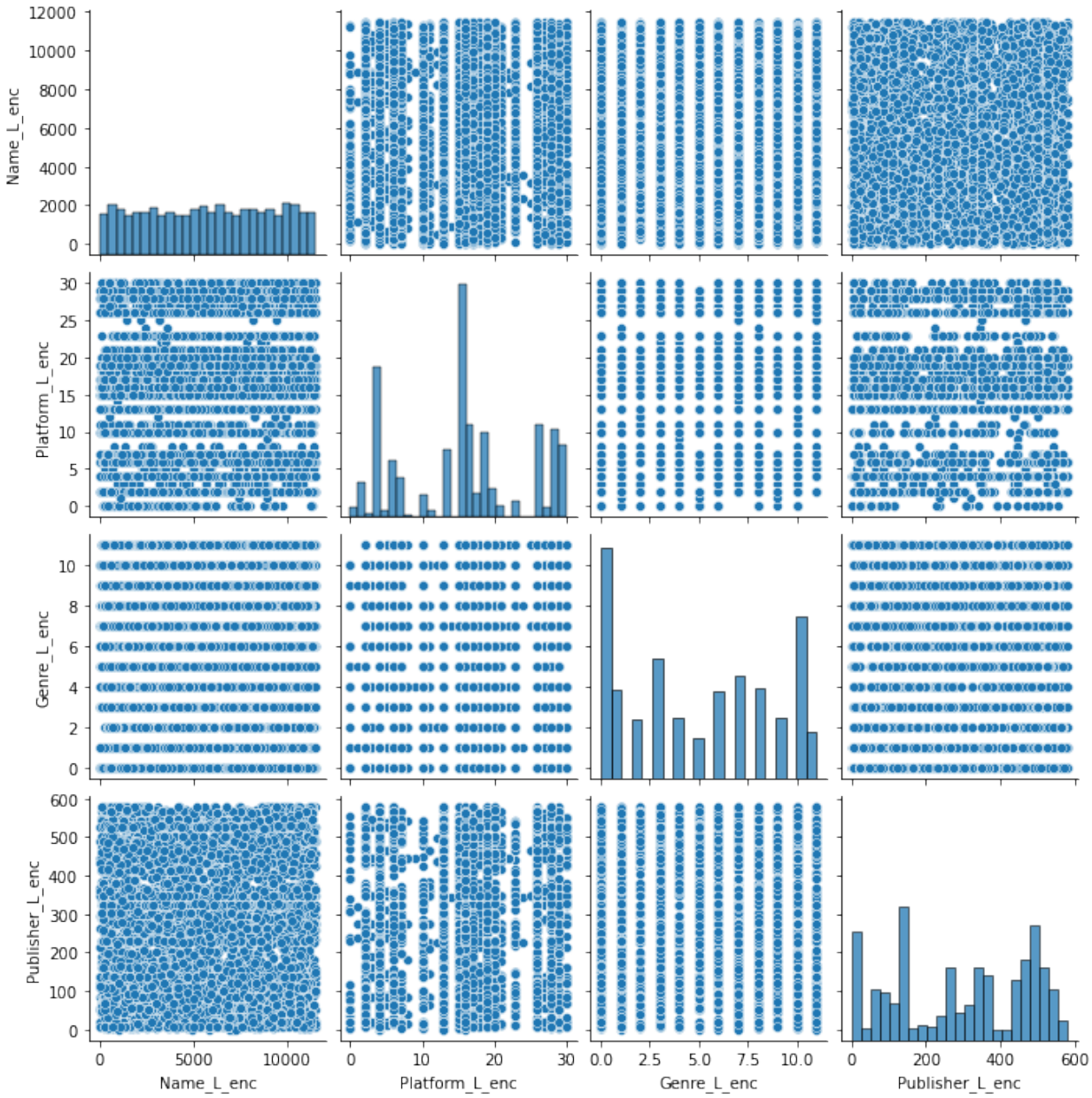
```
plt.figure(figsize = (16,9))
sns.heatmap(char_data_to_numerical.isnull())
```

```
<AxesSubplot:>
```



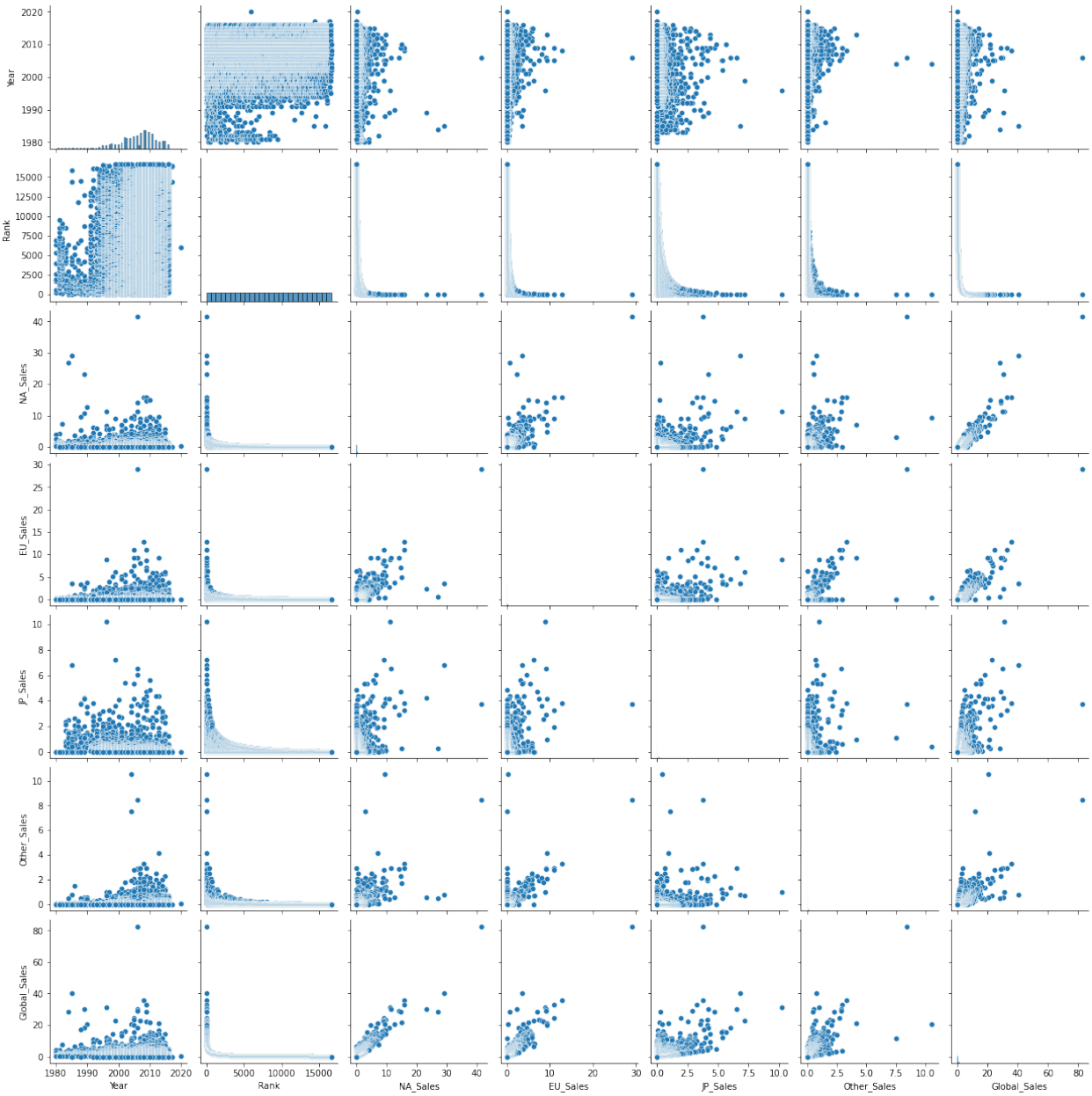

```
sns.pairplot(char_data_to_numerical)
```

```
<seaborn.axisgrid.PairGrid at 0x1fb0d4b2f70>
```




```
sns.pairplot(reprised_dataforheatmap)
```

```
<seaborn.axisgrid.PairGrid at 0x1fb1294e760>
```


Year
Rank
NA_Sales
EU_Sales
JP_Sales
Other_Sales
Global_Sales