```
# This Python 3 environment comes with many helpful analytics
libraries installed
import numpy as np # linear algebra
import pandas as pd # data processing

pd.set_option('display.max_columns', None)  # allows to display all
the columns

# Read the dataset
video_games_sales=pd.read_csv("C:\Areefa_Brunel\
C5703_Data_Visualisation\DV_Coursework_Submission\
Video_Games_Sales_as_at_22_Dec_2016.csv")

# To print first 5 values
print(video_games_sales.head(5))

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

   NA_Sales  EU_Sales  JP_Sales  Other_Sales  Global_Sales
Critic_Score  \
0     41.36     28.96      3.77         8.45         82.53
76.0
1     29.08      3.58      6.81         0.77         40.24
NaN
2     15.68     12.76      3.79         3.29         35.52
82.0
3     15.61     10.93      3.28         2.95         32.77
80.0
4     11.27      8.89     10.22         1.00         31.37
NaN

   Critic_Count User_Score  User_Count Developer Rating
0          51.0          8       322.0  Nintendo      E
1           NaN        NaN         NaN       NaN    NaN
2          73.0        8.3       709.0  Nintendo      E
3          73.0          8       192.0  Nintendo      E
4           NaN        NaN         NaN       NaN    NaN

#lets check if any missing data present in the dataset
#we can do this by checking individual values in dataset
```

```
print(video_games_sales.isna().sum())
#isna method is to check each individual value for missingness
```

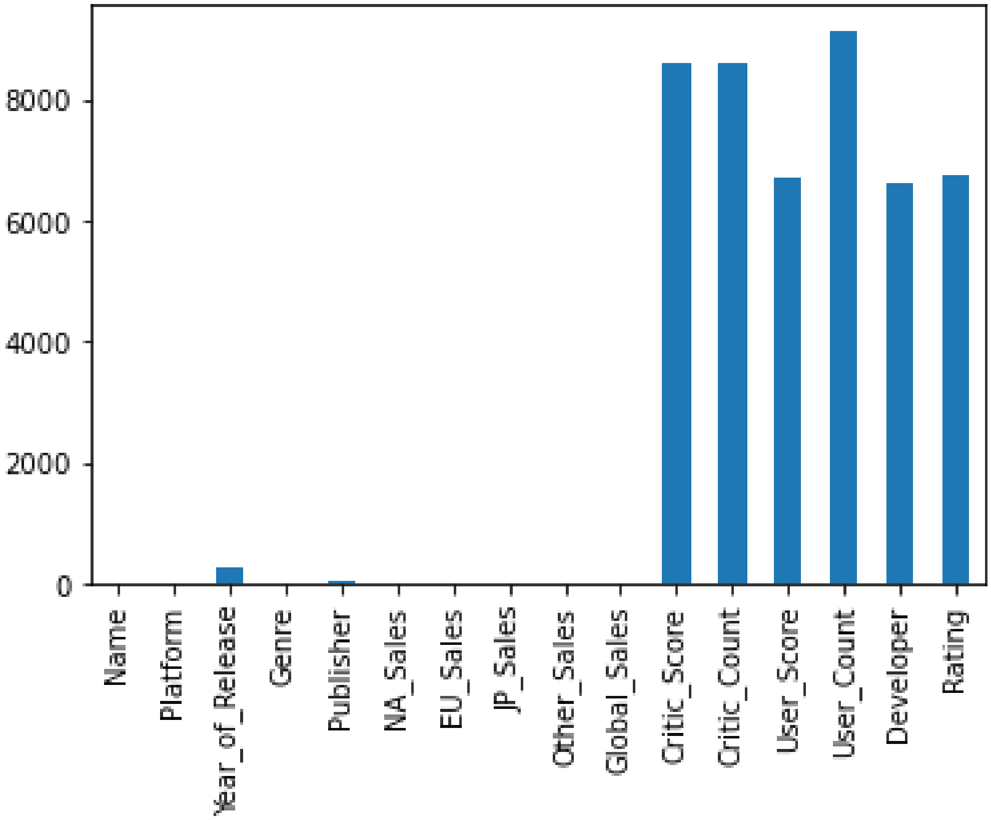
```
Name                  2
Platform              0
Year_of_Release     269
Genre                 2
Publisher            54
NA_Sales              0
EU_Sales              0
JP_Sales              0
Other_Sales           0
Global_Sales          0
Critic_Score       8582
Critic_Count       8582
User_Score         6704
User_Count         9129
Developer          6623
Rating             6769
dtype: int64
```

```
#"false" means no missing values, true indicates a missing values
# We can also print a summary to show if any value in each column is
missing or not
print(video_games_sales.isna().any())
#we add the any method to the previous code
```

```
Name                True
Platform           False
Year_of_Release     True
Genre               True
Publisher           True
NA_Sales           False
EU_Sales           False
JP_Sales           False
Other_Sales        False
Global_Sales       False
Critic_Score        True
Critic_Count        True
User_Score          True
User_Count          True
Developer           True
Rating              True
dtype: bool
```

```
#columns with missing values returns true and those without missing
values returns false
#we can also create a bar plot of the total number of missing values
in each column
import matplotlib.pyplot as plt
```

```
video_games_sales.isna().sum().plot(kind='bar')
plt.show()
```



```
#from the plot the last columns has so many missing values because Metacritics only covers a subset of the platforms.
# Also, a game may not have all the the fields recorded
video_games_cleaned=video_games_sales.dropna()


#Verify if we have dropped missing values:
print(video_games_cleaned.isna().any())
#we have successfully dropped rows with missing values.
```

```
Name               False
Platform           False
Year_of_Release    False
Genre              False
Publisher          False
NA_Sales           False
EU_Sales           False
JP_Sales           False
Other_Sales        False
Global_Sales       False
Critic_Score       False
Critic_Count       False
```

```
User_Score         False
User_Count         False
Developer          False
Rating             False
dtype: bool
```

```
# checking the information of the data
video_games_cleaned.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 6825 entries, 0 to 16706
Data columns (total 16 columns):
 #   Column           Non-Null Count  Dtype  
---  ------           --------------  -----  
 0   Name             6825 non-null   object 
 1   Platform         6825 non-null   object 
 2   Year_of_Release  6825 non-null   float64
 3   Genre            6825 non-null   object 
 4   Publisher        6825 non-null   object 
 5   NA_Sales         6825 non-null   float64
 6   EU_Sales         6825 non-null   float64
 7   JP_Sales         6825 non-null   float64
 8   Other_Sales      6825 non-null   float64
 9   Global_Sales     6825 non-null   float64
 10  Critic_Score     6825 non-null   float64
 11  Critic_Count     6825 non-null   float64
 12  User_Score       6825 non-null   object 
 13  User_Count       6825 non-null   float64
 14  Developer        6825 non-null   object 
 15  Rating           6825 non-null   object 
dtypes: float64(9), object(7)
memory usage: 906.4+ KB
```

```
# To save this file
video_games_cleaned.to_csv('video_games_cleaned.csv')
```