In [1]: **import** numpy **as** np #importing pandas library as pd import pandas as pd #reading the excel sheet to python dataset = pd.read\_excel("Movies\_Database.xlsx") #showing the content in dataset dataset #showing first 5 rows of the dataset dataset.head() #check the tail: dataset.tail(5) #showing the column header dataset.columns dataset.index dataset.values type(dataset) dataset.shape (14, 16)Out[1]: In [2]: dataset.info() <class 'pandas.core.frame.DataFrame'> RangeIndex: 14 entries, 0 to 13 Data columns (total 16 columns): Column Non-Null Count Dtype -----0 movie\_title 14 non-null object num\_critic\_for\_reviews 14 non-null int64 1 11 non-null duration float64 12 non-null 3 DIRECTOR\_facebook\_likes object actor\_3\_facebook\_likes 4 14 non-null int64 ACTOR\_1\_facebook\_likes 14 non-null int64 6 gross 14 non-null int64 num\_voted\_users 13 non-null float64 Cast\_Total\_facebook\_likes 12 non-null float64 facenumber\_in\_poster 9 non-null float64 10 num\_user\_for\_reviews 14 non-null int64 14 non-null 11 budget int64 12 title\_year 14 non-null int64 float64 13 ACTOR\_2\_facebook\_likes 13 non-null 14 imdb\_score 14 non-null float64 float64 15 title\_year.1 7 non-null dtypes: float64(7), int64(7), object(2)memory usage: 1.9+ KB In [3]: dataset['movie\_title'] Avatar?ÿ Out[3]: Pirates of the Caribbean: At World's End?ÿ 1 2 Spectre?ÿ 3 The Dark Knight Rises?ÿ John Carter?ÿ Spider-Man 3?ÿ Tangled?ÿ 6 Avengers: Age of Ultron?ÿ 8 Avengers: Age of Ultron?ÿ Harry Potter and the Half-Blood Prince?ÿ 9 10 Batman v Superman: Dawn of Justice?ÿ 11 Superman Returns?ÿ 12 Quantum of Solace?ÿ 13 Pirates of the Caribbean: Dead Man's Chest?ÿ Name: movie\_title, dtype: object In [4]: | dataset['movie\_title'] = dataset['movie\_title'].str.replace(r'[?ÿ]',"") print(dataset) movie\_title num\_critic\_for\_reviews \ 723 Avatar Pirates of the Caribbean: At World's End 302 2 602 Spectre 3 The Dark Knight Rises 813 John Carter 462 5 Spider-Man 3 392 6 Tangled 324 Avengers: Age of Ultron 635 8 Avengers: Age of Ultron 635 9 Harry Potter and the Half-Blood Prince 375 10 Batman v Superman: Dawn of Justice 673 11 Superman Returns 434 12 Quantum of Solace 403 13 Pirates of the Caribbean: Dead Man's Chest 313 duration DIRECTOR\_facebook\_likes actor\_3\_facebook\_likes \ 0 178.0 10 1000 1 NaN 563 2 148.0 20 161 3 22000 23000 NaN "475" 4 132.0 530 5 156.0 23 4000 6 NaN 15 284 141.0 10 19000 8 141.0 10 19000 9 153.0 282 10000 10 183.0 NaN 2000 11 169.0 NaN 903 12 106.0 395 393 13 1000 151.0 563 ACTOR\_1\_facebook\_likes num\_voted\_users \ gross 0 1000 760505847 886204.0 1 40000 309404152 471220.0 2 11000 200074175 275868.0 3 448130642 1144337.0 27000 640 73058679 212204.0 5 24000 336530303 383056.0 6 799 200807262 294810.0 26000 458991599 462669.0 8 26000 458991599 462669.0 321795.0 9 25000 301956980 10 15000 330249062 NaN 240396.0 11 18000 200069408 12 451 168368427 330784.0 13 40000 423032628 522040.0 num\_user\_for\_reviews \ Cast\_Total\_facebook\_likes facenumber\_in\_poster 0 NaN 4834.0 1 48350.0 NaN 1238 2 11700.0 1.0 994 3 106759.0 NaN 2701 4 1873.0 1.0 738 46055.0 NaN 1902 NaN 1.0 387 92000.0 1117 8 92000.0 4.0 1117 9 58753.0 3.0 973 10 24450.0 NaN 3018 11 NaN 2.0 2367 12 2023.0 1.0 1243 13 48486.0 1832 2.0 ACTOR\_2\_facebook\_likes imdb\_score title\_year.1 budget title\_year 0 237000000 2009 936.0 7.9 2009.0 30000000 2007 5000.0 7.1 NaN 1 2 2015 393.0 245000000 6.8 2015.0 2012 3 250000000 23000.0 8.5 NaN 4 263700000 2012 632.0 6.6 NaN 258000000 5 2007 11000.0 6.2 2007.0 6 260000000 2010 553.0 7.8 NaN 250000000 2015 21000.0 7.5 NaN 8 250000000 2015 21000.0 7.5 2015.0 250000000 2009 9 11000.0 7.5 NaN 10 250000000 2016 NaN 6.9 2016.0 11 209000000 2006 10000.0 6.1 NaN 200000000 2008 12 412.0 6.7 2008.0 2006 5000.0 7.3 2008.0 13 225000000 C:\Users\ssaji\AppData\Local\Temp\ipykernel\_16548\3964326471.py:1: FutureWarning: The default value of regex will change from True to False in a future version. dataset['movie\_title'] = dataset['movie\_title'].str.replace(r'[?ÿ]',"") dataset In [5]: movie\_title num\_critic\_for\_reviews duration DIRECTOR\_facebook\_likes actor\_3\_facebook\_likes ACTOR\_1\_facebook\_likes gross num\_voted\_users Cast\_Total\_facebook\_likes facenumber\_in\_poster num\_ Out[5]: 723 178.0 10 855 1000 760505847 886204.0 4834.0 Pirates of the 40000 309404152 1 Caribbean: 302 NaN 563 1000 471220.0 48350.0 NaN At World's End 11000 200074175 Spectre 602 148.0 20 161 275868.0 11700.0 1.0 The Dark Knight 813 NaN 22000 23000 27000 448130642 1144337.0 106759.0 NaN Rises 4 John Carter 462 132.0 "475" 530 73058679 212204.0 1873.0 1.0 Spider-Man 4000 336530303 392 156.0 23 24000 383056.0 46055.0 NaN 324 NaN 15 284 799 200807262 294810.0 NaN 1.0 Tangled Avengers: 141.0 10 19000 26000 458991599 462669.0 92000.0 4.0 635 Age of Avengers: 10 26000 458991599 Age of 635 141.0 19000 462669.0 92000.0 4.0 Ultron Harry Potter and the Half-375 153.0 282 10000 25000 301956980 321795.0 58753.0 3.0 Blood Prince Batman v Superman: 10 673 183.0 NaN 2000 15000 330249062 NaN 24450.0 NaN Dawn of Justice Superman 11 169.0 18000 200069408 240396.0 2.0 434 NaN 903 NaN Returns Quantum of 12 106.0 395 393 330784.0 2023.0 1.0 403 451 168368427 Pirates of the Caribbean: 151.0 563 1000 40000 423032628 522040.0 48486.0 2.0 313 Dead Man's Chest In [6]: dataset.info() <class 'pandas.core.frame.DataFrame'> RangeIndex: 14 entries, 0 to 13 Data columns (total 16 columns): Column Non-Null Count Dtype 0 movie\_title 14 non-null object 14 non-null num\_critic\_for\_reviews int64 1 11 non-null 2 duration float64 3 DIRECTOR\_facebook\_likes 12 non-null object 14 non-null actor\_3\_facebook\_likes int64 14 non-null 5 ACTOR\_1\_facebook\_likes int64 6 gross 14 non-null int64 num\_voted\_users 13 non-null float64 Cast\_Total\_facebook\_likes 12 non-null float64 facenumber\_in\_poster 9 non-null float64 10 num\_user\_for\_reviews 14 non-null int64 11 budget 14 non-null int64 12 14 non-null int64 title\_year 13 non-null float64 13 ACTOR\_2\_facebook\_likes float64 14 imdb\_score 14 non-null 15 title\_year.1 7 non-null float64 dtypes: float64(7), int64(7), object(2)memory usage: 1.9+ KB dataset.shape (14, 16)Out[7] pd.isnull(np.nan) Out[8]: pd.isnull(dataset) In [9]: movie\_title num\_critic\_for\_reviews duration DIRECTOR\_facebook\_likes actor\_3\_facebook\_likes ACTOR\_1\_facebook\_likes gross num\_voted\_users Cast\_Total\_facebook\_likes facenumber\_in\_poster num\_user\_ Out[9]: False False False False False False False False False True 1 False False True False False False False False False True 2 False 3 False False True False False False False False False True 4 False True 6 False False True False False False False False True False 7 False 8 False 9 False 10 False False False True False False False True False True 11 False False False True False False False False True False 12 False 13 False dataset.isnull().sum() 0 movie\_title Out[10] 0 num\_critic\_for\_reviews 3 duration DIRECTOR\_facebook\_likes 2 actor\_3\_facebook\_likes ACTOR\_1\_facebook\_likes gross num\_voted\_users 1 Cast\_Total\_facebook\_likes facenumber\_in\_poster num\_user\_for\_reviews budget title\_year 0 ACTOR\_2\_facebook\_likes 1 imdb\_score 0 title\_year.1 dtype: int64 In [11]: #filling notnull values with 0 dataset.fillna(0) movie\_title num\_critic\_for\_reviews duration DIRECTOR\_facebook\_likes actor\_3\_facebook\_likes ACTOR\_1\_facebook\_likes gross num\_voted\_users Cast\_Total\_facebook\_likes facenumber\_in\_poster num\_ Out[11]: 723 178.0 10 855 1000 760505847 886204.0 4834.0 0.0 Avatar Pirates of the 1 Caribbean: 1000 48350.0 302 0.0 563 40000 309404152 471220.0 0.0 At World's End Spectre 602 148.0 20 161 11000 200074175 275868.0 11700.0 1.0 The Dark 1144337.0 0.0 22000 23000 27000 448130642 106759.0 0.0 Knight 813 Rises 73058679 4 John Carter 462 132.0 "475" 530 212204.0 1873.0 1.0 Spider-Man 23 4000 336530303 383056.0 0.0 392 156.0 24000 46055.0 0.0 1.0 Tangled 324 0.0 15 284 799 200807262 294810.0 Avengers: 92000.0 635 141.0 10 19000 26000 458991599 462669.0 4.0 Age of Avengers: 635 141.0 10 19000 26000 458991599 462669.0 92000.0 4.0 Age of Ultron Harry Potter and 282 375 153.0 10000 25000 301956980 321795.0 58753.0 3.0 the Half-Blood Prince Batman v Superman: 0 10 183.0 2000 15000 330249062 0.0 24450.0 0.0 673 Dawn of Justice Superman 18000 200069408 11 434 169.0 0 903 240396.0 0.0 2.0 Returns Quantum of 403 106.0 395 393 451 168368427 330784.0 2023.0 1.0 Pirates of the Caribbean: 522040.0 563 1000 40000 423032628 48486.0 2.0 313 151.0 Dead Man's Chest In [12]: dataset.info() <class 'pandas.core.frame.DataFrame'> RangeIndex: 14 entries, 0 to 13 Data columns (total 16 columns): Column Non-Null Count Dtype 0 movie\_title 14 non-null object 14 non-null num\_critic\_for\_reviews int64 1 11 non-null 2 float64 duration DIRECTOR\_facebook\_likes 12 non-null 3 object actor\_3\_facebook\_likes 14 non-null int64 14 non-null 5 ACTOR\_1\_facebook\_likes int64 14 non-null 6 gross int64 num\_voted\_users 13 non-null float64 Cast\_Total\_facebook\_likes 12 non-null float64 9 non-null facenumber\_in\_poster float64 10 num\_user\_for\_reviews 14 non-null int64 11 budget 14 non-null int64 12 title\_year 14 non-null int64 13 non-null 13 ACTOR\_2\_facebook\_likes float64 imdb\_score float64 14 14 non-null float64 15 title\_year.1 7 non-null dtypes: float64(7), int64(7), object(2)memory usage: 1.9+ KB In [13]: dataset['DIRECTOR\_facebook\_likes'][4] =int(dataset['DIRECTOR\_facebook\_likes'][4].replace('"','')) C:\Users\ssaji\AppData\Local\Temp\ipykernel\_16548\3120792715.py:1: SettingWithCopyWarning: A value is trying to be set on a copy of a slice from a DataFrame See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user\_guide/indexing.html#returning-a-view-versus-a-copy dataset['DIRECTOR\_facebook\_likes'][4] =int(dataset['DIRECTOR\_facebook\_likes'][4].replace('"','')) type(dataset['DIRECTOR\_facebook\_likes'][1]) In [14]: Out[14]: dataset.fillna(0) In [15]: movie\_title num\_critic\_for\_reviews duration DIRECTOR\_facebook\_likes actor\_3\_facebook\_likes ACTOR\_1\_facebook\_likes gross num\_voted\_users Cast\_Total\_facebook\_likes facenumber\_in\_poster num\_ Out[15]: 723 178.0 4834.0 10 855 1000 760505847 886204.0 0.0 Avatar Pirates of 1 Caribbean: 302 0.0 563 1000 40000 309404152 471220.0 48350.0 0.0 At World's 148.0 20 11000 200074175 11700.0 1.0 2 Spectre 602 161 275868.0 The Dark 22000 0.0 813 0.0 23000 27000 448130642 1144337.0 106759.0 3 Knight Rises 4 John Carter 132.0 475 73058679 212204.0 1873.0 1.0 462 530 640 392 156.0 23 4000 24000 336530303 383056.0 46055.0 0.0 Tangled 324 0.0 15 284 799 200807262 294810.0 0.0 1.0 Avengers: 635 141.0 10 19000 26000 458991599 462669.0 92000.0 4.0 Age of Ultron Avengers: 10 19000 26000 458991599 92000.0 4.0 635 141.0 462669.0 Age of Ultron Harry Potter and 153.0 282 10000 25000 301956980 321795.0 58753.0 3.0 the Half-375 Blood Prince Batman v Superman: 183.0 0 2000 15000 330249062 24450.0 0.0 673 0.0 Dawn of Justice Superman 11 169.0 0 903 18000 200069408 240396.0 0.0 2.0 434 Returns Quantum of 12 403 106.0 395 393 451 168368427 330784.0 2023.0 1.0 Solace Pirates of the Caribbean: 13 1000 2.0 151.0 563 40000 423032628 522040.0 48486.0 313 Dead Man's Chest In [16]: dataset.info() <class 'pandas.core.frame.DataFrame'> RangeIndex: 14 entries, 0 to 13 Data columns (total 16 columns): Non-Null Count Dtype Column ----------0 movie\_title 14 non-null object num\_critic\_for\_reviews 14 non-null int64 duration 11 non-null float64 DIRECTOR\_facebook\_likes 12 non-null object actor\_3\_facebook\_likes 14 non-null int64 ACTOR\_1\_facebook\_likes 14 non-null int64 gross 14 non-null int64 num\_voted\_users 13 non-null float64 Cast\_Total\_facebook\_likes 12 non-null float64 facenumber\_in\_poster 9 non-null float64 14 non-null num\_user\_for\_reviews int64 10 budget 14 non-null int64 11 14 non-null int64 12 title\_year ACTOR\_2\_facebook\_likes 13 non-null float64 13 float64 14 imdb\_score 14 non-null title\_year.1 float64 15 7 non-null dtypes: float64(7), int64(7), object(2) dataset['DIRECTOR\_facebook\_likes'] = dataset['DIRECTOR\_facebook\_likes'].fillna(0).astype(int) In [17]: In [18]: dataset.info() <class 'pandas.core.frame.DataFrame'> RangeIndex: 14 entries, 0 to 13 Data columns (total 16 columns): Column Non-Null Count Dtype movie\_title 14 non-null object num\_critic\_for\_reviews 14 non-null int64 float64 11 non-null duration DIRECTOR\_facebook\_likes 14 non-null int32 actor\_3\_facebook\_likes 14 non-null int64 14 non-null ACTOR\_1\_facebook\_likes int64 14 non-null gross int64 13 non-null float64 num\_voted\_users Cast\_Total\_facebook\_likes 12 non-null float64 9 non-null float64 facenumber\_in\_poster 14 non-null num\_user\_for\_reviews int64 10 11 budget 14 non-null int64 12 title\_year 14 non-null int64 ACTOR\_2\_facebook\_likes 13 non-null float64 13 imdb\_score 14 non-null float64 title\_year.1 7 non-null float64 dtypes: float64(7), int32(1), int64(7), object(1) memory usage: 1.8+ KB dataset = dataset.fillna(0) In [19]: In [20]: dataset.info() <class 'pandas.core.frame.DataFrame'> RangeIndex: 14 entries, 0 to 13 Data columns (total 16 columns): Column Non-Null Count Dtype 0 movie\_title 14 non-null object 14 non-null num\_critic\_for\_reviews int64 duration 14 non-null float64 14 non-null DIRECTOR\_facebook\_likes int32 14 non-null actor\_3\_facebook\_likes int64 14 non-null ACTOR\_1\_facebook\_likes int64 6 gross 14 non-null int64 num\_voted\_users 14 non-null float64 14 non-null Cast\_Total\_facebook\_likes float64 14 non-null facenumber\_in\_poster float64 10 num\_user\_for\_reviews 14 non-null int64 11 budget 14 non-null int64 12 title\_year 14 non-null int64 13 ACTOR\_2\_facebook\_likes 14 non-null float64 14 imdb\_score 14 non-null float64 title\_year.1 14 non-null float64 dtypes: float64(7), int32(1), int64(7), object(1) memory usage: 1.8+ KB In [21]: dataset Out[21]: movie\_title num\_critic\_for\_reviews duration DIRECTOR\_facebook\_likes actor\_3\_facebook\_likes ACTOR\_1\_facebook\_likes gross num\_voted\_users Cast\_Total\_facebook\_likes facenumber\_in\_poster num\_ 178.0 4834.0 723 10 855 1000 760505847 886204.0 0.0 Avatar Pirates of 1 Caribbean: 302 0.0 563 1000 40000 309404152 471220.0 48350.0 0.0 At World's 11700.0 1.0 2 Spectre 602 148.0 20 161 11000 200074175 275868.0 The Dark 22000 23000 Knight 813 0.0 27000 448130642 1144337.0 106759.0 0.0 Rises 212204.0 4 John Carter 462 132.0 475 73058679 1873.0 530 640 1.0 Spider-Man 156.0 23 4000 24000 336530303 383056.0 46055.0 0.0 392 Tangled 324 0.0 15 284 799 200807262 294810.0 0.0 1.0 Avengers: 635 141.0 10 19000 26000 458991599 462669.0 92000.0 4.0 Age of Ultron Avengers: 141.0 10 19000 26000 458991599 462669.0 92000.0 4.0 Age of 635 Ultron Harry Potter and 153.0 282 10000 25000 301956980 58753.0 3.0 the Half-375 321795.0 Blood Prince Batman v Superman: 673 183.0 0 2000 15000 330249062 0.0 24450.0 0.0 Dawn of Justice Superman 11 434 169.0 903 18000 200069408 240396.0 0.0 2.0 Returns Quantum of 12 451 168368427 1.0 403 106.0 395 393 330784.0 2023.0 Solace Pirates of Caribbean: 1000 2.0 313 151.0 563 40000 423032628 522040.0 48486.0 Dead Man's Chest