REPORT

Hello there, I had a lot of fun doing this assignment. Here are my findings:

* Firstly, I understood how powerful Numpy actually is.

Numpy creates an ndarray (N Dimnesional Array) which supports homogenous data types. All sorts of functions and mathematical operations can be done on this ndarray. It also supports trigonometric functions, which can be very useful in wave analysis, signal processing using Fourier Transform.

One can also use these ndarray’s as matrices and perform matrix multiplication.

One can perform bitwise operation on these arrays like bitwise AND, OR, XOR, leftshift and rightshift.  
These arrays can be sorted with respect to one column or even multiple columns, as per our use.

Slicing is another powerful tool of numpy, as we can fetch the data from any interval. We can also fetch particular values by adding the necessary parameters.

These ndarrays can be reshaped and flattened too(i.e changing the structure of the array).

One of the most important feature of numpy is it’s Vectorized operations. It enables us to do array operations without the need of loops. Hence, making it faster, more efficient and hassle free.

* Pandas is a python library used for working with datasets. Pandas makes the data manipulation and analysis of large datasets very easy.

It creates data frames of different files.

One can perform various operations on this dataframe using the functions of Pandas. We can get all the information regarding the file using functions of this python library like .info() and .describe() .

We can filter columns and rows and print the data which we need using conditions.

We can also access the data using Indexing.

New columns can be created by performing calculations on the existing columns. Older columns and rows can be deleted.

One of the most important feature of Pandas is that it is built on top of the Numpy Library, hence by using vectorized operation on dataframes, we can get work done faster.

* Matplotlib is a python library built on Numpy, which is used to plot various forms of data on different types of graphs.

It’s a library through which we can represent huge data pictorially in a visually pleasing manner.

It helps to make our understanding of the data even deeper.

We can make line graphs, bar chart, histogram, scatter plot etc using matplotlib.

Seaborn is a visualization library built on top of matplotlib and is used to produce aesthetic graphs.

**INSIGHTS:**

* The given dataset is about different types of console games. It provides information on what console does the game run, what is the critic score of the game, the genre, the publisher, the developer of the game, the sales of the game in different areas. It also gives the release date of the game and when was the last update introduced in the game.
* From the histogram shown in the jupyter notebook, it can be inferred that as and as the value of total sales is increasing, the number of games or the frequency decreases.
* From the scatter plot shown in the jupyter notebook, it can be inferred that as the critic score increases, the total sales of games also increases. Hence, pointing out that the greater the critic score, the greater will be the sales.
* The mean value of Total Sales is 1.188, median is nan and the standard deviation is 1.844 .
* The mean value of critic score is 7.22

I did not face any such difficulties doing the assignment. The only challenge I faced was when I tried to gain insights on the release dates. I tried to compare it with a string, it flagged an error (comparison of int and str)

Then I tried to compare it using the datetime module. It again flagged an error.   
Then I checked the datatype of the column and it is termed as ‘object’. I tried to typecast it to str() and then I tried to use the strptime function of datetime module to convert string to datetime datatype.  
This was successful except for the last row, where there is no date given and the value is NaN, which is where the following error came up.



Then as a last resort I tried to delete the last row, using this method.

A black screen with green text

AI-generated content may be incorrect.

This too flagged an error of ‘Not found in axis’  
I searched up the meaning of this error

Apparently this error means that I’m trying to drop such columns which do not exist in the dataframe!!

Just when I was about to submit, it again clicked me that the index(i) here is counting the columns of 'missing in critic\_score' and 'total\_value' which I guess should be the problem.

And I'm stuck here