

Task 2 :

1. As of the First step I imported my necessary libraries and then uploaded and loaded the csv file
2. Then I see the structure of the data using data.shape to know about total number of columns and rows present
3. Then with the help of data.info(),it shows the data type of the each column like index,Rank are int64 type,Game Title,Platform,Genre,Publisher are object type,Year,North America,Europe,Japan,Rest of World,global,Review are float64 type
4. Then with data.describe(),it shows the each int64 and float64 types column's Statistical description like count,mean,std(Standard Deviation),min(Minimum),max(Maximum),percentile like 25%,50%,75% to know about how the columns vary between intervals
5. Then checked duplicate values and the missing values
 - data.isna() helps to find whether the dataset contains any null values and it like bool values(True or False)
 - data.isna().sum() helps to sum the total number of null values present in each columns in int64
 - So dataset got 29 null values in Year column and 2 null values in Publisher column
 - So we need to check whether the null values affect the other rows and columns ,If not then we can remove those rows or columns
 - But in our case it affect the other columns so we couldn't remove those columns and rows
 - So we can fill the null values with some other values
6. Then Analysing the dataset to what dataset can offers and what can we do
7. After analysing the dataset,I guess some information will be useful.So I made the objectives

Objective of the analysis :

1. Finding the total reviews in each genre
2. What are the top 5 genre which got maximum reviews
3. Find the Total Reviews By Each Year
4. Publisher with Most titles
5. Total sales by region
6. Find the maximum sales(all region) by year

Then I got the solution for the above objectives

1. Action - 21712.271, Adventure - 8962.822 , Fighting - 10358.293, Misc - 11570.414, Platform - 14736.235, Puzzle - 3269.506, Racing - 14527.607 ,Role-Playing - 14517.028 , Shooter - 16702.449 , Simulation - 6790.6910, Sports - 24282.5711, Strategy- 3297.49.
2. Sports genre got the maximum reviews of 24282.57
3. # we got the total reviews by the each year,so then we can calculate the top 10 reviews by year

we got the top 10 reviews by year.From this we can assume that the max no of games are sold in these years.And also we got 2008 has the most reviews so that we assume that max games were sold in 2008

As we assumed earlier ,**In 2008,most games were sold**

4. From this we got Electronic Arts publisher sold 341 different games
5. we got the max sales in North America 354 in sports genre
we got the max sales in Europe 221 in sports genre
we got the max sales in Japan 169 in Role-Playing genre
we got the max sales in Rest of World 71 in sports genre

Task 1 :

1. As of the First step I imported my necessary libraries and then uploaded and loaded the csv file
2. Then I see the structure of the data using data.shape to know about total number of columns and rows present
3. Used some commands to know about the structure of the dataset and the correlation between two columns and some insights
4. Then I do EDA for cleaning the dataset for more valuable insights and information to work
5. First of all i check the null values and then I fill with some other meaning full values.I didn't drop the row or column because it affect the other rows and columns so it could affect the information and insights we got
6. Then I converted the format of '**date released**' to date-month-year format which is already in object type
7. Then I extracted the date,month,year from the '**date released**' which is already in date-month-year to the specific column as 'date','month','year'
8. And then performed some insights about
directores with most flims
most movies released in a year