

Data Science with Python

Assignment 01



Instructions

- Use the given '**vgsales.csv**' dataset (Video Games Sales & Game Ratings Data Scraped from VzCharts) to answer all **10** questions.
- When you answer, use a **markdown cell** to indicate the question followed by your answer in a code cell.

Ex:

i) List of dictionaries using passenger details(Name, Sex, Age, cabin) and Print details related to the first 2 passengers.

```
In [22]: titanic=open('titanic.csv')  
         headings=titanic.readline().split(",")
```

Name: Braund Mr. Owen Harris Sex: male Age: 22 Cabin:

Name: Cumings Mrs. John Bradley (Florence Briggs Thayer) Sex: female Age: 38 Cabin: C85

- Rename the **ipynb** file with your answers, using the **name** (Eg : A.Perea.ipynb) and submit it **on or before 14th of October, 2023**
1. How many missing values are there in each column of the Video Games Sales dataset? Display the results in a bar chart.
 2. How can we handle missing values in the "Year_of_Release" column? Write a function to fill in missing values with the median year and show the first 10 rows in the dataframe.
 3. How does the distribution of video game genres look like in the dataset? Create a pie chart to visualize the count of games in each genre.
 4. How does the distribution of critic scores vary by genre? Create a boxplot to visualize the distribution of critic scores by genre.
 5. Is there a correlation between critic scores and user scores for video games? Create a scatter plot to investigate this relationship.

6. What is the average global sales for video games in each genre? Create a bar chart to visualize the average global sales by genre.
7. How many video games were released in each year, and what were the top 5 years with the highest number of game releases? Write a function to calculate this
8. How has the total global sales of video games evolved over the years in the "vgsales" dataset? Create a line plot to visualize the trend.
9. Visualize the correlation between global sales, critic scores, and user scores in the "vgsales" dataset using a heatmap?
10. Analyze the distribution of video game sales (e.g., global, North America, Europe, Japan, other regions) in the "vgsales" dataset and identify which region contributes the most to the total sales?