**Prog6280c**

**(Video Game Sales)**

**Program Description:** This lab program will guide you through using Pandas and Matplotlib in Python (or equivalent dataframe/plotting libraries in R, Julia, etc.) to analyze and visualize the “[Global Video Game Sales and Reviews](https://www.kaggle.com/datasets/thedevastator/global-video-game-sales-and-reviews)” dataset. The goal is to help you learn data wrangling and visualization techniques while uncovering interesting insights from the dataset.

1. Data Wrangling Exercises

* Remove leading/trailing spaces in column names.
  + df.columns = df.columns.str.strip()
  + Optionally, replace spaces in columns with underscores:

df.columns = df.columns.str.replace(' ', '\_')

* **Handling missing values (remove dirty rows or replace bad data).**

df = df.dropna() # Remove nulls/NaNs

* **Converting data types** (e.g., converting numbers to the correct format, like `year` to int).
* Normalizing data (if applicable).

2. Data Visualization Tasks

Each task should guide you to tell a “story” with the data using various types of plots; **this time, the objectives and plot types are not explicitly defined – you have to decide what ‘story’ to tell and how**:

**Task 1: Sales Analysis by Region**

* Columns to Use: North America, Europe, Japan, Rest of World.
* Suggested Plot Types: Bar plot, Stacked bar plot.
* Description: Analyze and compare the sales volumes of video games across different regions.

**Task 2: Trends Over Time**

* Columns to Use: Year, Global.
* Suggested Plot Types: Line plot, Area plot.
* Description: Investigate trends in global video game sales over the years.

**Task 3: Platform Popularity**

* Columns to Use: Platform, Global.
* Suggested Plot Types: Count plot, Pie chart, Donut chart.
* Description: Examine the distribution of global sales among different gaming platforms.

**Task 4: Genre Analysis**

* Columns to Use: Genre, Global, Review.
* Suggested Plot Types: Scatter plot, Box plot.
* Description: Explore the relationship between game genres, their sales, and review scores.

**Task 5: Publisher Performance**

* Columns to Use: Publisher, Global, Review.
* Suggested Plot Types: Heatmap, Bubble chart.
* Description: Evaluate how different publishers perform in terms of global sales and review scores.

3. Analysis and Interpretation

* Interpret the plots (one sentence each).
* Discuss any additional insights you may have found during your analysis.

**Data Location:** video\_game\_sales.csv

**Sample Output** **(not including interpretations):**

Task 1: Sales Analysis by Region

data:image/png;base64,

Task 2: Trends Over Time

A graph showing the growth of sales

Description automatically generated

Task 3: Platform Popularity

A graph of a number of games released on each platform

Description automatically generated

Task 4: Genre Analysis

A graph of colored dots

Description automatically generated

Task 5: Publisher Performance

A screenshot of a computer screen

Description automatically generated