GitHub Submission Files for Video Game Sales Analysis

## 1. Python Script: video\_game\_analysis.py

import pandas as pd  
import numpy as np  
  
# Load the dataset  
try:  
 df = pd.read\_csv("../Data/vgsales.csv")  
 print("Data loaded successfully.")  
except FileNotFoundError:  
 print("Error: 'vgsales.csv' file not found in '../Data/' directory.")  
 exit()  
  
# Display the first 10 records  
print("\nFirst 10 records:")  
print(df.head(10))  
  
# Drop missing values  
df.dropna(inplace=True)  
print(f"\nData shape after dropping missing values: {df.shape}")  
  
# Add a new column 'New\_Year' based on the release year  
df['New\_Year'] = np.where(df['Year'] > 2005, "Post 2005", "Pre 2005")  
  
# Calculate average global sales by 'New\_Year'  
avg\_sales\_by\_year = df.groupby('New\_Year')['Global\_Sales'].mean()  
print("\nAverage Global Sales by Release Period:")  
print(avg\_sales\_by\_year)  
  
# Conclusion: Which period had higher average sales  
if avg\_sales\_by\_year['Pre 2005'] > avg\_sales\_by\_year['Post 2005']:  
 print("\nConclusion: Average sales were higher before 2005.")  
else:  
 print("\nConclusion: Average sales were higher after 2005.")

## 2. README.md

# 🎮 Video Game Sales Analysis  
  
## 📘 Description  
This project analyzes global video game sales data to determine if average sales were higher before or after 2005. A new column is added to classify records as "pre-2005" or "post-2005" based on the release year, and average global sales are compared between these two periods.  
  
---  
  
## 🛠 Getting Started  
  
### ✅ Prerequisites  
To run this project, you need:  
- Python 3.x  
- Pandas  
- NumPy  
  
Install the required libraries using:  
pip install pandas numpy  
  
### 💾 Installing  
1. Clone the repository:  
 git clone https://github.com/your-username/video-game-sales-analysis.git  
 cd video-game-sales-analysis  
  
2. Place the vgsales.csv file in a Data directory under the project root (the file path used is ../Data/vgsales.csv).  
  
---  
  
## 🧪 Running the Tests  
  
### 🧩 Breakdown of Tests  
- Data Loading: Reads the CSV using Pandas.  
- Cleaning: Handles missing values.  
- New Column: Adds a New\_Year column marking each record as "Pre 2005" or "Post 2005".  
- Sales Analysis: Calculates and compares the average global sales for each category.  
  
---  
  
## 🚀 Deployment  
This is a standalone data analysis script and doesn't require deployment. You can simply run the .py file to see the output in a terminal or Jupyter Notebook.  
  
---  
  
## 👨‍💻 Author  
- Bhanushali Meet   
- Chacko Vilas   
- Chaudhary Kshitij   
- Khan Sharmeen   
- Poudel Saurav   
  
---  
  
## 📜 License  
This project is for educational use only. Do not reproduce without permission.  
  
---  
  
## 🙏 Acknowledgement  
- Thanks to the project supervisor and DATA 1102 course instructors for guidance.  
- Dataset: vgsales.csv (original source not specified in the report).

## 3. .gitignore

# Byte-compiled / optimized / DLL files  
\_\_pycache\_\_/  
\*.py[cod]  
\*$py.class  
  
# Data  
\*.csv  
\*.tsv  
\*.xls  
\*.xlsx  
  
# Virtual environments  
venv/  
env/  
  
# Jupyter Notebook checkpoints  
.ipynb\_checkpoints/  
  
# VS Code  
.vscode/  
  
# Mac  
.DS\_Store