Homework 4 Portfolio

# Introduction

The basis for this assignment was to use dimensionality reduction to analyze video games sales trends across the world. I wanted to see if there have been any trends with sales numbers related to genre. I used PCA and UMAP to create two static visualizations to discover if any trends exist. The data used includes sales data from North America, Europe, Japan, Other, and total Global sales. By reducing these columns, I was able to find some trends as I will explain below.

# Static Visualization 1: PCA

Below is the PCA reduction on the Video Games sales data:

A graph with many colored dots

Description automatically generated

The above graph provides some unique insights into the general sales trends of video games. Through PCA, the aim is to discover patterns and certain groupings in the different regions. By reducing it to two components to represent the variance in the dataset, it’s possible to explore if specific genres dominate different regions. Some key findings include clustering between sports and action games, which suggests consistent sales trends across regions for both genres. Also, role-playing games are more dispersed, indicating varying performance across different regions. Outliers are any games that have quite large global success and are most likely games that went above and beyond.

# Static Visualization 2: UMAP

Below is the UMAP reduction:

A graph with colorful dots

Description automatically generated

Through a UMAP reduction, the goal is to discover any similarities or differences in games sales across genres. The aim here is to see what outliers and clusters are present. Some of the findings here include the sports and action genre clusters, an additional dispersed genre in puzzle games (which was not easily shown in PCA), and again more outliers that differ from other games in their genre.

# Comparisons Between PCA and UMAP

Both PCA and UMAP are good for revealing clustering patterns, such as the Sports and Action clusters that indicate the two genres perform at the same rate in the same regions. PCA uses variance to highlight the dominant trends across regions and genres while UMAP takes a more focused approach emphasizing individual data points. Through both these visualizations it’s possible to see multiple different perspectives on the sales data for video games.