Final Report for STAT 4185 Project

Title: Comparative Analysis of Gaming Console Prices and Global Sales

Introduction:
This project is an in-depth analysis of gaming consoles worldwide, focusing on their prices and global sales. The objective was to understand the relationship between a console's price and its sales performance across different regions. This analysis aimed to provide insights that could inform pricing strategies and marketing decisions in the gaming industry.

Methodology:

Data Collection: The data was collected from two primary sources: VGChartz for global sales figures and inflationstation. Other sources weren't as good as these two sources.

Data Cleaning and Preprocessing: The data was cleaned using Python's pandas library, which involved handling missing values, normalizing data formats, and ensuring accuracy and reliability. Although the accuracy isn't as par as I wanted it to be.

Data Analysis Tools: Python's pandas for data manipulation, matplotlib for visualization, and scikit-learn for machine learning models.

Data Analysis and Results:

You can find out yourself in the project. But:
Best parameters (Grid Search): {'max_depth': None, 'min_samples_leaf': 4, 'min_samples_split': 2}
Best parameters (Randomized Search): {'min_samples_split': 2, 'min_samples_leaf': 4, 'max_depth': None}
Grid Search Best Model's R^2: 0.3010738002093295
Randomized Search Best Model's R^2: 0.3010738002093295

Global Sales Analysis:

The dataset included sales figures across North America, Europe, Japan, and the rest of the world, along with global sales totals.
A bar graph visualizing the top 10 gaming consoles by global sales was created, highlighting the dominant players in the market.

Console Prices Analysis:

A second dataset contained the prices of various gaming consoles.
A horizontal bar graph depicted the prices of these consoles, offering a visual comparison.
Merged Analysis of Sales and Prices:

The two datasets were merged based on console names, allowing for a combined analysis of sales and prices.
A scatter plot demonstrated the relationship between console prices and global sales, with a calculated correlation coefficient.
Modeling and Predictive Analysis:

A linear regression model was trained to predict global sales based on console prices, evaluated by Mean Squared Error (MSE) and R-squared metrics.
Decision tree regression was employed, revealing insights into the data's structure and further evaluating the sales-price relationship.
Hyperparameter tuning with GridSearchCV and RandomizedSearchCV was performed to optimize the decision tree model.
Discussion:

The project revealed a nuanced relationship between console prices and global sales, with a negative correlation coefficient suggesting a complex interplay of factors influencing sales.
The linear regression model provided a baseline understanding, while the decision tree regression offered more depth.
Hyperparameter tuning demonstrated the significance of model optimization in predictive accuracy.
Conclusion:
This project highlighted the intricate dynamics of the gaming console market. While pricing is a crucial factor, it's interwoven with other aspects like brand loyalty, game availability, and technological advancements. Future studies could delve into these additional factors or explore similar analyses for other gaming platforms.

References:

VGChartz website for global sales data.

https://www.vgchartz.com/charts/platform_totals/Hardware.php

Online source for console prices.
https://www.inflationstation.net/