Predicting Video Game's Scores Based on its Global Sales

By Nouf Alotaibi

Email: noufmitla@gmail.com

1. Introduction

Video games have become an integral part of the online culture. They became famous when their medium moved from large arcade machines to personal computers and consoles. Nintendo has been at the forefront of this online movement with video game consoles like the Gameboy, with Microsoft and Sony following closely behind with the Xbox and PlayStation. Now, video games have a wider audience. There are multitudes of multiplayer and online games, with many gaming conventions across the globe [1].

2. Problem Statement

In this project, I will be investigating the relationship between user and critic scores on the global sales of games to predict the games' score and popularity. Also, test the effect of user and critic scores related to the game's publisher, developer, and genre. Moreover, I will get familiar with Machine Learning, web scraping techniques, and modeling methods such as Linear Regression.

3. Dataset

For this project, I gathered the data from two sources:

- **Metacritic:** *Metacritic* is a website that aggregates reviews of films, TV shows, music albums, and video games. For each product, the scores from each review are averaged. It is also known as the leading online review aggregation site for the video game industry. I used *BeautifulSoup* library to scrap the data of the best video games of all times from [2], I gathered about 2,000 video games records with 11 features. The table below illustrates the Metacritic video games dataset's features and their types.

Column Name	Column type
game_name	Object
platform	Object

publisher	Object
developer	Object
release_date	Object
critics_rating	Object
num_critics_rating	Object
users_rating	Object
num_users_rating	Object
game_rate	Object
genre	Object

- **Video Games Sales from Kaggle:** *"Video Game Sales"* dataset that was uploaded to Kaggle.com, see [3]. The data set contains 11 features and 16,598 records, each of which is a game released between 1980 and 2020.

Column Name	Column type
Rank	int64
Name	Object
Platform	Object
Year	float64
Genre	Object
Publisher	Object
NA_Sales	float64
EU_Sales	float64
JP_Sales	float64
Other_Sales	float64
Global_Sales	float64

4. Tools

These are the technologies and libraries that I will be using for this project:

- **Technologies:** SQL, SQLlite, Python, Jupyter Notebook.

- **Libraries:** NumPy, Pandas, Matplotlib, Seaborn, AQLAlchemy, Requests, BeautifulSoup, Selenium, Statsmodels, Scikit-learn.

5. Resources

[1] VideoGameCons.com. (2018). 2018 Video Game Convention Calendar| VideoGameCons.com. [online] Available at: https://videogamecons.com/calendar/calendar.php?year=2018

[2] "Best Video Games of All Time", Metacritic, 2021. [Online]. Available at: https://www.metacritic.com/browse/games/score/metascore/all/all/filtered?page=0.

[3] "Video Game Sales", Kaggle.com, 2021. [Online]. Available at: https://www.kaggle.com/gregorut/videogamesales.