Project proposal

This Project proposal is for the project of T5 Bootcamp requirements.

Video Game Sales

Abstract

In this Project I will perform Exploratory data analysis, data visualizations and Modling of the video game sales.For EDA,I am using Python programming language, for data visualization,I am using Matplotlib and Seaborn and for data modling,I am using sketlearn.

Questions

I will try to answer the following questions:

1. Global Sales Distribution.
2. The video games produced in a each year.
3. which genre sold the most globally?
4. The platform,game and the publisher which has the top sales in global sales
5. The platform,game and the publisher which has the low sales in global sales
6. How does the Genre affect the Global_Sales?
7. What are most games produced in a specific Gaming Platform?

Design

This project is one of the T5 Data Science BootCamp requirements. Data provided by Kaggle has been used in this project.

Data

The dataset is provided in .csv format.This dataset contains a list of video games with greater sales. It contains 16,598 records, each record has 11 features. The most relevant feature to this project is Total worldwide sales.This feature is extracted from other features such as Sales in North America,Sales in Europe,Sales in Japan, and Sales in the rest of the world.

Fields include

• Rank - Ranking of overall sales

• Name - The games name

• Platform - Platform of the games release (i.e. PC,PS4, etc.)

• Year - Year of the game's release

• Genre - Genre of the game

• Publisher - Publisher of the game

• NA_Sales - Sales in North America (in millions)

• EU_Sales - Sales in Europe (in millions)

- JP_Sales - Sales in Japan (in millions)

- Other_Sales - Sales in the rest of the world (in millions)

- Global_Sales - Total worldwide sales.

Tools

- Data processing :Pandas , Numpy.
- Modelling: Scikit-learn
- Visualizations Matplotlib and Seaborn

Algorithm
- Cleaning the data using EDA process
- For modelling Linear Regression