Analysis of Video Game Sales

Aymen Soussi

Abstract

We will explore the data set of video game sales to try to predict the global sales of video games.

Motivation

In this project we will make an analyze for video game sales data set with seaborn. We will make a data visualities at first with matplotlib and then we will use machine learning algorithms to predict the global sales according to the actual sales provided.

Dataset

-Data Source: Video Game Sales Dataset
-Location: https://www.kaggle.com/gregorut/videogamesales
-Filename: videogamesales.zip

Data Preparation and Cleaning

At a high-level, what did you need to do to prepare the data for analysis?

We need at first to clean up our data from null values so that we eliminate outliers and we can make a better analyze for our data.

Describe what problems, if any, did you encounter with the dataset?

If we don’t clean our data we will get weeker analyze results because we will consider outliers as part of our data.

Research Question(s)

Our research question in this project is:

According to the actual sales can we predict the global sales ?

Methods

We have used actual sales to predict global sales.

Findings

We have found that using actual sales we can predect glabal sales with a great accurency (99%)

Limitations

For our work, we have put a lot of time to understand what is the best feature to focus on it to predict the global sales of video games so we can not estimate that the acctuale sales can be used in every sales dataset to predict the global sales so every time we need to make a good analyse to find the best feature.

Conclusions

Using xgboost we have predect global sales of video games based on the actual sales.

Acknowledgements

I have collected data by my self from Kaggle, I haven’t got feedback because of limit time (this is the last day of the course)

References

https://xgboost.readthedocs.io/en/latest/

https://matplotlib.org/

https://seaborn.pydata.org/