Dataset name :Video\_Games\_Sales\_as\_at\_22\_Dec\_2016

 It contains [**16719] rows** and [**16] features/columns**where the features include:

* **NA\_Sales, EU\_Sales, JP\_Sales:** Sales in North America, Europe and Japan .
* **Other\_Sales:** Sales in other parts of the world
* **Global\_Sales:**Total worldwide sales
* **Rating:** The [**ESRB**](https://www.esrb.org/)ratings. And other features

For dataset handling:

1. Drooping the columns will not useful fot predicting the out put
2. Handling the numerical columns (missing values with mean& outliers)
3. Nonmumeric(missing values with mode&)
4. The **StandardScaler()** function from **sklearn.preprocessing** :

used to **standardize** or **normalize** features in a dataset by **scaling them** to have a **mean of 0** and a **standard deviation of 1**.

5-The **OneHotEncoder** from **sklearn.preprocessing:**

is used to convert **categorical variables** (non-numerical data) into a **numerical format** that machine learning models can work

how:

will create **one binary column for each unique category**. It works well for **nominal categorical features** regardless of the number of categories.

However, for cases where the number of categories is very large (high cardinality), **other techniques** can be more efficient. Let’s discuss **two scenarios**:

The general idea of machine learning is to :

Predicting video game General\_Sales feature

Using 2 Models:

1-Linear Registration

2-k-Nearest Neighbors (kNN)

🡺Key Evaluation Metrics

1-Mean Absolute Error (MAE)

2-Mean Squared Error (MSE)

3. Root Mean Squared Error (RMSE)

4. R² Score (Coefficient of Determination)

For linear regression Linear-

🡺MSE: 0.02141593019138756,

🡺RMSE: 0.14634182652744074,

🡺MAE: 0.07700490430622009,

🡺R²: 0.5784859854148776

For knn

🡺KNN - MSE: 0.0207409209280303

🡺RMSE: 0.14634182652744074

🡺MAE: 0.07434833532695374,

🡺R²: 0.5917716966558625