Dilan Kusuma Final Project

Video Game Sales

Welcome!!

Initiating Presentation

Final Project / Video_Games_Sales.csv - Executive Summary

Business Background

We can conclude that the Video Games Sales data has served a company which have around 16.700 users. In response it has feature such as Name, Platform, Release Date, Genre, Publisher, Sales, Critics, Users, Developers and Ratings

Problems Statements

An immediate needs of treatment regarding the sharp decline of sales since the year of 2010

Objective

Figure the effecting factor and possibilities regarding the matter of declining sales

Proposed Solutions

The deploying of Machine Learning (ML) in predicting a possible leverage from the data in a mission to raise the sales that's been declining, also provide a fit assistance to the problem.

Result:

Analysis Results :

- ➢ As for the Sales, NA are by far the best sales
- ➢ EA and Activision are the most promising Publisher
- ➢ The most hyped genre are Action and Sport genre as
- ➢ the have the most sales in comparison to other genre

Points mentioned above require appropriate treatment

ML Result :

- Methods of regression used are Linear Regression bears the result of 97% R Square along with MSE score of 0.0016 accuracy

Business Benefit

- ❖ It is possible to be used as a bas eline of product treatment
- ❖ The Data have been dealt and an alyzed in regards to the User Beh avior, Interest, and Preference of Games
- ❖ Lay a data with definitive prove regarding the Game Sales as of 1991 until 2016, and are qualified to use for further Marketing Plan
- ❖ Found a provable solution in which causing the declining of sales

Work Flow

IMPORT LIBRARY

First of all we must import all the relevant library so we can proceed further

READ DATA

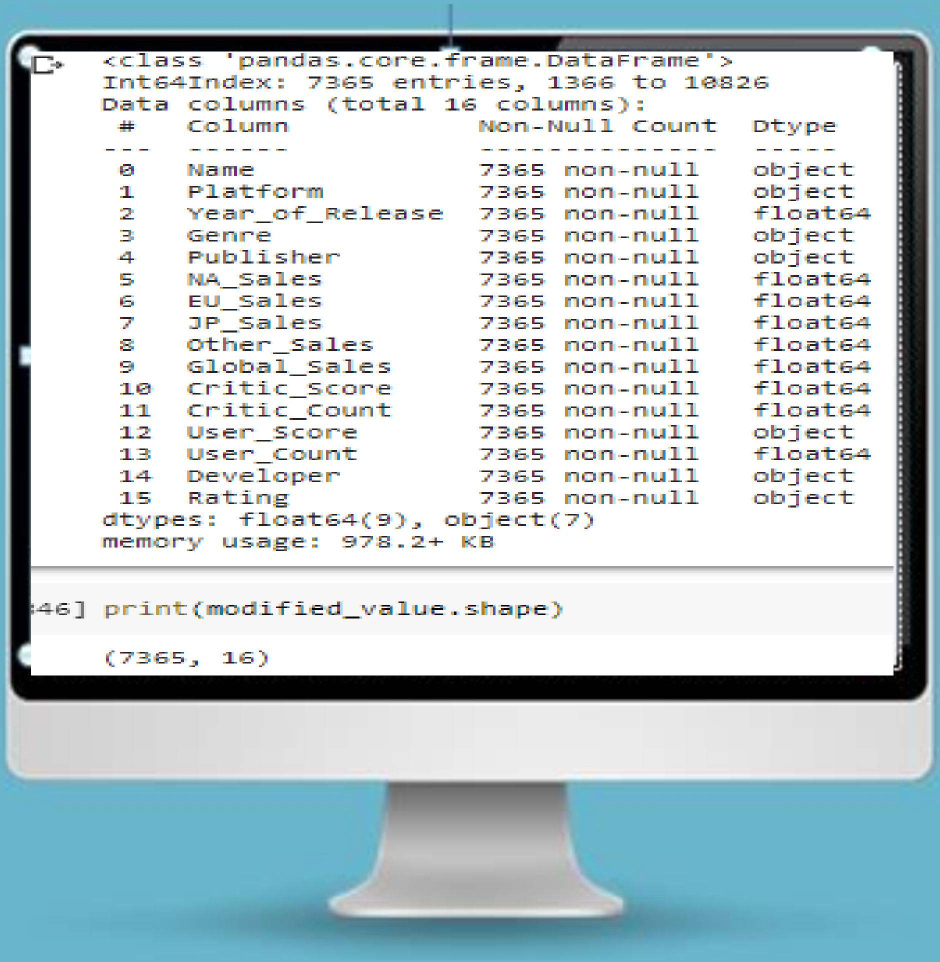
Get the data into Google Colaboratory

EXPLORATORY DATA

Probe the data as much as we can so data become a lot more understandable

REGRESI

Find the most befitting R square and MSE

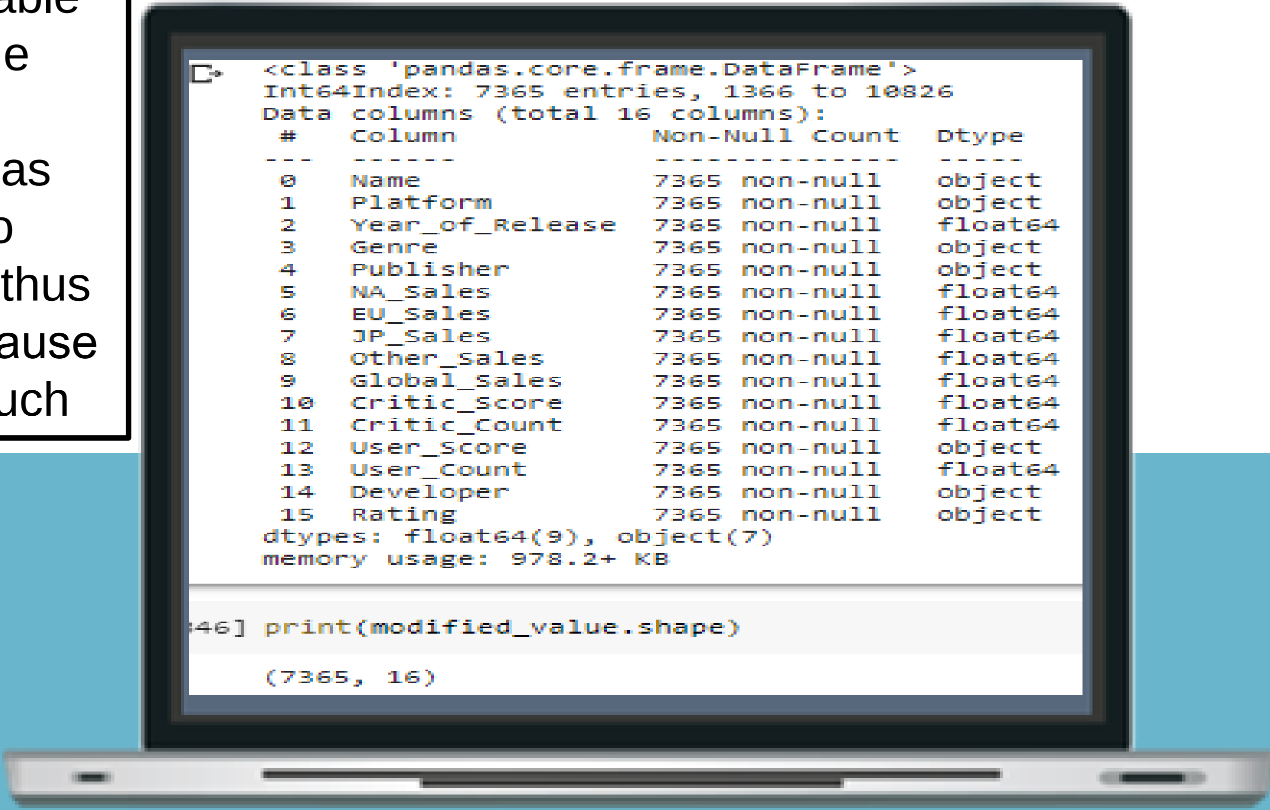


```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 7365 entries, 1366 to 10826
Data columns (total 16 columns):
 #   Column           Non-Null Count  Dtype  
---  ------           --------------  -----  
 0   Name             7365 non-null   object 
 1   Platform         7365 non-null   object 
 2   Year_of_Release  7365 non-null   float64
 3   Genre            7365 non-null   object 
 4   Publisher        7365 non-null   object 
 5   NA_Sales         7365 non-null   float64
 6   EU_Sales         7365 non-null   float64
 7   JP_Sales         7365 non-null   float64
 8   Other_Sales      7365 non-null   float64
 9   Global_Sales     7365 non-null   float64
 10  Critic_Score     7365 non-null   float64
 11  Critic_Count     7365 non-null   float64
 12  User_Score       7365 non-null   object 
 13  User_Count       7365 non-null   float64
 14  Developer        7365 non-null   object 
 15  Rating           7365 non-null   object 
dtypes: float64(9), object(7)
memory usage: 978.2+ KB

46] print(modified_value.shape)

(7365, 16)
```

As we know that encountering impure d ata are common in data science, therefore we must clean the data before we proceed. Though unfortunate, our data seems to be a little underperformed as it has many null value.

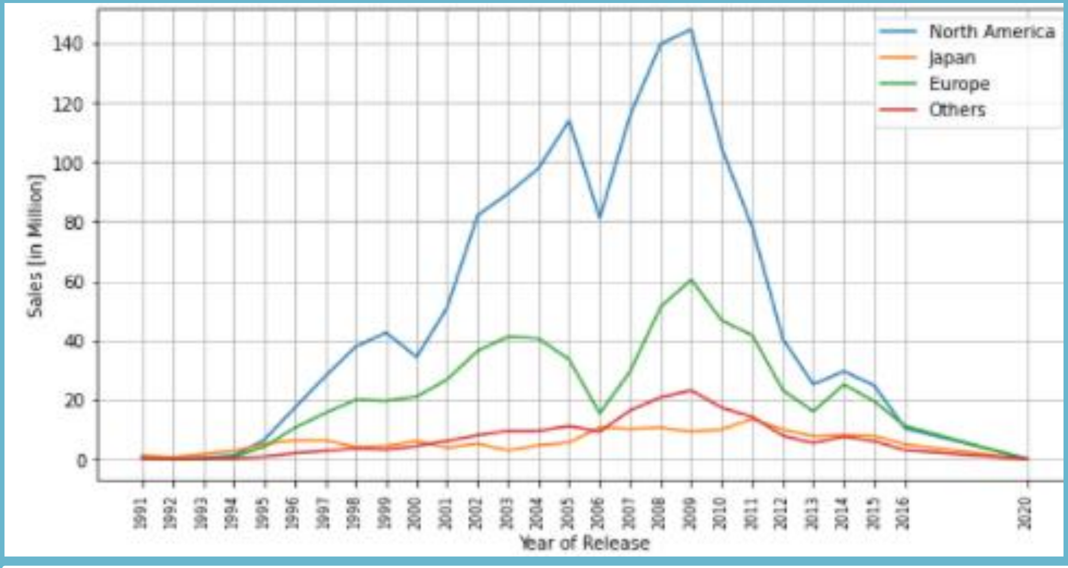
Thus I decide to do the outlier first because I believe that doing so later would cause a disrupt in the data

By doing the outlier first enable us to fill the null or nan value using ***bfill*** method when handling the missing value as the methods used Means to process the missing value, thus if there’s still outlier, it will cause the Means to spread too much

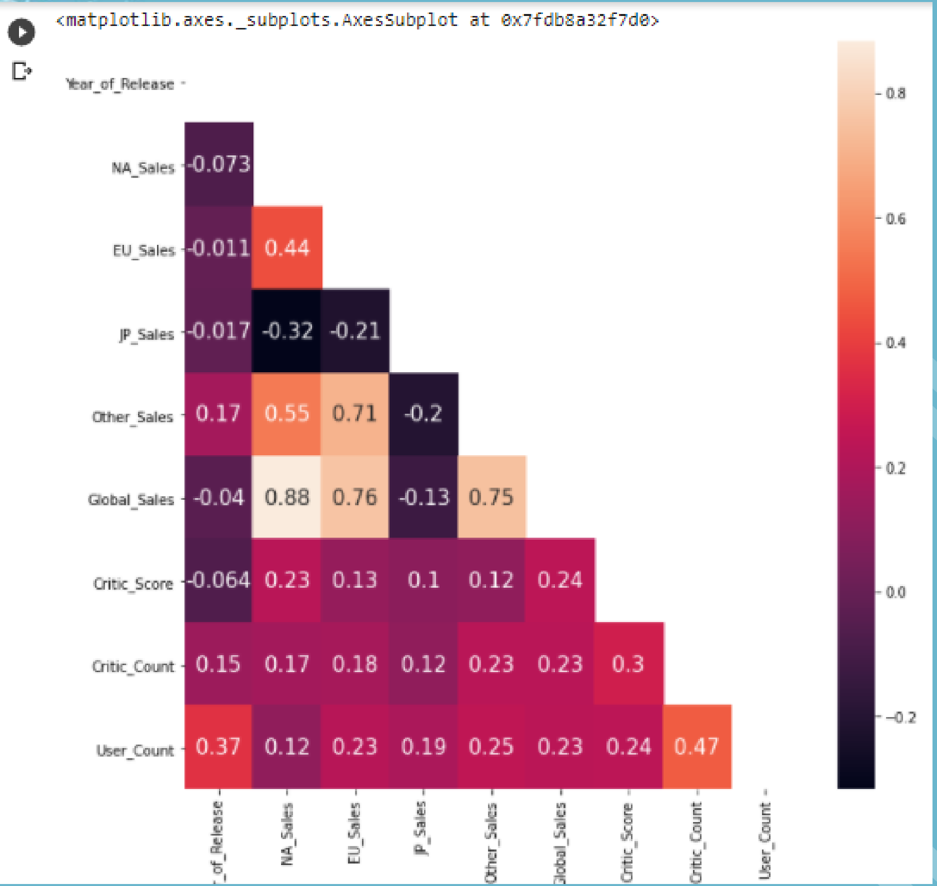


GRAPH OF SALES DISTRIBUTION

Graph based on the Years the game launched



From this graph we can conclude that there’s been a sharp decline in Japan Sales.



HEATMAP CORRELATION

From what I have tested and observed I believe that we can exclude JP_Sales as it not a significant Variable for our *used feature* later in order to further improve my analysis

```
# Evaluasi Model dengan Mean Square Error (MSE) dan R squared
print("MSE :", metrics.mean_squared_error(y_test,y_test_pred))
print("R squared :", metrics.r2_score(y_test,y_test_pred))
```

```
MSE : 3.15843608838748e-05
R squared : 0.9994666376262938
```

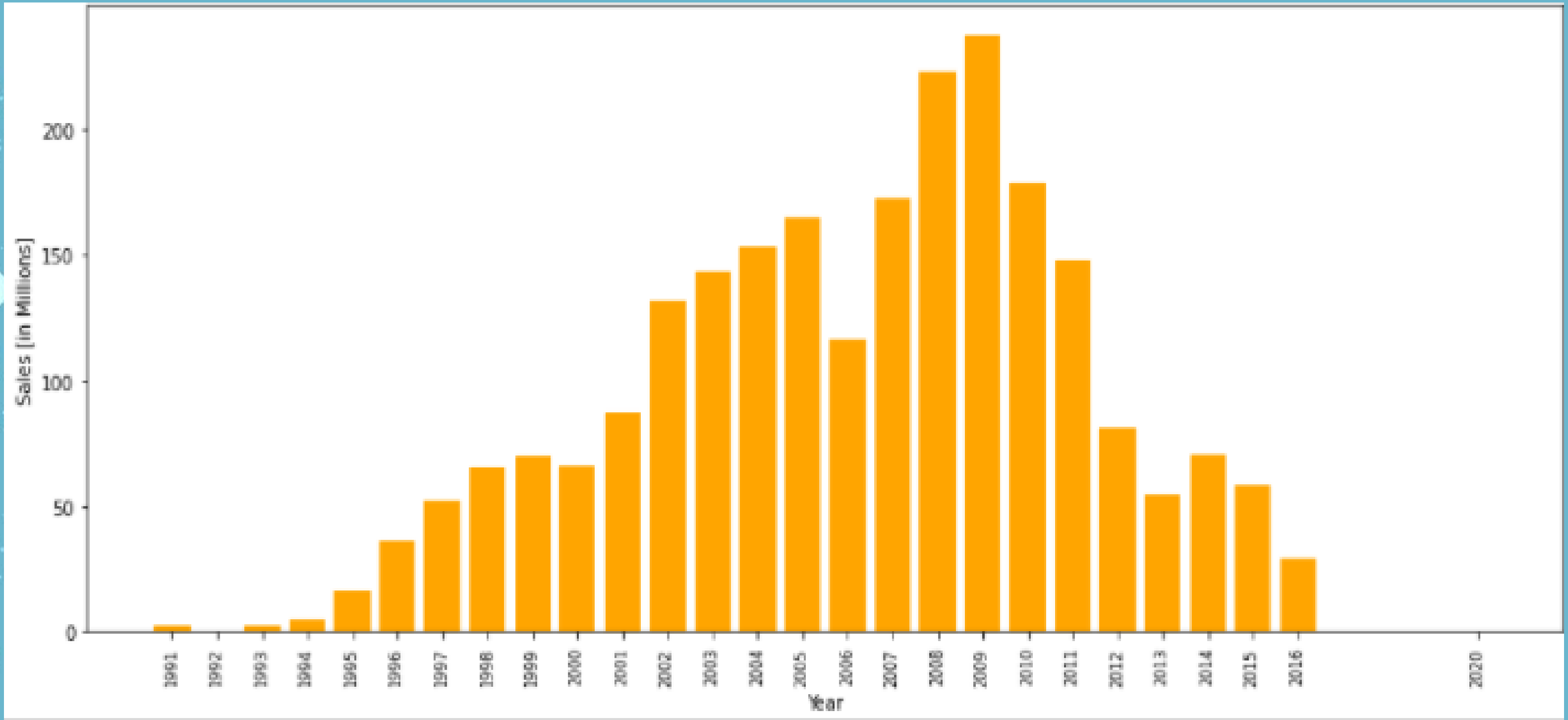
With JP_Sales

```
[228] # Evaluasi Model dengan Mean Square Error (MSE) dan R squared
print("MSE :", metrics.mean_squared_error(y_test,y_test_pred))
print("R squared :", metrics.r2_score(y_test,y_test_pred))
```

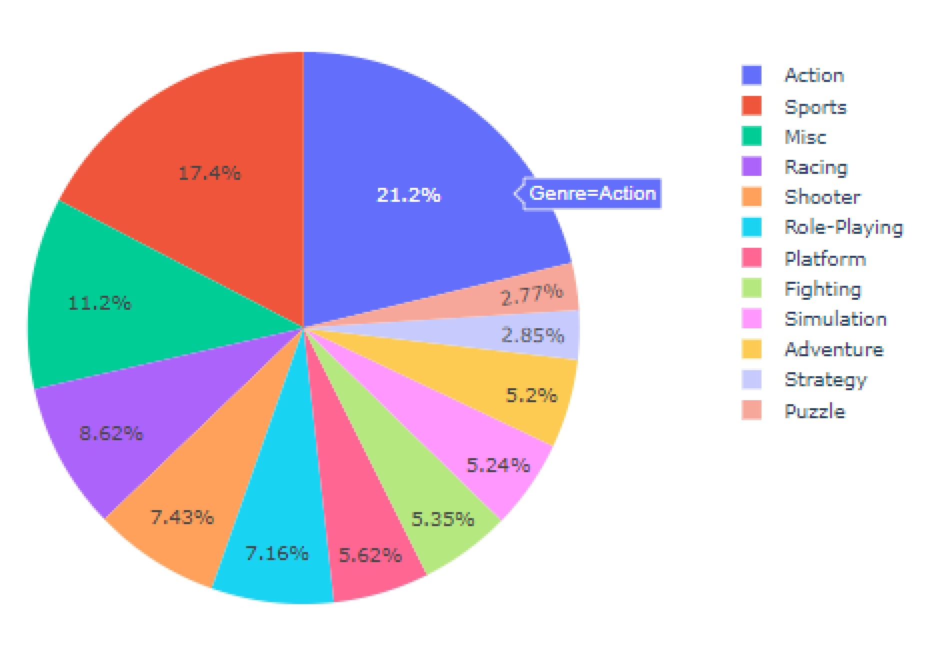
```
MSE : 0.001680154702264934
R squared : 0.9716273726896582
```

Without JP_Sales

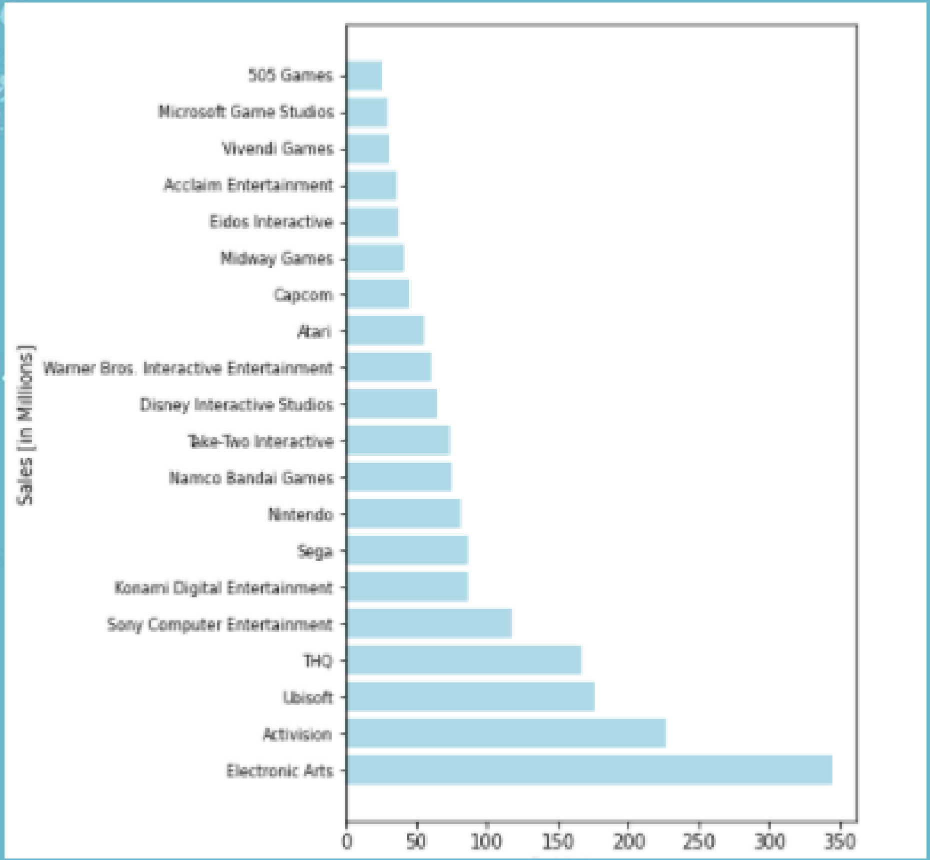
The removal of JP_Sales has caused a drastic improvement for MSE and R Squared leading to a reliable and accurate result



As the graph above explained there's been a sharp decline of game sales ever since year of 2010 up to the end of the data. Signaling for an immediate response and treatment



This pie graph shows an overwhelming disparity in the Genre Feature. It also suggest that the Action and Sport Genre are what majoring the most hyped Genre, thus contributing in Sales



I believe there are some factor in which possibly affect the sales, such as Publisher. The graph shows the most contributing Publisher are EA (Electronic Art) by a far margin, followed by Activision

Thank you

I’m very grateful for your enthusiastic attention