# **EXPLORATORY DATA ANALYSIS REPORT**

**(VIDEO GAME SALES DATA)**

**CONTENTS:**

* **Introduction**: Our understanding of the video game sales industry
* Initial hypotheses
* Relevant questions for the analysis
* Detailed process description, using visualization

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

* **THE VIDEO GAME INDUSTRY**

The video game industry according to Wikipedia.org is the economic sector involved in the development, marketing and monetization of video games. It encompasses dozens of job discipline and its component parts employ thousands of people worldwide. As of 2018, the video games industry generated sales of US$134.90 billion annually worldwide. The video games industry has since its inception seen a significant growth in sales as the years goes by as its popularity grows bigger.

it has numerous genres namely sports, racing, action games, adventure etc and many publishers in the industry, some of which include Nintendo, Ubisoft, Sony, sega etc.

* **HYPOTHESIS**

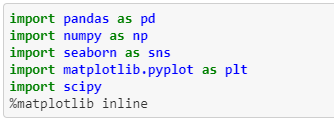
There has been an increase in video game global sales in recent with the Play Station Platform and North American region taking the highest share.

* **QUESTIONS**

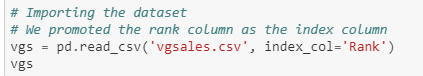
1. What is the trend of global video game sales over the years?
2. Which years had the highest and lowest video game sales?
3. What is the distribution of sales across genres, platform, and publishers, year on year and decade on decade?
4. What is the distribution of video game sales across the various regions over the years?
5. What has been the most popular video game genres/platforms over the years?

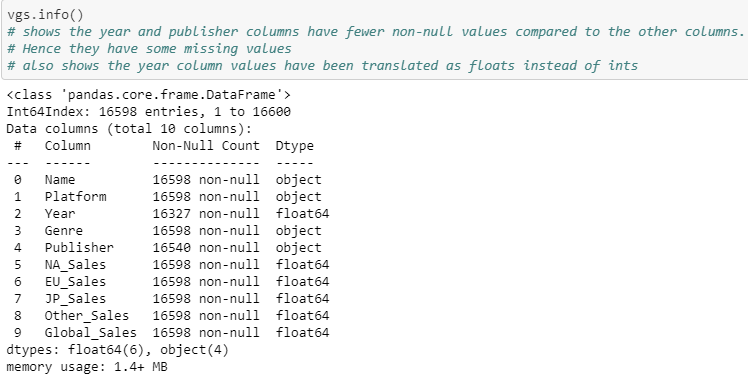
* **DETAILED PROCESS DESCRIPTION**
* First of all, we imported all of the libraries that we would need for the analysis and conducted a preliminary analysis as shown below:

**UNDERSTANDING THE DATASET**

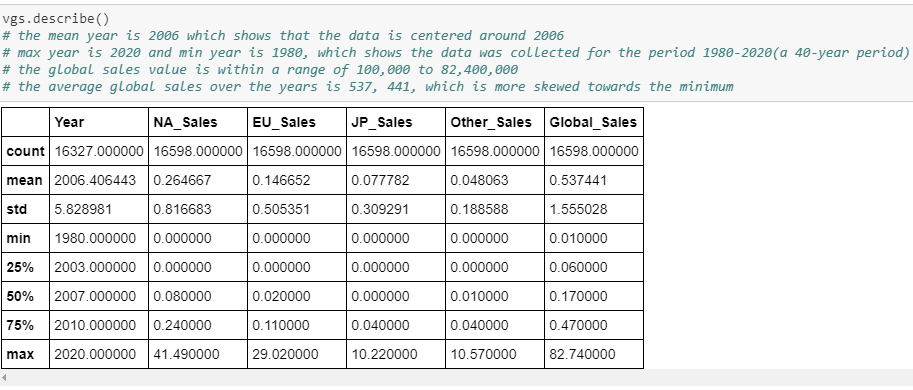


* Next, the dataset was loaded into python and inspection of all columns (variables) done:



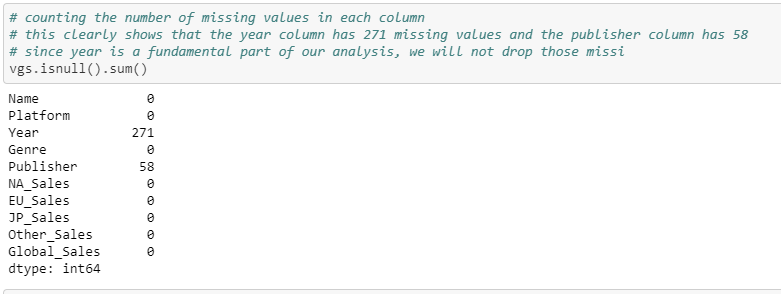


* And then a basic description of the whole datasets was carried out.



**CLEANING THE DATASET**

* After thorough inspection, and further exploration missing values were identified;



* Upon identifying these missing values, steps were taken to refill these missing values for the **year** column and the **publisher** column as indicated above as they play a major role in our analysis. (Details can be seen in the jupyter notebook attached)



**ANALYZING RELATIONSHIP BETWEEN DATASET**

**(DATA VISUALIZATION)**

* Data Visualization with the help of the Matplotlib imported in python was used to help answer the relevant research questions raised for analysis. Diagrams such as the histogram, pie chart, scatter plot, bar plots etc were utilized.