

T5 Boot camp

Data science

Project proposal

Games sales dataset

mohammad alajmi

Instructor:

Mejdal Al-Qahtani

abstract:

The main goal of this project is to study the video games market and analyze it to help the client to have a good understanding about the market and the sales of the video games to help the company establish their own game.

The Data set:

The used data set in this project is "Video_Games_Sales_as_at_22_Dec_2016", it is about the sales of games in the latest years in the different regions all over the world, the dataset covers all the sales before 2016, The data could be found on Kaggle website, the data has about 15 thousand row and about 11 columns. The used columns are Name, Platform, Year_of_Release, Genre, Publisher, NA_Sales, EU_Sales, JP_Sales, Other_Sales, Global_Sales

The used tools:

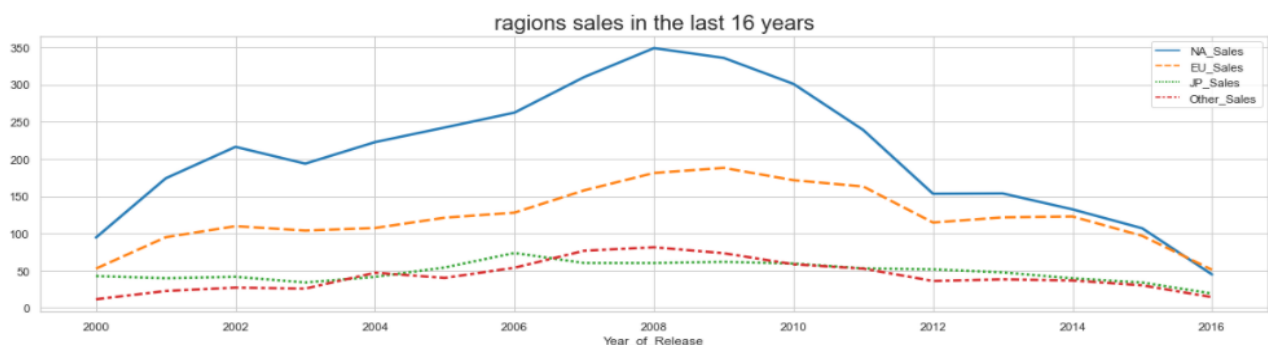
I have used different tools to achieve this goal such as Jupyter notebook, Python, Numby library, Pandas library, Matplotlib library and seaborn library.

algorithms:

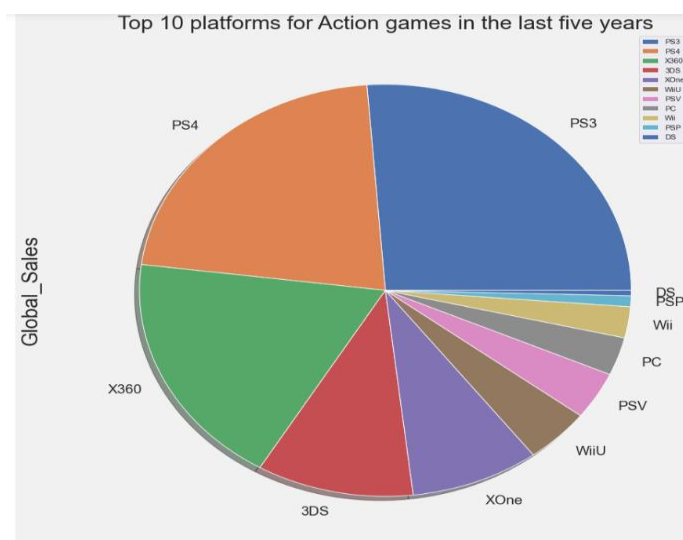
- fixing year of release to int value
- substitute all missing values with values
- fixing the number of indexes

Communication:

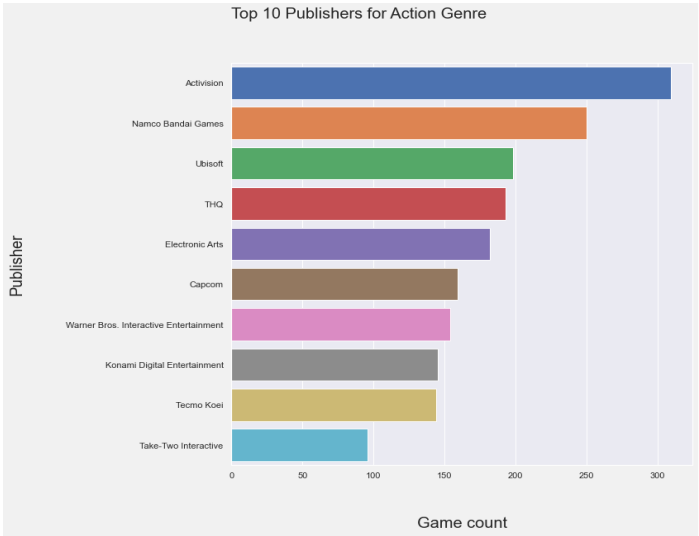
In figure 1 the used plot is called lineplot to represent the sales of every region in the last 16 years based on feature called "Year_of_Release".



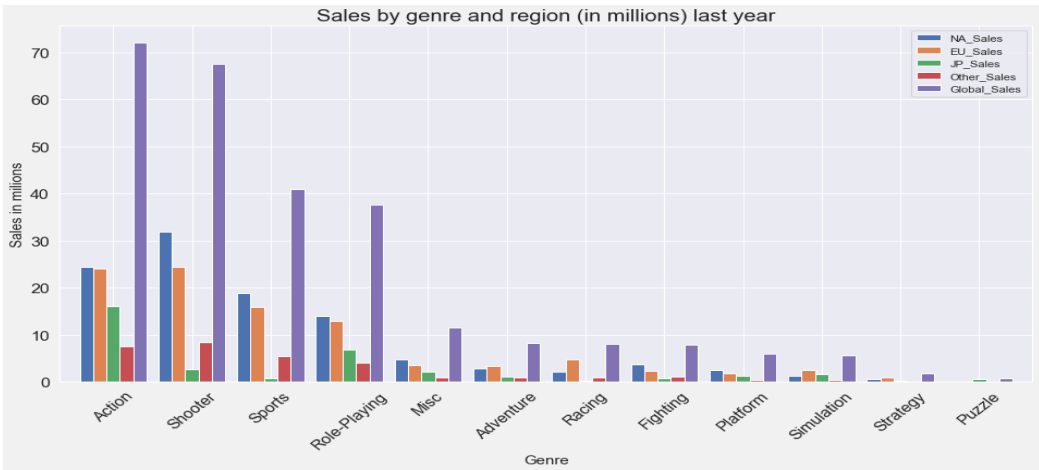
In figure 2 the used plot is called pie chart plot to represent the top 10 platforms for Action games in the last five years based on feature called "Global Sales".



In figure 3 the used plot is called BarH to represent the top 10 publishers for Action games based on feature called "Publisher".



In figure 4 the used plot is called Bar chart to represent the sales by genre and region for Action games on last year based on feature called "Global_Sales".



In figure 5 the used plot is called circle pie (donuts) chart to represent the top 20 video games sales based on feature called "Global_Sales".

