7PAM2000 Applied Data Science 1
Assignment 1: Visualization

Name: Sreegovind Sreekumar

MSc Data Science

Student ID: 22079644

Link to repository: https://github.com/SreegovindSreekumar/Assignment-1

INTRODUCTION

To visualize the data, we use a variety of graphical methods, in this visualization process we mainly use three types of visualization. All the datasets that are used here are taken from open sources. In each section details of the dataset are included. All the datasets are read by using the 'pandas' function and the graphs are plotted with the help of 'pyplot' function

DATASET

Dataset that is used here is taken from an open-source database called "Kaggle". During my research I came across a variety of different types of datasets, then I finally decided to use the dataset which contain the information of video game sales. Since the main aim of this project was to learn about visualization, the authenticity of the specified data is not checked. This dataset contains the information like name of the games, console in which the game was run, year of release of each game, category of the game, it's publisher and sales of the game in North America, Europe, Japan and other countries. Since it is a big data, it was not possible to include all in the simple plot graphs. So, I have taken only few relevant data from the set.

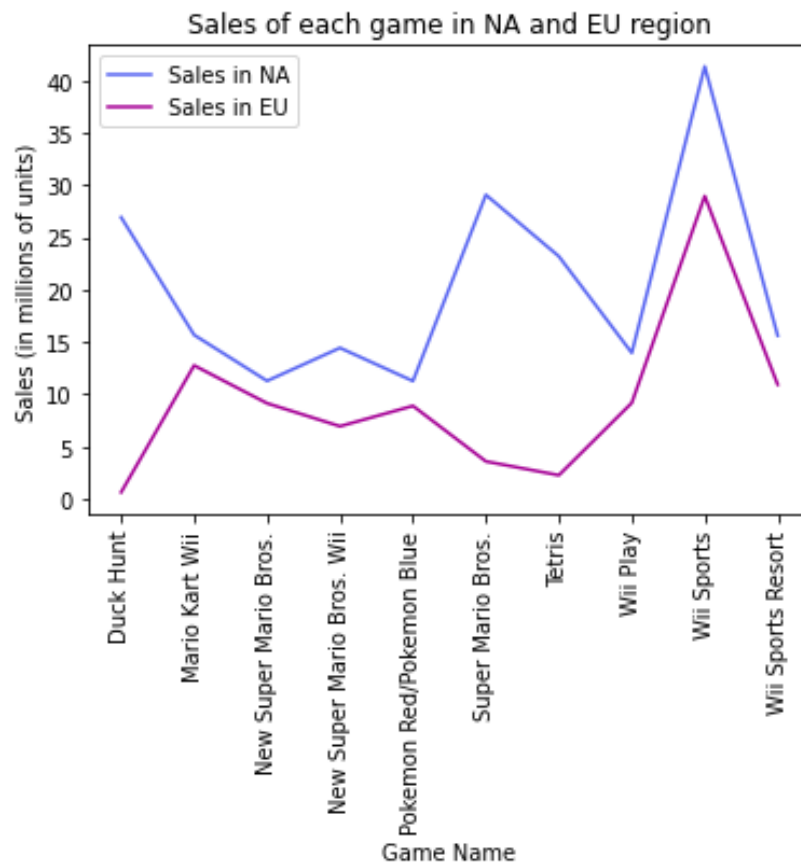Dataset is taken from the following link:

https://www.kaggle.com/datasets/rush4ratio/video-game-sales-with-ratings

Visualization 1: LINE PLOT (SALES OF EACH GAME IN NA AND EU REGION)

Line plot is made using the pyplot function. Since the dataset I had preferred enlist thousands of games, I am taking only the first ten games to denote the plot functions. In the graph X-axis represents the "Game name" and the Y-axis represents the "Sales" of the games. Two plot lines are made here. The blue plot line denotes the sale of a game in the North American region and the purple plot line denote the sale of a game in European region. From the below graph we can understand that Wii Sports game are the most sold game in both North American region and in European region. We can also figure out from the graph that the games New Super Mario Bros and Pokemon Red/Pokemon Blue is least sold game in North America while Duck Hunt game is the least sold game in Europe.
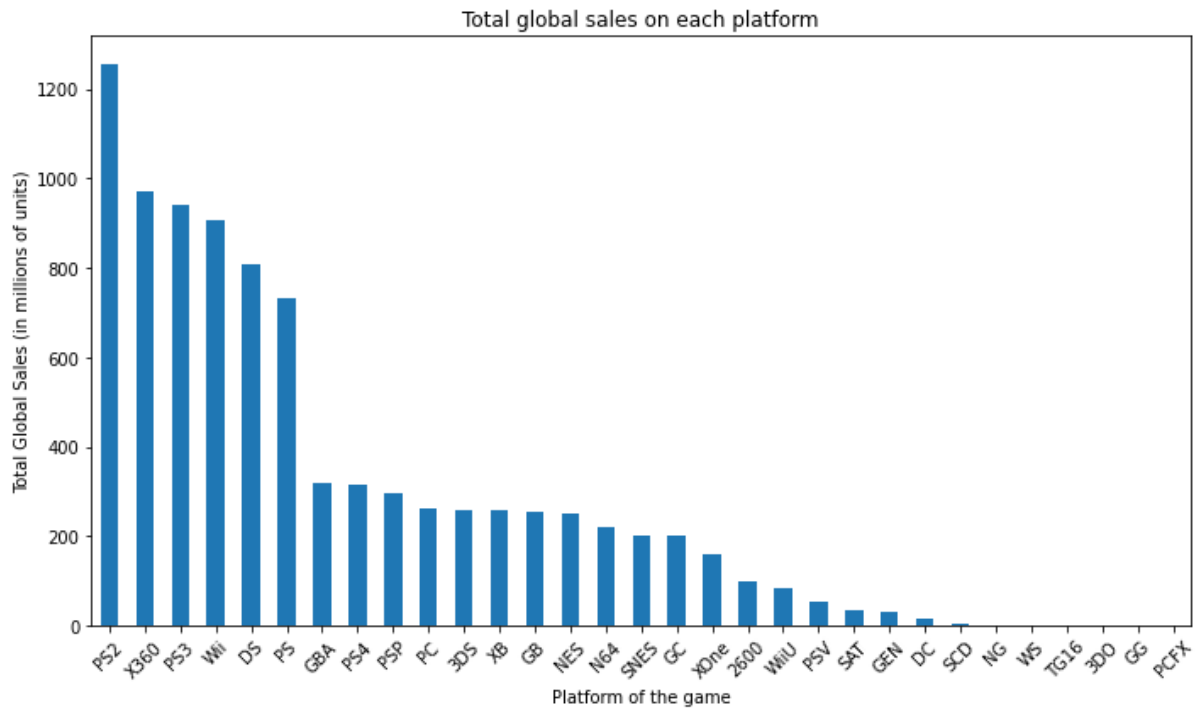
Even though Duck Hunt game was not so popular in Europe, it achieved pretty good sales number in North America. Similarly, the game Super Mario Bros was

quite popular in North America but didn't have good time in Europe. The games like Mario Kart Wii, New Super Mario Bros, Pokemon Red/Pokemon Blue and Wii Sports Resort had similar sales both in Europe and North America
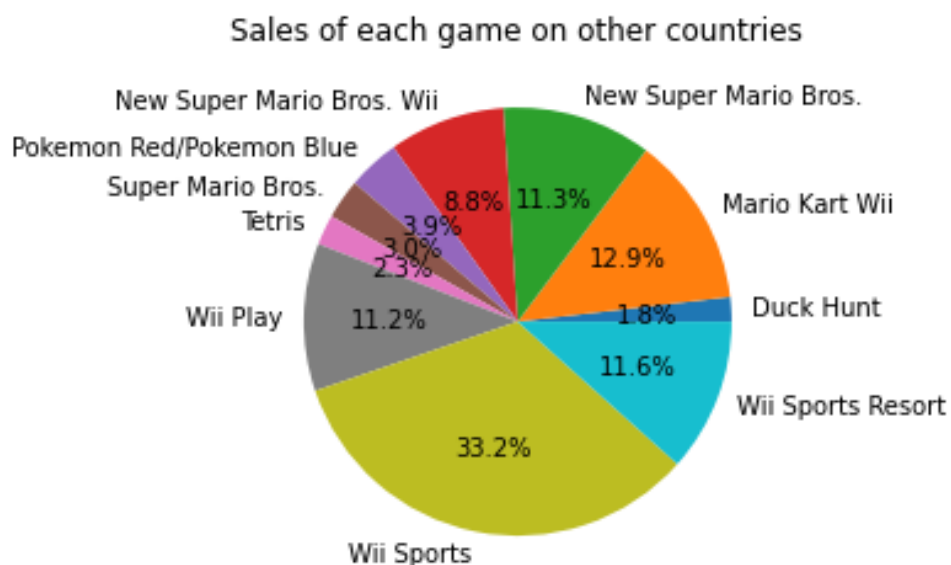

Visualization 2: BAR GRAPH (TOTAL GLOBAL SALES IN EACH PLATFORM)

To denote a bar graph, we use the same pyplot fuction. In order to visualize a bar graph we use the same dataset we used earlier with a slight difference. In here we use the X-axis to represent the platform of the game and the Y-axis used to represent sales of the game. From the graph we can clearly depict that PS2 has the maximum number of sales which is over 1200 million units, followed by X360 with around 1000 millions units, PS3 and Wii approximately the same which is around 900 million units, DS with 800 million units, PS with 700 million units, then comes GBA, PS4 and PSP with around 300 million units, follows in that order PC, 3DS, XB, GB, NES, N64, SNES, GC, XOne, 2600, WiiU, PSV, SAT, GEN, DC and finally SCD.

Total global sales on each platform

Visualization 3:PIE CHART(GLOBAL SALES OF EACH GAME)

To produce a pie chart, other sales data is taken for each game. The main objective of the pie chart was to find out which game was sold the most in other countries. To visualize a pie chart the pyplot function is used.

Sales of each game on other countries

From the above pie chart, it is clear that Wii Sports is most globally sold game with sums up 33.2% of the global sales, followed by Mario Kart Wii with 12.9%, Wii Sports Resort with 11.6%, New Super Mario Bros with 11.3%, Wii Play with 11.2%, New Super Mario Bros with 8.8%, Pokemon Red/ Pokemon Blue with 3.9%, Super Mario Bros with 3.0%, Tetris with 2.3% and finally Duck Hunt with only 1.8% sales in other countries.

CONCLUSION

In this assignment three types of visualization methods are used those are, line plot, bar chart and pie chart

Each visualization offers different perspectives of the given data such as line plot gives the number of sales of each game in NA and EU region, bar chart represents total sales on other countries achieved by different platforms and pie chart depicts the sales of each game in other countries.

The CSV file is read using pandas and graphs are plotted using matplotlib.