**Data Analysis Project 2021**

**Oleg Glingeanu X00161834**

**Video Games Sales Dataset**

For my data analysis project, I have chosen the Video Games Sales dataset. The Dataset comes with 2 comma separated files which are the PS4 and XBOX game sales datasets. The third data set is a game sale data set that is from the years 1980 to 2016 and it includes tens of other consoles which do not relate to my hypothesis and the data set is too old to work with. The 2 data sets show multiple variables which will help me carry out this project. The Variables are Name of the game, Year released, Genre, Publisher, North America sales, Europe sales, Japan sales, Rest of the world sales and Global Sales. You can see the data head below.

Graphical user interface

Description automatically generated with medium confidenceGame sales are shown in millions of sales so for example 0.60 sales would be 600 thousand game sales for a specific region.

My Research Question is: “Does the Genre the game is released affect the sales of the game” and I’m going to work on my research question by using The Global and Game variable and compare them with the 2 datasets to see what the difference is. My hypothesis is that the action genre will be the genre with the most sales as I think it’s the most popular genre 2 consoles and game sales as the PS4, and XBOX are big competitors so it will be interesting to carry out a data analysis project on this subject.

My Analysis will be carried out using PyCharm Community edition and the following python packages:

* Pandas
* Matplotlib
* NumPy