Video Game Sales Analysis

Presentation By: Emily Fagan
Last Updated: November 22, 2023

Description:

During my Data Analytics Bootcamp with Auburn University, I completed many projects. The first couple of projects involved Excel - learning how to clean data, analyze data, create and manipulate pivot tables and graphs.

Learning Excel for 8 weeks may seem unnecessary, however, 60 percent of businesses still prefer to use Excel in their workplace. (Source) Learning Excel allowed me to use my skills in my current workplace and become more efficient with data cleaning.

This presentation highlights the assignment I submitted for the second week of my analysis with the Video Game Sales Analysis dataset.

Task 1: Find the game with the highest sales in the Europe region in the Sports and Action genres

I filtered the working copy to get my results.

Grand Theft Auto V's sales were the highest overall with $9.71 million in the Action category.

FIFA 18's sales were the second highest with $8.64 million in the Sports category.

Game	Year	Genre	Publisher	North_America_S	Europe_Sales_M$	Japan_Sales_M$	Other_Sales_M$	Global_Sales_M
Grand The	2016	Action	Rockstar Games	6.06	9.71	0.6	3.02	19.39
FIFA 18	2019	Sports	EA Sports	1.27	8.64	0.15	1.73	11.8

Task 2: Specify the top three genres with the highest median sales in the North America region

I filtered the working copy of the dataset to obtain my results.

I then copied the appropriate columns and pasted them into a new worksheet.

The top three genres are Action-Adventure, Sports, and Shooter.

The top three genres with the highest median sales in the North America region are Action-Adventure at $0.26 million, Sports at $0.23 million, and Shooter at $0.22 million.

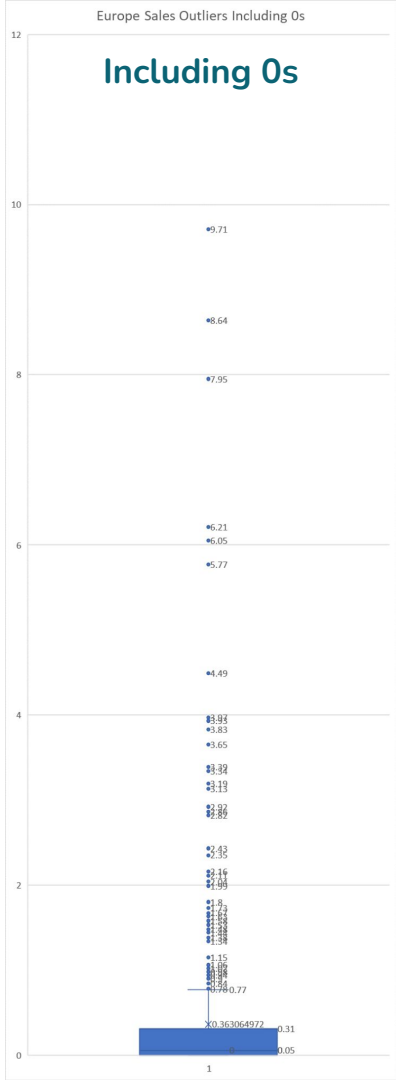
Genre	MedianNorthAmericanSales($M)
Action-Adventure	0.26
Sports	0.23
Shooter	0.22
Fighting	0.105
MMO	0.09
Platform	0.09
Music	0.08
Action	0.07
Role-Playing	0.06
Racing	0.055
Adventure	0.05
Misc	0.05
Party	0.045
Simulation	0.04
Puzzle	0.025
Strategy	0.02
Visual Novel	0

Task 3: Are there any outliers in the Europe region? If yes, visualize the outliers

The two box and whisker plots are made in Excel. In a real-life situation, I would not use Excel for my plots because of the size of this particular dataset. For the purposes of this assignment I included them in Excel format.

The outliers in the Europe Sales category (including 0s) are any sales above $0.77 million because the upper fence is any value above 0.77.

The outliers in the Europe Sales category (excluding 0s) are any sales above $1.18 million because the upper fence is any value above 1.18.



Task 4: Identify the publisher with the highest average global sales. Calculate the median value of the global sales for the identified publisher and compare it with their average global sales.

First I created a pivot table to include the publisher in the rows and the average of global sales in the values column. I then sorted the pivot table in descending order by average sales. I discovered that the publisher with the highest global sales is Rockstar Games with an average of $11.31 million in global sales.

Continued on next slide...

Pivot table screenshot

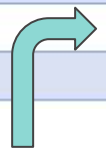
Publisher	Average of Global_Sales_M$
Rockstar Games	11.31
Blizzard Entertainment	3.68
Warner Bros. Interactive	3.28
Electronic Arts	2.748
EA Sports	2.641666667
Activision	2.587142857
2K Sports	2.48
Hello Games	2.26

Task 4 Continued

To find the median of global sales, I went back to the working dataset and filtered the publisher to show only Rockstar Games. I sorted the global sales column in descending order.

The median value of Rockstar Games' Global sales is $13.94 million. **This means there is a left skew distribution.**

Game	Year	Genre	Publisher	Global_Sales_M$
Grand Theft Auto V	2016	Action	Rockstar Games	19.39
Red Dead Redemption 2	2020	Action-Adv	Rockstar Games	13.94
L.A. Noire	2019	Adventure	Rockstar Games	0.6



Median sales of Rockstar Games

Task 5: Calculate the yearly mean and median global sales. Determine whether there is any skewness in the data.

In each year, the average global sales are higher than the median sales. That means there is a **right skewness** in the data for each year 2015-2020.

Year	Average Global Sales		Median Global Sales
2015	1.7975	>	1.71
2016	1.48525641	>	0.59
2017	0.963956835	>	0.22
2018	0.675223881	>	0.13
2019	0.547364017	>	0.1
2020	1.434285714	>	0.29