

# Guided Project from CareerFoundry: Analysis of Video Game Sales Data

## Prompt:

You are a data analyst hired by GameCo (A fictitious company) to analyze the data from the inception of video games to the “present” (this project would take place in early 2017, so data ranges from the 1980s to 2016.) The company currently has an understanding that all video game sales from each region stay consistent and wants to create a marketing campaign focusing their relative efforts on each region. Your task is to use the “vgsales” data (see attached excel sheet) and analyze the data to understand the trends within the data and either confirm or challenge the understanding held by GameCo.

## Purpose of Project:

To highlight Excel Skills and showcase Data Analysis Practices in a circumstance where potential Stakeholders in an industry demand a solution derived from data. As such, this project serves more of a skill showcase rather than an actual report on video game sales. For the actual report on video game sales analysis, see attached PowerPoint Presentation “Video Game Sales Report 2017”.

## Skills Highlighted:

### Excel

- Data Cleaning (*Removal of Duplicates*)
- Data Formatting (*Header row formatting, cell width formatting, sort, filter*)
- Data Aggregation (*Pivot Tables, Sliders*)
- Data Visualization (*Pie Chart for comparison of categories, Line Charts for time-based analysis*)

### PowerPoint

- Presenting Findings (*Uploading static charts from Excel with context*)
- Making Recommendations (*Determining best course of action based on analysis*)

## Initial Info on Data:

Data originated from Kaggle

(URL: <https://www.kaggle.com/datasets/kedokedokedo/vgsales>)

Rows: 1 Header Row + 16324 Observations, Columns: 11

Column Names (*And Type of Data*):

- Rank (*Categorical, can serve as Row Identifier*)

- Name (*Categorical*)
- Platform (*Categorical*)
- Year (*Categorical, could be treated as Numerical if necessary*)
- Genre (*Categorical*)
- Publisher (*Categorical*)
- NA\_Sales (*Quantitative*)
- EU\_Sales (*Quantitative*)
- JP\_Sales (*Quantitative*)
- Other\_Sales (*Quantitative*)
- Global\_Sales (*Quantitative*)

## Procedure of Analysis:

1. Data Cleaning
    - a. Formatting Headers
    - b. Adjust Cell Widths
    - c. Sort by Rank
    - d. Include Filters
    - e. Freezing Top Row
    - f. Removal of Duplicates
    - g. Removal of Missing Data in Rows or Columns
    - h. Any Other Data Cleaning Practices
  2. Exploratory Data Analysis
    - a. Exploring Counts of Unique Values for Categorical Data
    - b. Exploring Statistical Figures for Quantitative Data
  3. Answering the Prompt
    - a. What to analyze to answer prompt
    - b. Further Exploration
    - c. Results and Recommendations
  4. Telling the Story
    - a. Design
    - b. Visualizations
    - c. Flow of Content
    - d. Transitions and Animations
- Reflections on The Project (Not technically part of analysis)
    - How to Improve upon Analysis
    - Why this project was created

## Data Cleaning

In trying to understand where the data comes from, we find out that the data from Kaggle is something that could be scraped from the website [www.vgchartz.com](http://www.vgchartz.com), but since missing

data or duplicate data may appear, we need to clean through the data to make sure our findings are as accurate as possible as well as readable for anyone who wishes to understand the data presented.

When we first see the data (see first tab “vgsales” in attached “vgsales” Excel file), we notice that the data has many characteristics making it difficult to read. The names column does not have the full name shown as parts of the names are hidden within other data. The data is sorted by year of publication with the “Year” column in the middle instead of all the way to the left, which leaves the first “Rank” column showing numbers in a seemingly random order, which is distracting. To fix this, we first format the headers by increasing the font of the first row to size 14, putting the text in bold, highlighting the header with a light blue color, and adjusting the cell widths by clicking the top right corner to select the whole sheet and double clicking the line between two columns to automatically adjust the column widths. After this, I decided to sort the data by Rank instead of year for easier readability, including filters for any exploration later in the future, and froze the top row so that even when scrolling down the sheet, you would not lose track of what each column represents.

From here, we must view the data to see if there are any rows of missing data or duplicate rows. To remove duplicate rows, I went to the data ribbon, clicked the remove duplicates button, and it gave me options of what to consider when looking for duplicates based on the columns available. If I selected all columns, no data is a duplicate, and when we only remove rank from consideration, which could be inferred to be a unique value for each row, no rows get removed either. Removing any other columns from consideration would have resulted in removing unique data and therefore skewing the data by unintentionally only taking a part of the whole data when attempting to analyze the data as a whole. As such, we conclude that there are no more duplicates. After that, to observe any rows that contain missing data, the filter function can show whether or not certain values are N/A values or blank values. After checking several columns, we see that all the quantitative data does not contain blank values, and while there are many circumstances where N/A values appear in the “publisher” column with the “misc” value appearing subsequently in the “genre” column, there is enough information in all other columns from those rows that warrant the consideration of those rows overall. Nothing needs to be deleted.

Any other data cleaning practices would be to check through data to make sure that formatting was normalized, meaning it all had the same format whether it be a date format, numerical format, percentage format, currency format, etc. From looking at the numerical data, we see that the number format stays consistent with each number showing itself up to at most two decimal places. From looking at the categorical data with relatively little values compared to the more unique identifiers, there don’t appear to be any misspellings or case adjustments that would potentially differentiate the same value as two distinct values. (One that may have been considered was 3DO with 3DS in the Platform column, but as it turns out, the 3DO was a platform that existed in the 1990s while the 3DS first showed itself after the year 2010, so a mistake could not have been made there, and as

such, they are inferred to be two different platforms.) We have checked that there are no duplicates, no rows almost completely full of missing values, no misspellings, and consistent formatting in the data values. For the purposes of our analysis, our data is clean.

## Exploratory Data Analysis

Some of the first things we need to do when analyzing data is exploring the data that you are working with, which involves observing statistical figures such as the mean, median, minimum, and maximum while observing any potential outliers for quantitative data and observing counts and identifying modes for categorical data. Some information is available when you aggregate the data using pivot tables while other data is available by using built-in statistical functions from Excel. We will go through each column and work from there.

Rank:

Lowest Value: 1

Highest Value: 16600

Number of unique values: 16324

Highest (and lowest) count of a value: 1

What can be inferred: From what used to be a list of 16600 different video games, we may have many missing observations. The number of missing values would account for 1.66% of the 16600 different video games that supposedly should be listed. A more exhaustive search for the missing data would help to fill the gaps and give us a better representation of the data, but this is data based on 2016. If we attempted to find the data, we would also gain data all the way into 2025 by the time of creating this document, which would drastically affect the potential ranking overall. For the purposes of this project set in 2017, we will work with the data given to us.

Name:

Number of unique values: 11,358

Name with Highest Count: "Need for Speed: Most Wanted" (Count: 12)

Lowest Count of a name: 1

Number of Names with count of 1: 8,642

What can be inferred: We have several circumstances of names appearing more than once. One could expect the names to be all unique values, but as it turns out, the data separates games as observations not just based on how it is titled, but also on which platform it was created and what year it was created. For instance, in "Need for Speed: Most Wanted", you can see that the game was created for the same platform once in 2005 and the other time in 2012, and the rest of the games listed as "Need for Speed: Most Wanted" take all other available platforms into consideration.

Platform:

Number of Unique Platforms: 31

Platform with Highest Count: DS (Count: 2,133)

Platforms with Lowest Count: PCFX and GG (Count: 1)

What can be inferred: The DS has the most unique games available for play, which tells me that there could have been a huge campaign on Nintendo's end (The creator of the DS) to publish as many games as possible for the DS, which would be a bigger campaign than most other gaming consoles. Looking at the years of the observations that hold the PCFX and GG as their Platform, I saw that they each were used during the 1990s. Doing some deeper research from the internet, I found that the GG refers to the Game Gear, while the PCFX is its own system that would work as the successor to PCs as a gaming unit. Unfortunately, the latter did not receive much success due to its expensiveness compared to other systems that existed at the time. As for the former, I imagine there were several other games that were available for the GG that simply aren't listed here. Perhaps that information could be found within the missing data alluded to earlier.

Year:

Oldest Year: 1980

Most Recent Year: 2016

Year with Highest Count of Video Games: 2009 (Count: 1433)

Year with Lowest Count of Video Games: 1980 (Count: 9)

What can be inferred: Looking at a table of counts for years alone, you notice that the number of video games has a rather consistent incline from 1980 to 2009, and from 2010 onwards, you notice an interesting decline. This finding is apparent if you were to look at the numbers by themselves but becomes even more apparent when this information is available as a line chart.

Genre:

Number of Unique Genres: 12

Genre with Highest Count: Action (Count: 3,252)

Genre with Lowest Count: Puzzle (Count: 571)

What can be inferred: There seemed to have been an overall desire by many companies to create action games that can often be described as dynamic and/or fast paced.

Conversely, there seemed to have been an overall desire by many companies to not make as many puzzle games that can often be described as games that force you to be slow and methodical. Also, something to note when performing further analysis is that since we have very little unique genres compared to the other columns so far, it may be easier to categorize the data according to genre to make observations than if we were to look at any other categories as variables.

Publisher:

Number of Unique Publishers: 577

Publisher with Highest Count: Electronic Arts (Count: 1339)

Number of Publishers with Lowest Count of 1: 192

What can be inferred: It is surprising to see that there were this many publishers with only one publication on record. It would be interesting to see what platform they published their games to see if there was any relationship between the success of the platform the publishers published for and the success of the publishers themselves.

Sales Data:

The Categorical Data can provide little when it comes to numerical analysis, but the same cannot be said for quantitative data such as the sales data for North America, Europe, Japan, Other Countries, and Global. We were able to determine using Pivot Tables and other statistical functions in Excel the following:

- Average (found through Pivot Tables)
- Standard deviation (calculated as population standard deviation using STDEVP function as opposed to sample standard deviation that takes the number of observations into account to calculate standard deviation)
- Median (Used MEDIAN function as there was not an easily accessible field for Median in Pivot Tables)
- Minimum (Pivot Tables)
- Maximum (Pivot Tables)
- First and Third Quartiles (calculated using QUARTILE function)
- Interquartile Range (IQR calculated as  $Q3 - Q1$ )
- Thresholds for outliers in the data (calculated as  $Q1 - 1.5 \times IQR$  for bottom outliers and  $Q3 + 1.5 \times IQR$  for top outliers, though the former calculation isn't shown in Excel file due to understanding that the minimum and Q1 values are already extremely close to each other, if not already the same)
- Number of outliers in the data (Looked through the data after sorting according to the column from largest to smallest until we found the last value that surpasses the threshold for outliers, then used row numbers to reference how many values exceed the threshold. I understand that there are more efficient ways to find this information, especially when dealing with big data, but this method felt somewhat appropriate for the small amount of data listed here. Another method I could employ would be to filter out any rows that had their value within the threshold, then use the COUNT function to count how many rows remain which are by default, outliers. Alternatively, I could use the COUNTIF function to count how many rows have values that exceed the threshold. This method is probably the simplest and most efficient method out of the others I just listed in terms of figuring out the information in the least amount of time, but would still probably take some

computing power to gain the information given the amount of steps necessary for the computer to parse through the information and check whether or not it satisfies the criteria.)

Here are the results of finding the above information for the sales figures from the attached Excel file:

Sales	Sum	Average	St Dev	Min	Q1	Median	Q3	Max	IQR	Outlier Cutoff (Right)	Count of Outliers
NA_Sales	4333.43	0.265464	0.821633	0	0	0.08	0.24	41.49	0.24	0.6	1660
EU_Sales	2409.12	0.147581	0.508793	0	0	0.02	0.11	29.02	0.11	0.275	2054
JP_Sales	1284.25	0.078673	0.311575	0	0	0	0.04	10.22	0.04	0.1	2407
Other_Sales	789.01	0.048334	0.189896	0	0	0.01	0.04	10.57	0.04	0.1	1645
Global_Sales	8820.31	0.540328	1.565812	0.01	0.06	0.17	0.48	82.74	0.42	1.11	1812

It is rather striking that there is a very large number of outliers in the context of the data. The outliers alone appear to consist of somewhere between 10-15% of the entire data set. Additionally, it is very clear to see that the data is skewed right, which means that the outliers of the data pull the average much higher than the median to a significant degree.

One thing that I noticed immediately when looking at the video game sales figures directly from the clean data was that the first ranked game “Wii Sports” appeared to have game sales that beat out all other games by a significant margin. As such, it felt appropriate to see how the data would change if this one observation was taken out of the dataset. Here is what I found:

Sales w/o Wii Sports	Sum	Average	St Dev	Min	Q1	Median	Q3	Max	IQR
NA_Sales	4291.94	0.262938	0.755647	0	0	0.08	0.24	29.08	0.24
EU_Sales	2380.1	0.145813	0.455865	0	0	0.02	0.11	12.88	0.11
JP_Sales	1280.48	0.078446	0.310242	0	0	0	0.04	10.22	0.04
Other_Sales	780.55	0.047819	0.178123	0	0	0.01	0.04	10.57	0.04
Global_Sales	8737.57	0.535292	1.427567	0.01	0.06	0.17	0.48	40.24	0.42

And here is a table of how much each of the aggregated data values changed due to the removal of Wii Sports:

Difference made in...	Sum	Average	St Dev	Min	Q1	Median	Q3	Max	IQR
NA_Sales	41.49	0.002526	0.065987	0	0	0	0	12.41	0
EU_Sales	29.02	0.001769	0.052928	0	0	0	0	16.14	0
JP_Sales	3.77	0.000226	0.001333	0	0	0	0	0	0
Other_Sales	8.46	0.000515	0.011773	0	0	0	0	0	0
Global_Sales	82.74	0.005036	0.138244	0	0	0	0	42.5	0

You will notice that changes are made in the averages and standard deviations on all aspects, but the max only changed for NA, EU, and Global sales when Wii Sports was taken out, which tells us that Wii Sports was only the top game in sales for NA and EU, but for JP and other parts of the world, other games were considered top sellers instead.

The biggest thing to note is that the minimum and the first three quadrants are not affected by the removal of the biggest value in the data, but the sums, averages, and standard deviations were. This is in-keeping with the property that averages and standard deviations are affected by outliers much more than the first and third quartiles and medians are.

## Answering the Prompt

Coming back to the original prompt, we want to either confirm or challenge the following notion: “All video game sales from each region stay consistent.” This can be viewed either from an absolute perspective of values or from a relative perspective of proportions. We will attack the prompt both ways.

If we are looking for the consistency of sales, we want to understand things from a temporal perspective, which means we will be grouping the data by year and observing trends by following changes from each year. This will require the creation of line charts. We will create a line chart for sales from each region as well as for the proportion of sales from each region.

Looking at the line chart of sales within regions over time, between the years of 1980 and 1995, North America’s sales have oscillated greatly while all other regions have stayed rather consistent; after which point, there is a drastic rise in Video Game Sales for N.A. and Europe with a slight rise for Japan until around 2010, when sales start to decline for all regions.

Looking at the line chart of the proportion of sales within regions over time, there is an oscillation between the proportion of North America’s Sales and Japan’s Sales before the 2000s but is then followed by relative consistency within sales for both regions until recent years when North America started to drop again. Meanwhile, the proportion of Europe’s sales and the sales of other regions have consistently shown an increase. This challenges the notion that sales have remained consistent over time. Note that the oscillation for Japan is more so a reflection of how dynamic North America’s sales were rather than Japan actively catching up to North America in sales.

After seeing the line chart that looks at the values and proportions changing over time, we understand that overall, not all regions’ sales remain consistent. There are regions that show consistent fluctuations relative to the rest of the world and some regions that are progressing relative to the rest of the world. We could leave it at that, but a data analyst must be inquisitive. There is more we can analyze. For instance, we can observe the trends of values and proportions for each genre to see if there is any consistency between trends in each genre and the overall trends. Let’s look at the top 5 genres. To figure out which genres those would be, let’s analyze the global sales per genre and compare values and proportions from there. Comparing values, we can use a bar chart. Comparing proportions, we can use a pie chart. Making both, we understand that the top five from highest to lowest are as follows: Action, Sports, Shooter, Role-Playing, and Platform. From here, we can



compare the trends of regional sales of each genre to the overall trends and see where each region is recently at in terms of sales.

Looking at the line charts from the Action Genre, we see that trends for North America and Europe both seem to correlate well between action sales and overall sales; however, it appears that in most of the 2000s, other regions slightly yet more consistently dominated over Japan in action sales instead of having oscillations in performance. In 2016, NA, EU, and JP each made 30% of Action Sales with Other Regions at 10%.

Looking at the line charts from the Sports Genre, we see that there is a low number of sales until around 1996 when almost all regions experienced a rise, similarly to overall sales trends. Some major differences, however, is that North America and Europe saw two big jumps in sales in 2006 and 2009. Additionally, Japan dominated over most of the 1980s until the middle of the 1990s, then saw a decline and became the lowest in sales, even to other regions. In 2016, while everyone sees a decline, relatively speaking, Europe is in a better place than everyone else.

Looking at the line charts from the Shooting Genre, we see that sales from the 2000s and onward seem to be dominant in North America and Europe, which correlates with overall sales trends. A big difference is that there was a huge jump in NA sales in 1984, and then very little activity between then and the middle of the 1990s as opposed to small growth or fluctuation. Additionally, Japan is also seen as very weak in Shooter sales throughout time, especially during the 2000s and onward compared to overall performances. NA and Europe are relative to each other at around 40% of Shooter sales in 2016 while other regions are beating out Japan.

Looking at the line charts from the Role-Playing Genre, we see that North America and Europe see increases starting from the middle of the 1990s with all regions seeing decline in recent years, which correlates with overall sales trends. A huge difference to note is that Japan has seen tremendous success in RPG Sales before the middle of the 1990s and remained competitive with NA and EU Sales even afterwards. Other regions outperformed only in the beginning in 1986, but then consistently stayed at the bottom. In 2016, Japan is at the top with NA and EU relative to each other.

Looking at the line charts from the Platform Genre, we see that Japan was somewhat competitive in sales before the 2000s, but then Europe was competitive afterwards. This pattern tracks with overall sales trends. A key difference we see is that North America hit their prime during the middle of the 1980s in Platform Sales, and since then, plummeted near the bottom and oscillated towards growth into the 2000s with decline in recent years, coming close to but never hitting the records they set in the 80s. As of 2016, North America and Europe are relative to each other while Other Regions have overtaken Japan for now, but we could easily see an upward oscillation from JP given their track record of oscillation in sales between years.

From these observations, we notice a couple of things:

- While all of the world seems to be seeing a decline in sales, relatively speaking, Europe has been consistently growing in sales overall, now starting to be relative to North America, which has been dominant in video game sales for a decent amount of time.
- Looking at the genres individually tells different stories than if you were to look at the overall sales figures by themselves.
  - For Action games, North America, Europe, and Japan are currently in relative positions, so marketing can focus equal amounts of effort on all three but expect Europe to respond better considering the more consistent growth relative to most others.
  - For Sports games, Europe is starting to pull ahead relatively speaking, so if we could bolster sales on that front with marketing campaigns focusing on that region.
  - For Shooter games, while North America has historically been the most dominant in this Genre, Europe is starting to come about. Either we could attempt to reinvigorate sales on the North American front by focusing marketing campaigns there, or we could bolster sales on the European front by focusing Marketing efforts there.
  - For Role-Playing Games, Japan has been historically the most dominant in this Genre while staying competitive within recent years. If we aim to focus a marketing campaign for Japan, it would be most beneficial to focus on marketing for Role-Playing Games since they appear to be the most successful in that region, especially as of 2016.
  - For Platform Games, we saw an outlier during the 1980s in a year that sold so many platform games in North America, and to this day, that record has not been broken. I would advise for further analysis on that year of platform sales specifically to understand the circumstances that led to that success and see if it is possible to market new platform games in North America to possibly replicate or even surpass that level of success.
- If our desire was to understand how to approach each region, focus on the genres that have proven to be or will be more marketable.
- If our desire was to see which region we should focus efforts on more than any other region, I would suggest focusing on Europe specifically because of it's increasing relative success.

## Telling The Story

It's one thing to be able to relay this information to your stakeholders, however a big part of being able to communicate your findings is by finding ways to make it easier to connect with your audience. As such, a plain presentation is not good enough. We need to focus on the details.

The first detail to focus on was the design of the PowerPoint itself. The design I chose was the “Circuit” design, which is a gradient of blue that has a circuit like pattern at the edges of each slide. Considering that my stakeholder was GameCo, a Video Game Company that is by default all about technology, I figured a design that incorporates the idea of technology was a must.

The next detail after this was the data visualizations. I knew I was going to need to input the visualizations I created in Excel into the PowerPoint itself. When I created the visualizations in Excel, I made sure to include the title of the chart, the titles of the axes, and spaced the visualizations in a way that the years are not written overlapping each other to allow for easier readability. From there, I colored the line series based on regions, and found through Google that North America is often associated with Green, Orange, Red and Yellow, Europe is often associated with Blue, Japan is often associated with Red and White (their flag colors), and I figured that Other Regions should be given a neutral color so as to not associate the color with a specific nation when covering what is essentially all other nations, so black or gray would be my go to colors. Ultimately, I settled on Green for NA, Blue for EU, Red for JP, and Black/Gray for Others. From there, I realized that there is a possibility some viewers of this presentation could be color blind and decided from there that I needed to make some colors darker while others would be lighter. Japan’s red is a specifically standard color of red, and altering this shade of red did seem like it would cause some confusion, so I left JP to hold the standard color of Red, which means that NA and EU would need their color shading to be changed accordingly. Considering that NA was the most dominant in sales for the most part, it would be best to darken the green and brighten the blue. Lastly, for Others, I decided that black was the best choice in case the level of shading for a gray would be too like the shading level of any of the other colors, which would make seeing it way too confusing. Those were the changes I made on Excel, but a problem came when I tried to incorporate those graphs into PowerPoint. If it was incorporated in a way where the design merges with the design of the PowerPoint, the color scheme changes completely for the visuals, and the intent behind the colors gets lost in translation. The best course of action at that point was to paste the images of the charts with the white background that comes with it.

Next is the flow of the story being presented. Ultimately, to make the information more digestible, it needs to flow within a structure. To do this, first, after the title slide of the presentation, we would state why this presentation is being given by reminding the stakeholders of what was asked of me, which was to perform analysis of data to either confirm or challenge the current understanding of the company and give recommendations based on our findings. From there, after a disclaimer on the quality of analysis based on the available data provided, I show a slide that holds the sales figures in absolute and relative terms for overall video game sales, coupled with some notes on the side to guide what details of the visuals should be given consideration and what we can infer from those details. After giving the analysis approached from the overall perspective, we guide the story into perspectives by genre using the bar chart and pie chart for the genre analysis, which would segway into analyzing the trends by genre and comparing those

trends to the overall trends we bring to attention in the beginning. Once all those trend analyses are presented, we will summarize our findings in the last slides and give out various recommendations so that they have some options to work with. The stakeholders can decide for themselves how they wish to proceed based on those recommendations and what their ultimate goals are.

And lastly, the slide transitions and animations. When given the PowerPoints directly, this will not be something of importance. However, in the middle of the presentation, it is important to guide your audience through the information one piece at a time for easier digestibility. As such, we don't want to showcase all the information at once whenever a slide is shown, so we included animations that allow the information to come in one part at a time. Transitions between slides also assist in the pacing of the presentation, instead of an immediate switch needed in terms of paying attention from one slide to another.

## Reflections on The Project

With this, the essential part of our project has been completed. As stated earlier, the purpose of this project is to demonstrate Excel skills, PowerPoint skills, and Data Analysis practices in terms of data cleaning, data exploration, answering the concerns of stakeholders, and communicating our findings. I feel that this project does exactly that, so I would consider this to be a success.

There is always room for improvement, however. One thing that should be addressed is the missing datasets. An attempt could have been made to rediscover the data that was missing from the dataset given, but one of my biggest concerns was how rankings would be affected by the addition of new information from 2017 and onwards. Thinking about it now, one way to navigate through this would be to filter the data scraped from the original website to only account for games from 2016 and before and manually create a ranking system by sorting it according to global sales and providing the rankings from there, assuming all the information from the website still remains and that the end result of the many records matches or exceeds the number 16,600.

The biggest reason this was not done was because the biggest purpose of this project was to highlight specific skills such as Excel, PowerPoint, other data cleaning and analysis practices, and my mental thought process when performing and presenting my analysis for Stakeholders using the most basic tools. There are other ways I could have presented my insights, such as with Tableau and Microsoft Power BI. There are other ways I could have analyzed the data, such as by uploading these into a database and querying through SQL. These skills will be highlighted in other projects which will be published soon enough. This project simply highlights what I am capable of with Excel and PowerPoint, but not what I am limited to.

Why video game sales? There are two reasons. The first: Many years ago, shortly after graduating with a degree in Applied Math and a minor in Stats, I was interested in Data

Analysis as a career, so I purchased a limited-time program for Data Analytics Certification from CareerFoundry, and my first project to introduce me to Data Analytics had me observe video game sales. I was able to complete the project in 2021, as well as a few other projects which tackle Tableau and SQL, but I never finished the program and got my certification simply because I had burnt out and lost the drive to seriously continue after having went through the rigorous experience that was my college education. I found myself comfortable in the workspace I was in at the time and still am in now, but after many years of living, having a wife, and having a son, I wanted to be better for myself, which includes correcting the mistake I made in putting off a data analytics career for so long.

In trying to submit the project completed long ago into my portfolio, I found that a lot of relevant files were missing, including the original dataset given by CareerFoundry from all those years ago. I managed to find a dataset that closely mimicked the dataset from Kaggle, but I understood that this was still a different dataset altogether, so that meant that in order to reference this dataset, I had to reperform the analysis that I performed from scratch. Now that I understood how the analysis was supposed to go after doing it once before and having more exposure to analysis practices, I was able to perform the analysis at a much faster rate. This version of the project took 14 active hours (three days total) to get to where it is now versus the month it took doing it the first time when I spent most of my time trying to learn the Excel skills needed to use it as an effective analysis tool. I will acknowledge that the speed of this project was in part due to how I had performed it in the past and knew what to do and what to look for, which is why I will still acknowledge this as a guided project from CareerFoundry. That being said, this was also a decent opportunity to come back to the basics from how I was introduced to this. I could have chosen another dataset altogether, but I wanted to make the most out of the work I put in a long time ago. That's the first reason. The second? I used to grow up playing Mario and Sonic when I was much younger, so I have a connection to video games from my childhood. Why not analyze video games again and submit this project to my portfolio? All in all, I believe I did myself justice with what I have presented, but there is so much more data to analyze in several ways, and so much more of my abilities to showcase. I look forward to continuing to show you what I am capable of.

Last updated: 07/11/2025