

Step 1: These are the steps that I conducted to obtain insights based on the given data:

A. Obtained the data

- Obtained the data from a reputable source (VGchart) – this information was already given through the exercise, but as an avid gamer, my understanding is that this is one of the credible sources for game sales data.

B. Reviewed the data to understand

- What does each field represent? Most of the fields were self-explanatory; however, the unit of the quantitative fields were ambiguous at the start. This had to be clarified first (e.g., are sales in \$'s or number of units sold? Are the figures rounded to the nearest thousands or millions?)
- Reviewed the data collection methodology to understand any limitations with the data and noticed that the data collected is based on official shipment/sales data that is made available by developers and publishers only.

C. Cleansed the data

- For data completeness - there were incomplete fields (blank or N/A fields) or data items that do not have the complete set of data (e.g., 2017 and 2020 data only had few entries), such items were removed applying the analyst judgment.
- For data accuracy - While the accuracy of the data is mostly dependent upon the collector as there are no other data sets to cross reference and verify the accuracy of the data presented; however, a minor check was applied to ensure the accuracy and consistency of the data. The sum of all regional sales should equal to the global sales figure that is already available and this was not always the case based on the initial data set given and for the most of instances noted, this inequality to was due to missing sales figure for the one particular region's sales figure. Such instances were corrected by calculating the missing regional sales figure by subtracting the sum of all available regional sales figures from the global sales figure.

D. Analyzed the data

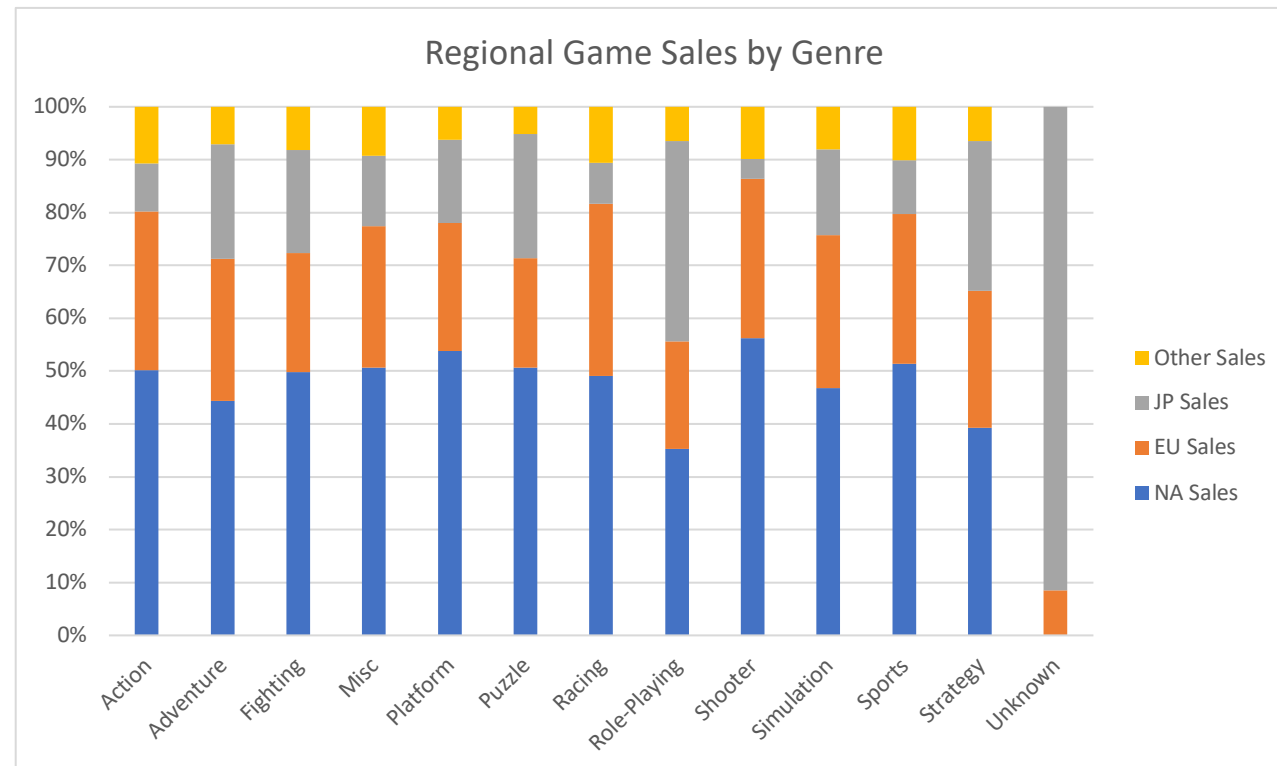
- With the clean data available, an Explanatory Data Analysis was performed to understand the overall background and then, a couple of descriptive analysis were conducted through the exercises available, which provided additional insights to what I initially expected as an avid video gamer. The two key expectations that I had were:
- **Expectation 1:** NA has been and is the dominant market for the game sales
 - a. Steps conducted to obtain insights over this expectation:
 - i. Using a PivotTable's functionality, added calculated fields to obtain each region's % contribution to the overall global sales and grouped each region's contribution %'s by each year

- ii. Created a line graph by plotting the three variables (NA, EU, JP game sales contribution to the global sales) for a visualization to have a better glance at the overall trend changes. Refer to the chart below for details.



- b. Based on this insight, I noted that my initial expectation was not right and revised the expectations. Following insight details were noted:
- North America has not been always the dominant market of the game sales. The NA's contribution of the game sales has been dominant for certain years and has been dominant from 1996 to 2015, but their contribution % has been volatile over the years.
 - Interestingly EU market has not been less volatile in terms of the game sales and their contribution % has been steadily increasing over the years. In 2016, EU market accounted the most in terms of the global game sales
 - Another interesting insight noted based on this chart is that generally game sales tend to be inversely related between NA and JP game sales

- **Expectations 2:** Based on the region-specific preference of the game genres, Japanese market will be the dominant contributor of the role-playing game sales
 - a. Steps conducted to obtain insights over this expectation:
 - i. Using a PivotTable's functionality, added calculated fields to obtain each region's % contribution to the overall global sales and grouped each region's contribution %'s by genre
 - ii. Created a 100% stacked bar chart by stacking 4 variables in one single bar that adds up to 100% for each genre to review the trend of regional game sales by each genre. Refer to the chart below for details.



- b. Based on this insight, I was able to confirmed that my initial expectation was correct where the Japanese market contributed the most in terms of role-playing game sales.

Step 4:**Types of Visualizations Used with Reasons:**

- **Stacked Bar Chart** for Number of Games Sold Grouped by Selling Buckets: to illustrate the mode (central tendency) of the data without using the term “mode”
- **Pie Charts** for Sales by Genre and by Platform – to provide quick, at a glance, representations of the data by category groups
- **Line Chart** for Regional Accountability to Global Sales – showing the regional sales trend in a quick, easily digestible format
- **100% Stacked Bar Charts** for regional sales contribution for each genre – to illustrate the genre popularity by region in a quick manner
- **Stacked Bar Charts** for Platform preference based on regional sales and Console preference (a breakdown of platform preference) – to illustrate the material platforms where the most games sold at a glance. Platform chart was grouped in high-level broad categories (Handheld, PC, and Console) and the Console chart was the show the detail breakdowns of the particular (material) platform group that the management might want to understand further
- **100% Stacked Bar Charts** for platform-level sales contribution for each genre – to illustrate the insights over specific genre preferences by different platform users in a quick manner
- **Stacked Bar Charts** for Number of Titles Released vs. Copies Sold by Genre and by Platform – to illustrate the profitability (# games sold per release) based on genres and platforms for easy digestion

Key Similarities and Differences – between the visualizations in Step 1 vs. visualizations in the presentation:

- **Similarities**
 - Two charts mentioned in Step 1 were used in the final presentation with modifications to show the (1) trend – line chart and (2) composition breakdown (100% bar charts). The main intention for the chart usage did not change
- **Differences**
 - The major difference for all charts used in the presentation is the scope limitation. The charts used in Step 1 displays the data for all periods (1980-2016); however, for the presentation, all data was limited for the period 2012-2016 based on the judgmental criteria (analysis based on the game sales occurred during the current console generation) to have the insights more relevant.
 - Applied cosmetic changes to the charts in the presentation for better legibility and for better illustration (data labels)
 - Additional charts, apart from the ones listed in Step 1, were added to explore availabilities of different insights to meet the management’s objectives.
 - Numerous stacked bar charts were used and considerations were given on whether to use the original stacked bar charts vs. 100% stacked bar chart (show the materiality and composition vs. purely composition)