**Deepak sai Paty**

**INFO 5810**

**Fall 2020**

**Video game sales analysis on genres, regions, and sale year**

**Data Goals**

1. Highest global video game sales were done in which year?
2. In last three decades, Which region has the highest sales in ‘racing’ genre, and which has the lowest sales in ‘fighting’ genre? Suggest a genre we need to concentrate with respect to each region to increase the sales?
3. Which platform made good sales in 1996 and in 2014. If a publisher would like to invest more, which platform can we suggest in future? (Globally)
4. Considering from 2000-2015, global video games sales has lowest average in which year?
5. Highest games were released by publisher named ‘Activision’ in which Genre? Name the other genre in which Activision can invest more in future?

**Pivot Table**

* Pivot tables is a functionality provided in excel that can be applied on the dataset to solve many business questions.
* We can perform many operations such as sort, total, count on the table data.
* We can also categories the data into different groups (depending on data), count the items in each group, sum the items in each group, finally we can also provide an average of each group.
* We can transform rows into columns and columns into rows.
* It supports analysis even with large amounts of data.

**Steps to create Pivot table**

1. Initially select any cell of the dataset and then click on the insert tab on top of excel.
2. Click on the pivot table then a below dialogue box opens,

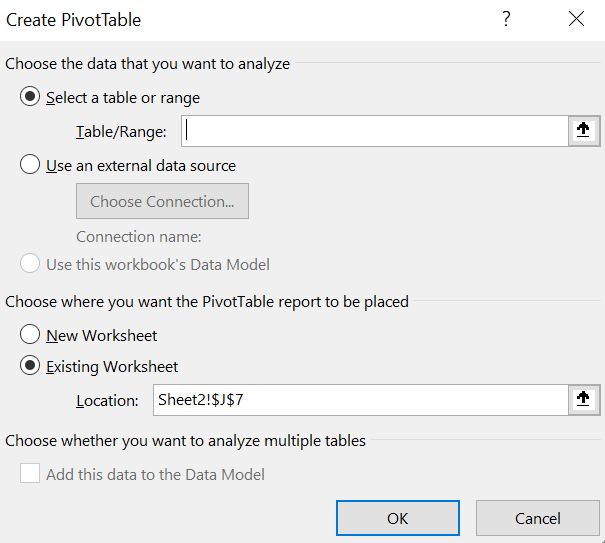


Figure 1 Range input taken to create pivot table

1. Now we need to select the data cells, or we can specify the cell numbers directly in the range input. We can also select to create pivot table in existing worksheet or new worksheet.
2. Then we get an option to drag the fields on to rows, columns, filters, and values. We need to drag and drop the fields according to our analysis needs.

**Applications of Pivot table**

1. We can summarize the data according to the categories.
2. We can list out the unique values of any column in the pivot table.
3. We can further visualize the pivot table data into pivot charts.
4. Performing operations such as filter, sort etc. without writing its actual formula.
5. We can connect the data to a pivot table from some external sources.

**Pivot Charts**

* Pivot charts are basically visual representation of pivot table. We are free to choose the required fields for visualization of table.
* It will make a user easily understand what data all is about and the patterns that are lying between the fields.
* Large data can be visualized easily using the pivot charts. There are different visualizations included with this functionality such as,
  + Line chart
  + Pie chart
  + Bar chart
  + Area chart
  + Treemap.. etc

**Steps to create Pivot Chart**

1. Initially click on the insert option and then click on pivot chart option available on the menu.
2. Select the range of the pivot table or we can directly specify the table dimensions in range input (similar to Figure 1). We also have an option to create the pivot chart in same worksheet or the different worksheet. After selecting the options click on ‘ok’.
3. Then it will show us the pivot chart fields i.e filters, Legend, Axis and Summation values.
4. Click on the ‘+’ sign available on the pivot chart right end corner to check the

* Axes
* Axis titles
* Chart title
* Data labels
* Data table
* Error bars
* Gridlines etc.

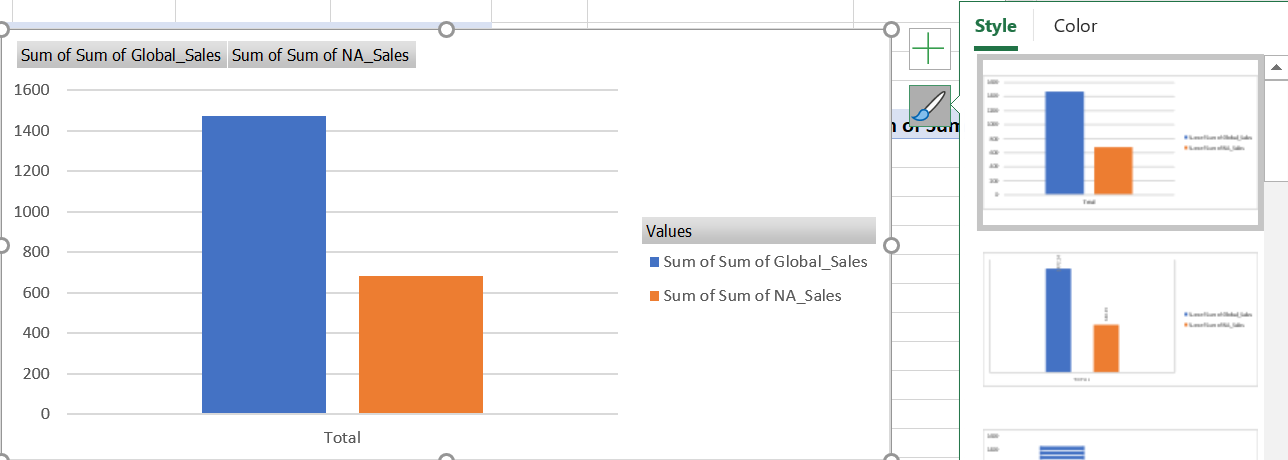
1. We can also change the style and color of the chart by clicking on the option below ‘+’ sign. 

Figure 2 Example showing the style and color options available for pivot chart

**Applications of Pivot chart**

* Pivot chart makes it easy to understand the pivot table data and allows us to know the patterns included between the fields.
* Various representations are available to visualize the data which is allowing for an easy data analysis.
* As it supports large volume of data, many organizations use it for summarizing the business reports.
* Categorized pivot tables can be visualized easily using pivot charts.

**Goal Results**

1. Highest global video game sales were done in which year?

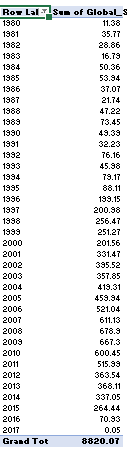


Figure 3 Goal 1 pivot table

**Rows**- Year

**Sum of values**- Global sales



Figure 4 Goal 1 pivot chart 1



Figure 5 Goal 1 pivot chart 2

* From the above two charts (bar and line) we can clearly observe that highest sales were done in the year 2008. Apart from that we can also observe there is a linear relation between global sales and the year.

1. In last three decades, Which region has the highest sales in ‘racing’ genre, and which has the lowest sales in ‘fighting’ genre? Suggest a genre we need to concentrate with respect to each region to increase the sales in next decade?

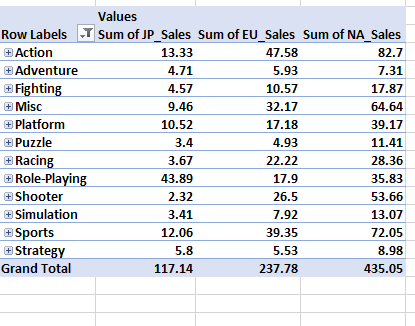


Figure 6 Goal 2 Pivot table

**Rows**-Genre, Year

**Columns**- Sum of values

**Sum of values**- Japan, Europe, and North America Sales

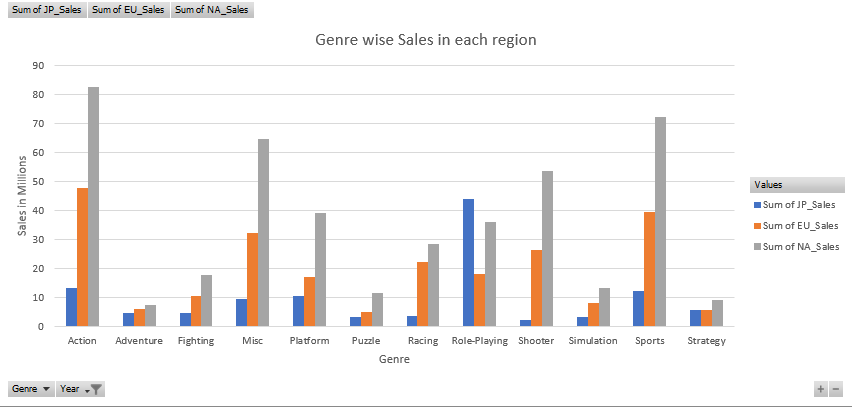


Figure 7 Goal 2 Pivot chart 1

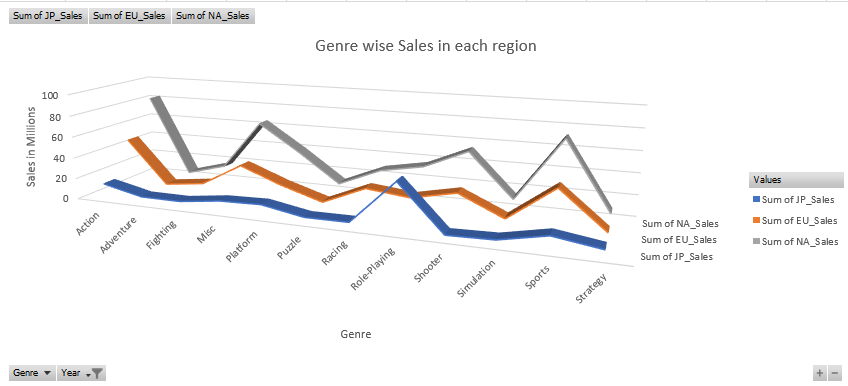


Figure 8 Goal 2 pivot chart 2

* From the above charts we can observe that, North America region has highest sales in racing genre and Japan has the lowest sales in fighting genre considering the period from 1980-2010.
* We can expect good sales if we can invest more in these genres,
  + Japan- Role-playing
  + Europe- Action
  + North America- Action, Sports

1. Which platform made good sales in 1996 and in 2014. If a publisher would like to invest more, which platform can we suggest in future? (Globally)

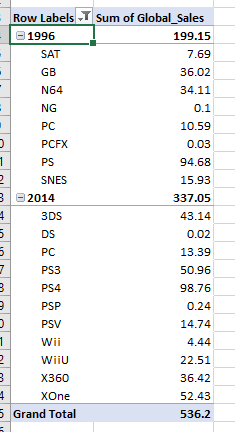


Figure 9 Goal 3 pivot table

**Rows**-Year, Platform

**Values**- Sum of global sales

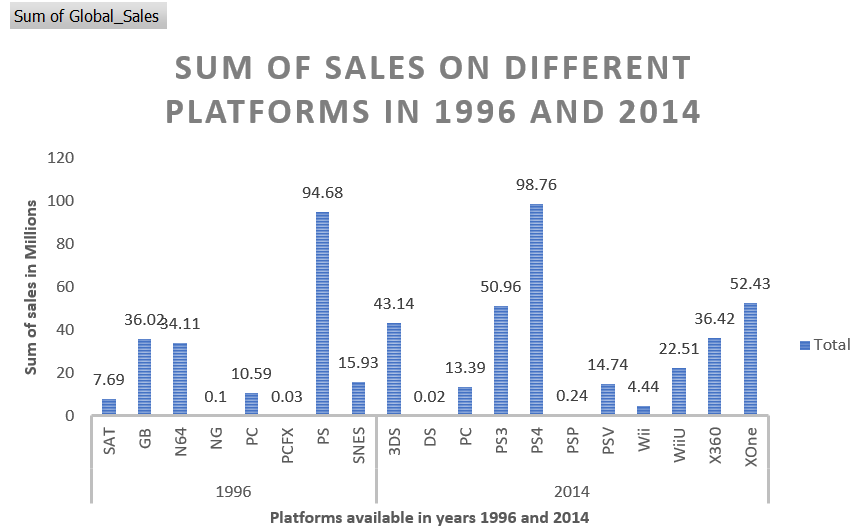


Figure 10 Goal 3 pivot chart 1

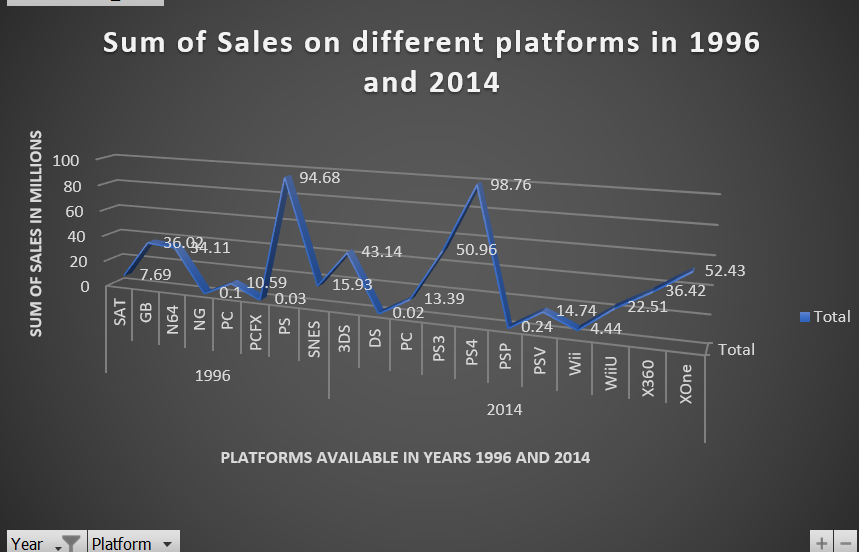


Figure 11 Goal 3 pivot chart 2

* Observing the above graphs, we can clearly say that in 1996 ‘PS’ platform got the highest sales and also in 2014 ‘PS4’ platform got highest sales. As PS platforms are doing well in the market irrespective of region, we can suggest to invest more on it.

1. Considering from 2000-2015, global video games sales has lowest average in which year?

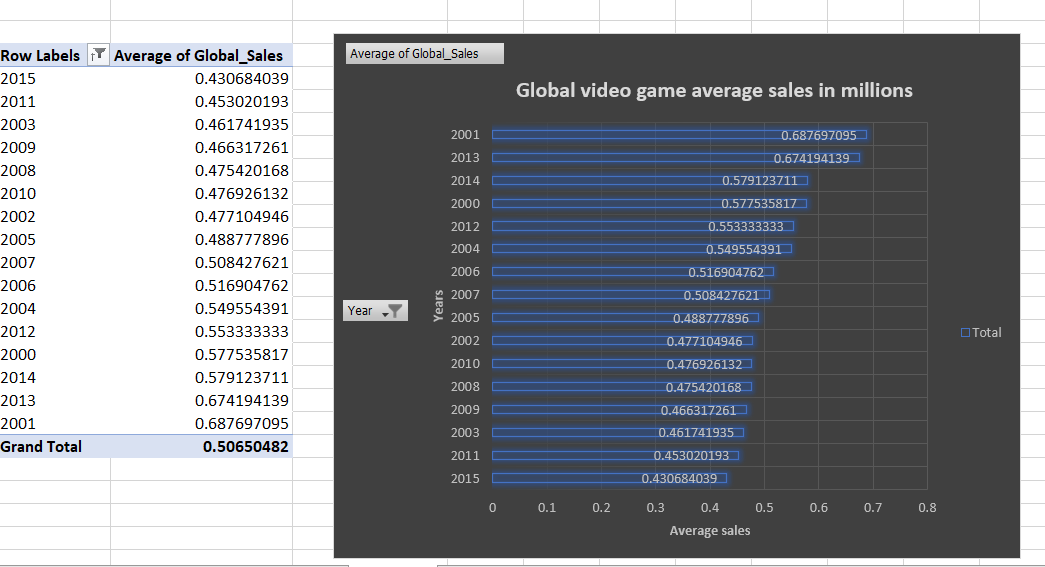


Figure 12 Goal 4 pivot table and pivot chart

**Rows**- Year

**values**- Average global sales

* From the above graph we can observe that lowest average is observed in year 2015.

1. Highest games were released by publisher named ‘Activision’ in which Genre? Did it release at least one game in all available Genres?

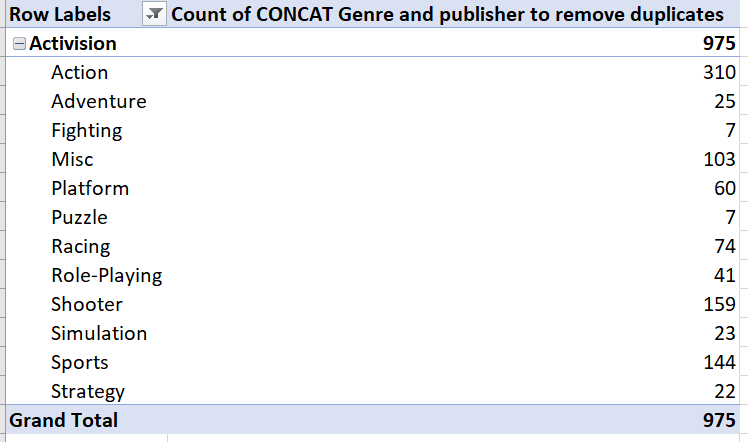


Figure 13 Goal 5 pivot table

**Rows**- Publisher, Genre

Here I have concatenated two columns Genre and publisher and then counted the values as I want to remove duplicates.

**Values**- Count of (Concatenated column)

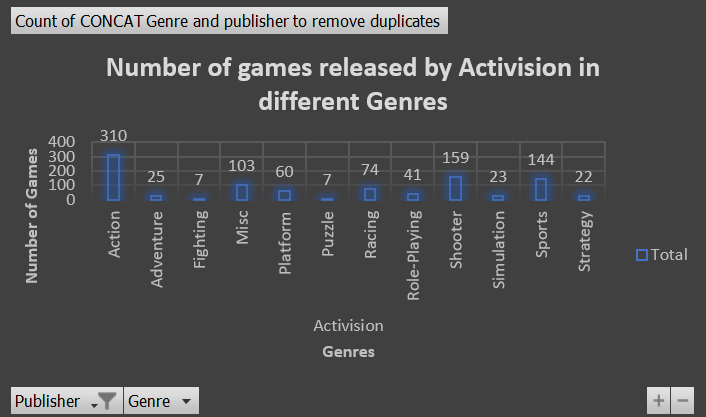


Figure 14 Goal 5 pivot chart 1

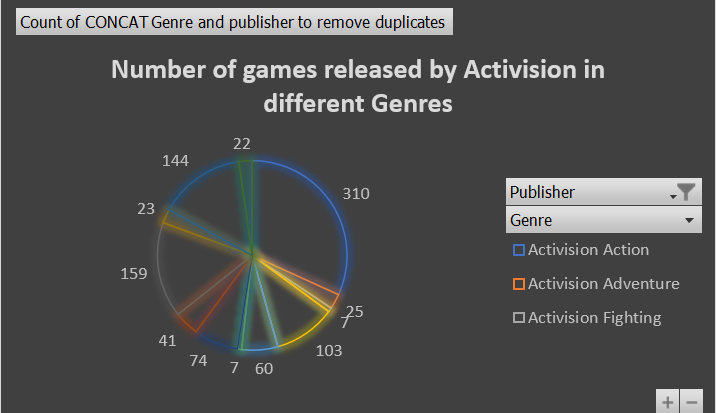


Figure 15 Goal 5 pivot chart 2

* Activision released highest number of games in genre ‘Action’ with count 301. Yes it released at least one game in all the available genres. The least games were released in genres fighting and puzzle.

**RapidMiner**

RapidMiner is a powerful software tool that supports complex data analysis tasks. I am planning to use RapidMiner tool to perform data analysis on my video game sales data. Firstly, I have gone through the download and installation steps and followed them one by one. Installation procedure was so easy, and I was able to install the application in my first attempt.

**Download and Installation steps**

1. Go to below link initially, and then click on the register tab. Fill in all the required information and click on submit.

<https://my.rapidminer.com/nexus/account/index.html#login>

1. An email confirmation will be sent to the registered email address. Confirm it.
2. Once the email confirmation is done, website redirects you to home page. Click on download and select the OS platform.
3. Once download is done, install RapidMiner application. Open the application and login with your credentials.

**Tutorial Learnings**

* RapidMiner provides user friendly interface with different views (design, results, and auto model), repositories, operators, parameters. We have different operators under Data access, blending, cleansing, modeling, scoring, validation, utility, and extensions. We can just drag the required operator onto process panel.
* We can install various extensions through marketplace by just clicking on the extensions tab on the top menu. The installed extensions will be updated in operators’ extensions on the left bottom panel.
* We can also see the statistics associated with each column of the data and various filter operations can be performed.
* Visualizations can be performed on the data using visualization workbench which contains various types of charts.
* We can import excel data in RapidMiner, merge two or more datasets and then perform analysis over it.
* One thing that interested me on this tool is we can make some external connections to load the data. For example, we can make connection with a file or with a database or with cloud API etc.
* The advantage of this toolkit is it is open source. We can download it perform analysis for free. With the increasing demand for analysis in present day, I think the toolkit serves the users in best way to get the future predictions.