ISM 6419

Data Visualization

Video Game Sales Analysis

Kshitija Nandkishor Dound

U97616470

Introduction:

The video games industry has been prevalent for a long time now. My interest in this field makes me inquisitive about how the popularity and sales of games get affected by various factors. The factors that influence the video games sales are country or the market, the genre of the game, platforms, and the publisher of the game which eventually decide the success of the games.

Video Games have proven to be an ever-increasingly popular source of entertainment not just for teens but for a wide range of audiences. A lot of gamers not only play games but also stream while playing which helps them make money from it. With the advent of Covid-19, a lot of people invested in online games as per the online reports thus, increasing the demand for games. In 2021, the net global revenue of video games accounted for $138.4 billion which was a little over $7 billion than 2020's market value.

To analyze the data trend of video game sales and understand more about this multibillion-dollar industry, I decided to create this data visualization project which addresses the following questions:

1. Who made the most sales? Is it true that the game market is booming?
2. How do gaming companies target their users and distribute or market their products? To answer this question, I'll be researching factors like user and critic score, genre, gender, age group, and countries where they were released.
3. Which streaming platforms are more preferred for streaming?

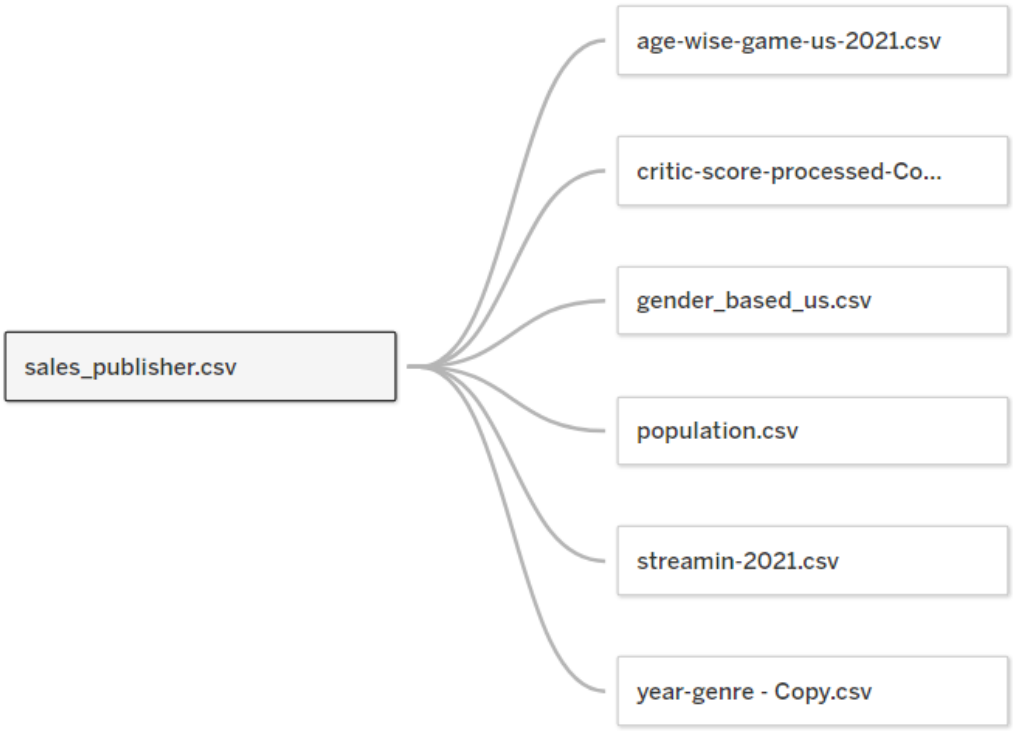
To make this project ambitious, I collected data from different sources to include different factors to help make sense of the data. For example, collecting the country-population data, normalizing it, and creating calculations to find per capita sales, and collecting gender-wise country population and finding per capita sales for males and females in the respective country.

Methodology:

Step 1: Finding datasets
I found Video Game Sales data collected from Vgchartz and ratings from Metacritic on Kaggle in the form of CSV files. I found more data from Statista, United Nations World Population Prospects. The descriptions of the data sources is as follows:
Kaggle: Kaggle is a reputed data source. Here I found web scraped data from VGChartz and Metacritic which are the sites which keeps the video game sales data and the video game rating data.
United Nations World Population Prospects: This serves as a reputed source for population estimates and projections that have been prepared by the Population Division of the Department of Economic and Social Affairs of the United Nations Secretariat.
Statista: Statista is a leading provider of market and consumer data for various industries.

Step 2: Cleaning and Pre-processing the data in Excel, Tableau-Prep Builder and OpenRefine
I used Excel to add columns that were needed and get rid of unnecessary rows which were meta-data. I used Tableau-Prep Builder to get rid of the null values which got rid of almost 9000 records. I filtered the year released to 2009-2016 which further reduced more records. The scoring system range was different for the critic score and the user score, so, I multiplied the user

scores by 10 as the critic score's high bound was 100 and the user score's high bound was 10 to make the scoring system consistent. I used OpenRefine to transpose cells across columns and perform a split operation on them. I used excel to clean up the rest of the data which included meta-data for different columns in other CSV files and kept the data which was required for the visualizations.
After cleaning up, I imported the CSV files to Table and performed join operations on the tables as shown below:



Data Tables :

1. **Sales_publisher:** This table includes following columns:
    - Name – Name of the game
    - Publisher – Publisher of the game
    - Market – Countries (North America [NA], Europe [EU], Japan [JP])
    - Sales – Video Game sales in million dollars
2. **Year_Genre:** This table includes following columns:
    - Name – Name of the game
    - Platform – Platforms where the games were released (3DS, DS, GBA, GC, PC, PS2, PS3, PS4, PSP, PSV, WiiU, X350, XOne)
    - Year of Release – Year of the game release
    - Genre – Genre of the game

3. **Critic_Score:** This table includes following columns:
    - Name – Name of the game
    - Source – Type of Score (User or Critic) collected by Metacritic
    - Score – Actual scores
4. **Populationnormalized:** This table includes the following columns:
    - Market – Countries (North America [NA], Europe [EU], Japan [JP])
    - Population – Normalized Population count in million
    - Year – Year of the population estimate
    - Male – Normalized Male population
    - Female – Normalized Female population
5. **Streaming-2021:** This table includes the following columns:
    - Market – Country for which this data was collected
    - Year – Year this data was collected
    - Streaming platform – Name of the streaming platform
    - Share of response – (%)Percentage of share
6. **Age-wise-game-us-2021:** This table includes the following columns:
    - Market – Country for which this data was collected [US]
    - Year – Year in which the data was collected
    - Age Group – Age groups
    - Percent – Percentage(%) of people in that age group
7. **Gender_based_us:** This table includes the following columns:
    - Market – Country for which this data was collected[US]
    - Year1 - Year in which the data was collected
    - Male(%) – Percentage of male gamers in the US
    - Female(%) – Percentage of female gamers in the US

Analysis:

1. **Simple Visualization with filter:**
   The visualization Figure 1 shows the trends of video game sales over the years for different markets/countries. The visualization can also be filtered by country which further helps to check country specific trends. It seems like the sales are descending over time for all markets. But if we see the 2022 statistics on the internet, the video game industry has increased 14.4% from 2020. In 2016, it was predicted that the industry will be worth $90.07 billion in 2022, but as we can see there is a huge difference of 76.8% between the two figures. But the data here tells a different story. The possible explanation for this would be the market in the dataset doesn't include the markets which is adding to the global sales. In addition to that, this dataset doesn't contain mobile as the platform.
   According to this visualization, highest number of sales is generated Globally followed by North America, Europe, Other and lastly Japan in 2009, and we can see it's been decreasing since. This partially answers the first question.

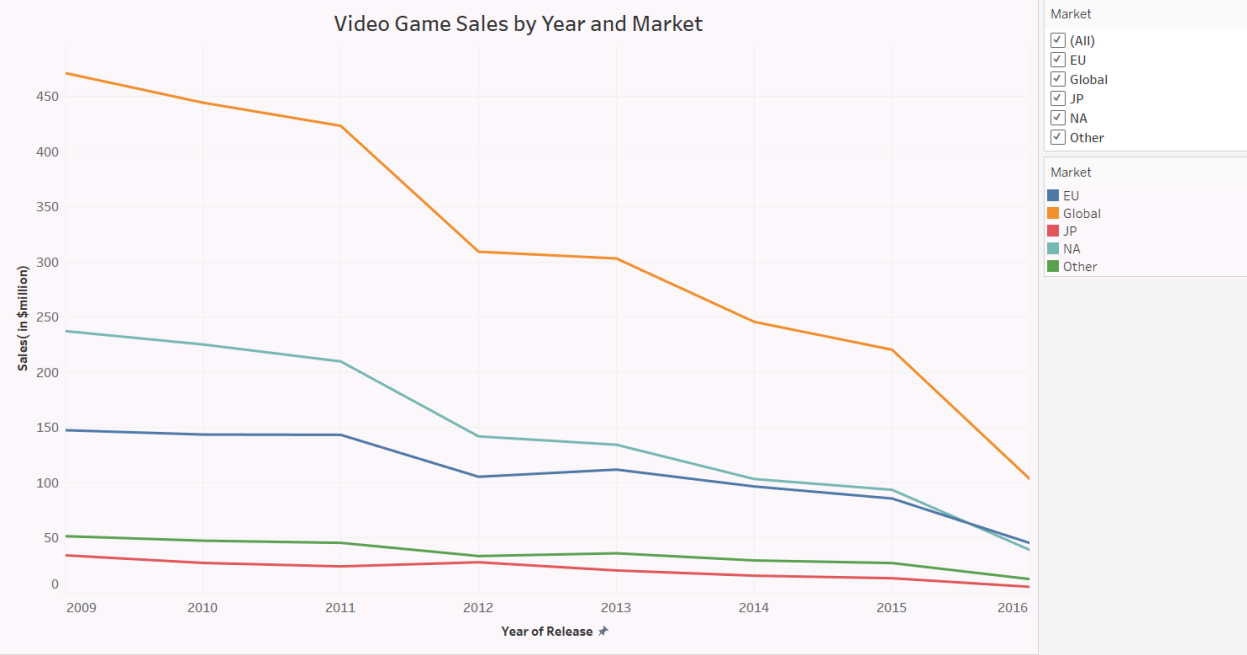


Figure 1. Video Game Sales by Year and Market

2. Simple Visualization:

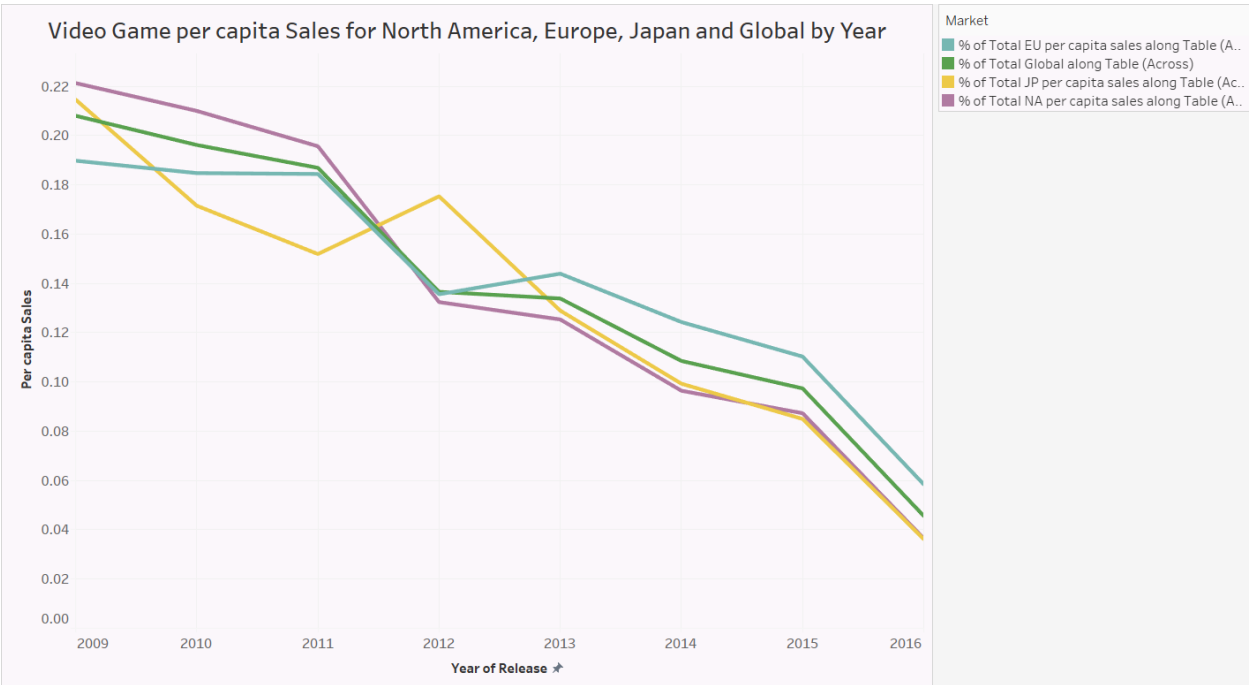


Figure 2. Video Game per capita Sales by Year and Market

Figure 2. shows the per capita sales generated by video games in different markets. To make more sense of the data, I normalized the population as the high global population would have significantly affected the visualization. As we can see here, North America was leading in per capita sales in 2009 and Europe was leading in 2016. As you can see there is an increase in per capita sales for Japan in 2012 and decrease

in per capita sales for Europe. Again, the explanation for the decline could be the lack of different markets and different platforms which are not included in the dataset which explains the inaccuracy in the visualizations.

3. Simple Visualization:

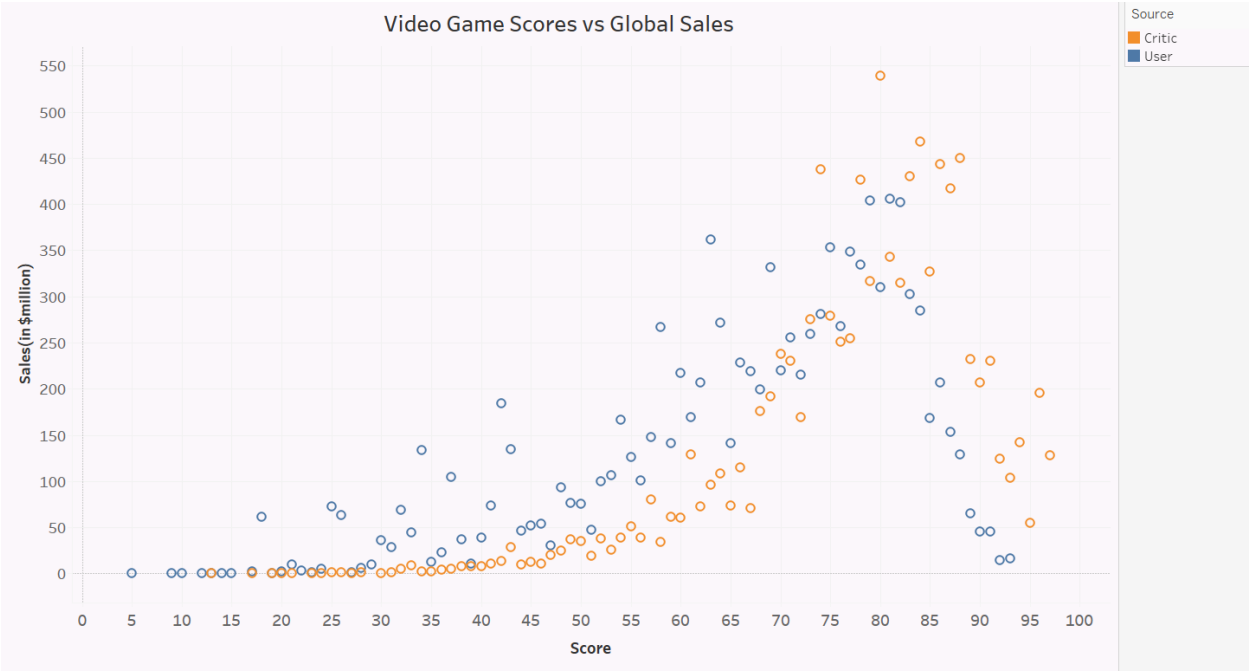


Figure 3. Video Game Scores vs Global Sales

In the visualization above, we can see the correlation between critic scores, user scores and the video game sales which seems to be non-linear. It basically means that there is no noticeable trend with the increase in score and sales. The discrepancies in the data like lack of data point for a specific rating can affect the visualization and thus, the correlation between scores and the sales. This partially answers the second question.

4. Simple Visualization:

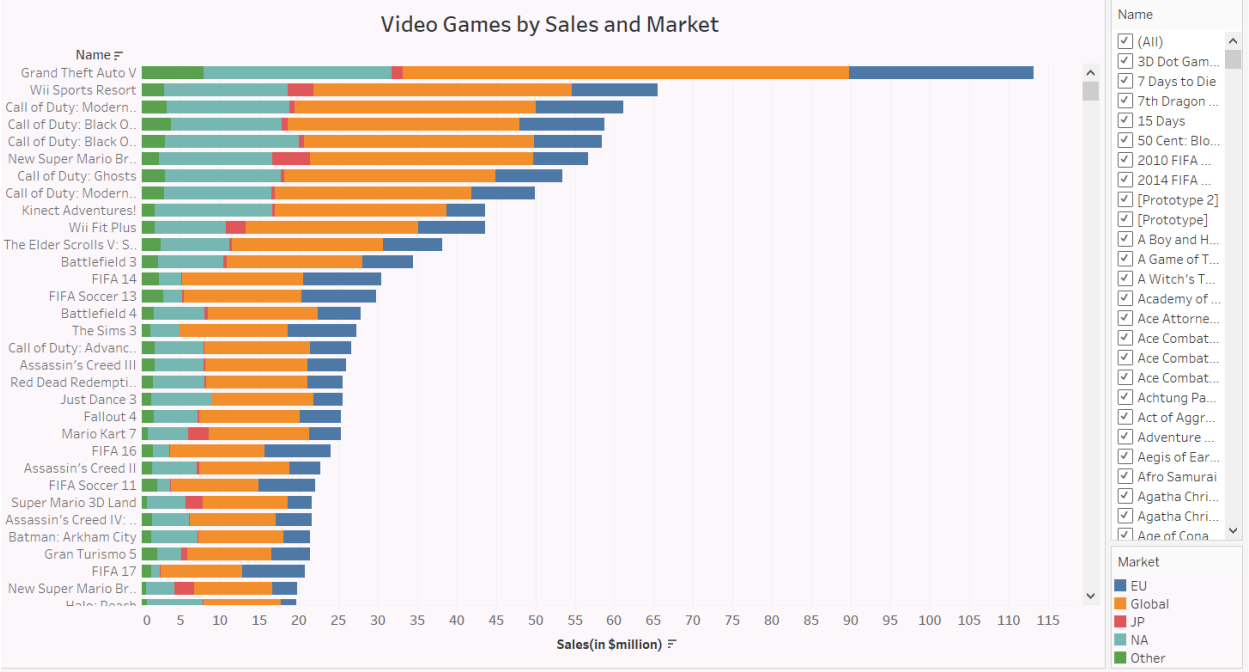


Figure 4. Video Game by Sales and Market

We can see from the Figure 4 that the winner is Grand theft Auto V followed by Wii Sports Resort and Call of Duty: Modern Warfare 3. This answers the first research question.

5. Simple Visualization:

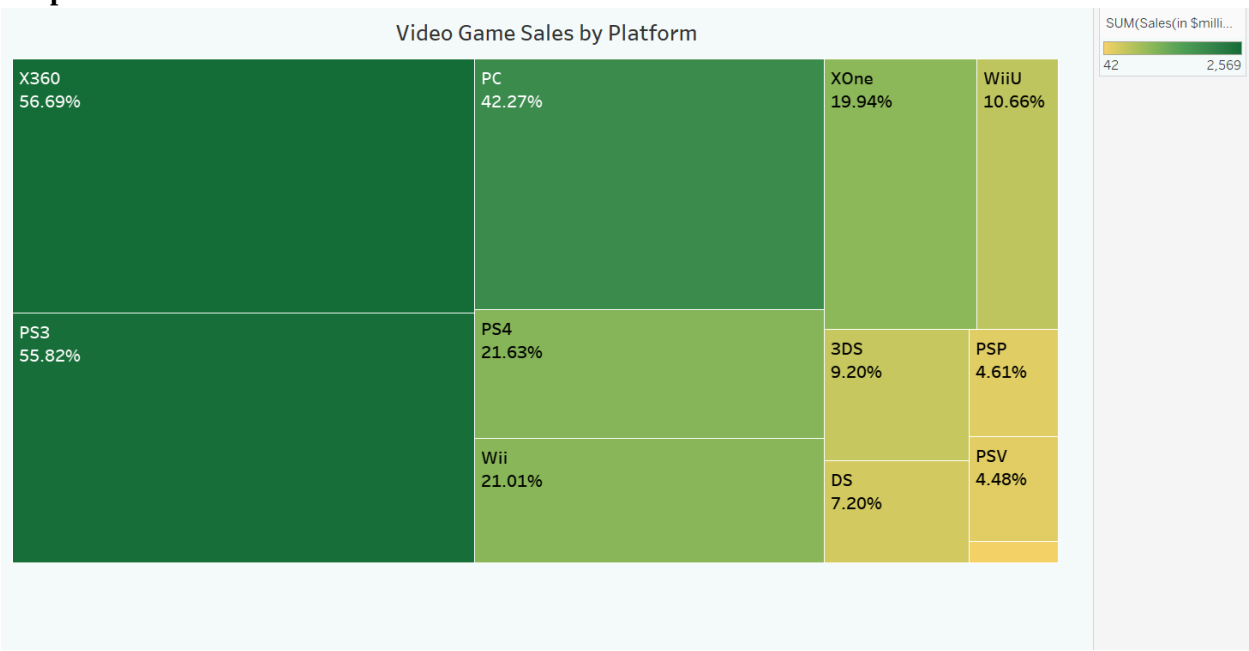


Figure 5. Video Game Sales by Platform

The visualization above shows that the platform X360 made the most sales followed by PS3 and PC. The colors move to the lighter shade as the sales decreases. This answers the first question.

6. Simple Visualization:

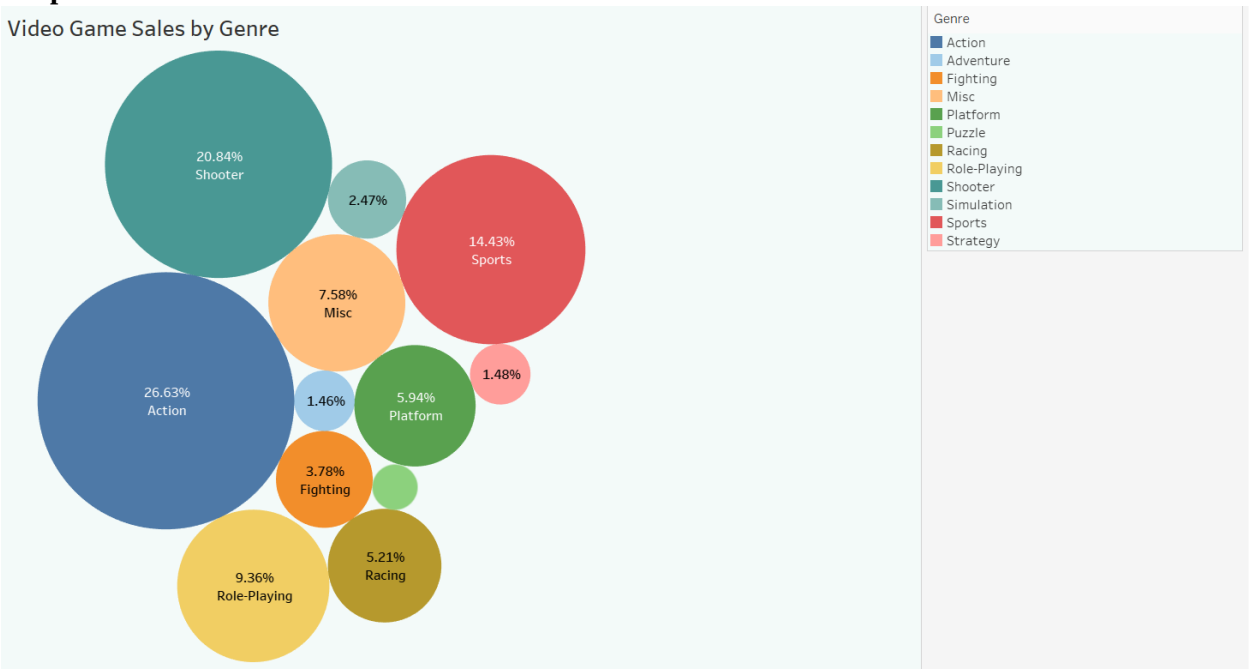


Figure 6. Video Game Sales by Genre

To answer the second question, I have made the visualization above to see which genre accounts for more sales. The size of circle represents the amount of global sales and genre is represented by colors. The top 3 genres are Action, Shooter and Sports. The gaming companies can target the gamers who buy the top genres, and release more games in the genre which is played the most. This partially answers the second question.

7. Advanced Visualization:

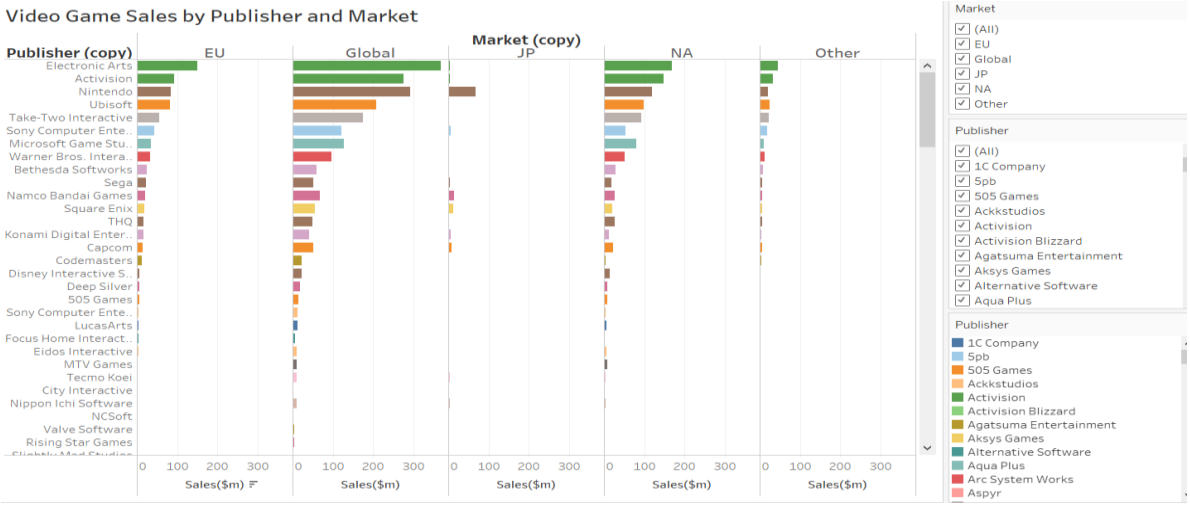


Figure 7. Video Game per Sales by Publisher and Market

The figure above shows Market wise publishers that dominated the sales. Electronic Arts dominates most of the markets except for Japan, followed by Activision, and Nintendo is the top publisher in Japan. The next up in Japan is Sega which is somewhat like Nintendo. Thus, these publishers can target the specific markets to generate more sales. EA and Activision would face loses if they continued to sell their games in the Japan market if they won't use better marketing strategies. This partially answers the second question.

8. Simple Visualization:

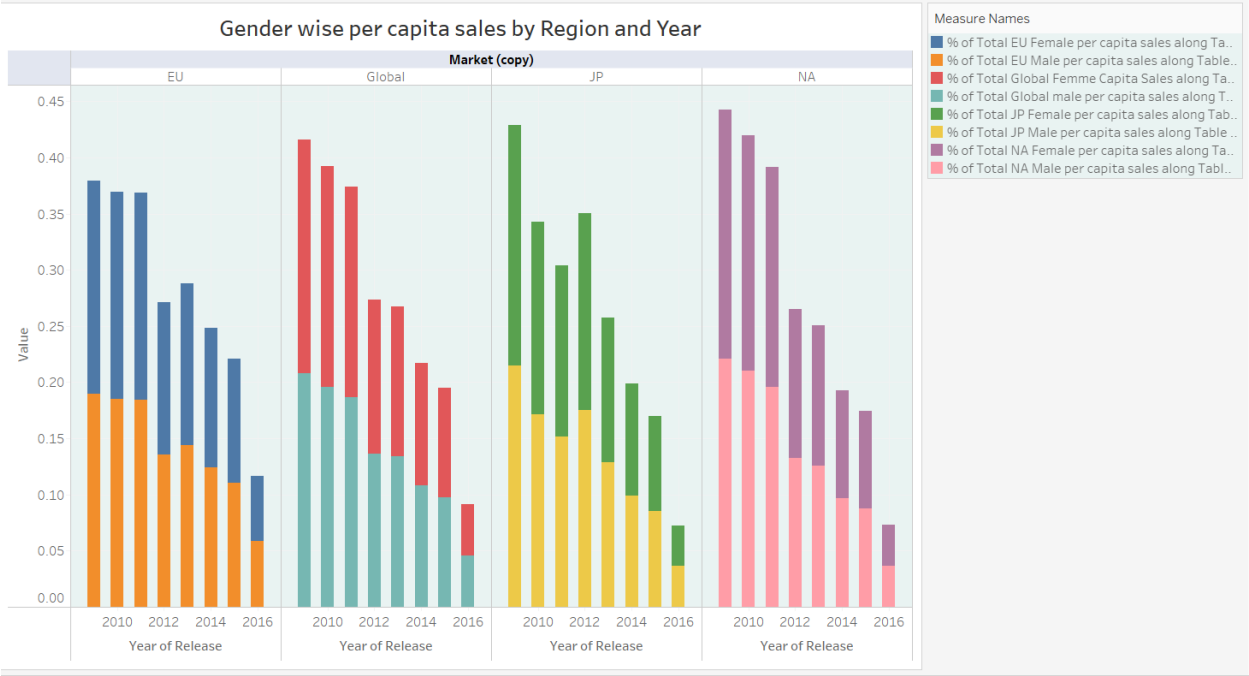


Figure 8. Gender wise per capita sales by region and year

The graph above shows the gender wise per capita sale for the different markets by year. The highest per capita sales is in NA followed by Japan, Global and Europe. The gaming companies can thus need to market their product more in the countries where the sales are comparatively less. They can use strategies like cheaper rates based on the per capita income to sell their products. This partially answers the second question.

9. **Simple Visualization:**

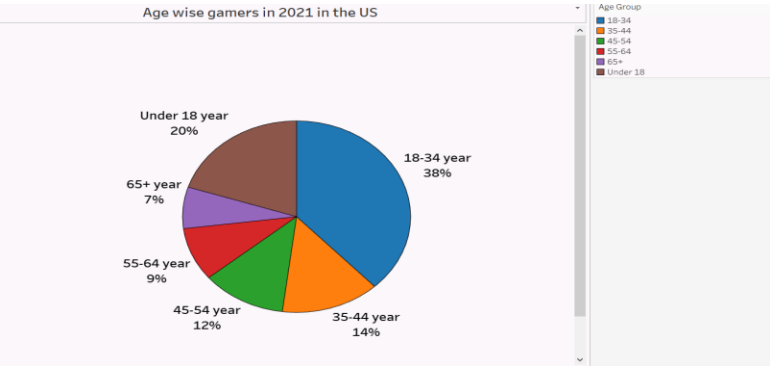


Figure 9. Age wise gamers in 2021 in the US

The pie chart above shows the share of gamers in the year 2021 in the US market according to the age of the gamers. It's the gamers with the age range from 18-34 account for 38% of the share followed by gamers under 18 and age range of 35-44 years. The companies can use strategies suitable for the age group to market their products to the specific age group. Also, it is quite noticeable to see that there is just 4% difference in the gamers who are under 18 and 35-44.

10. **Simple Visualization:**

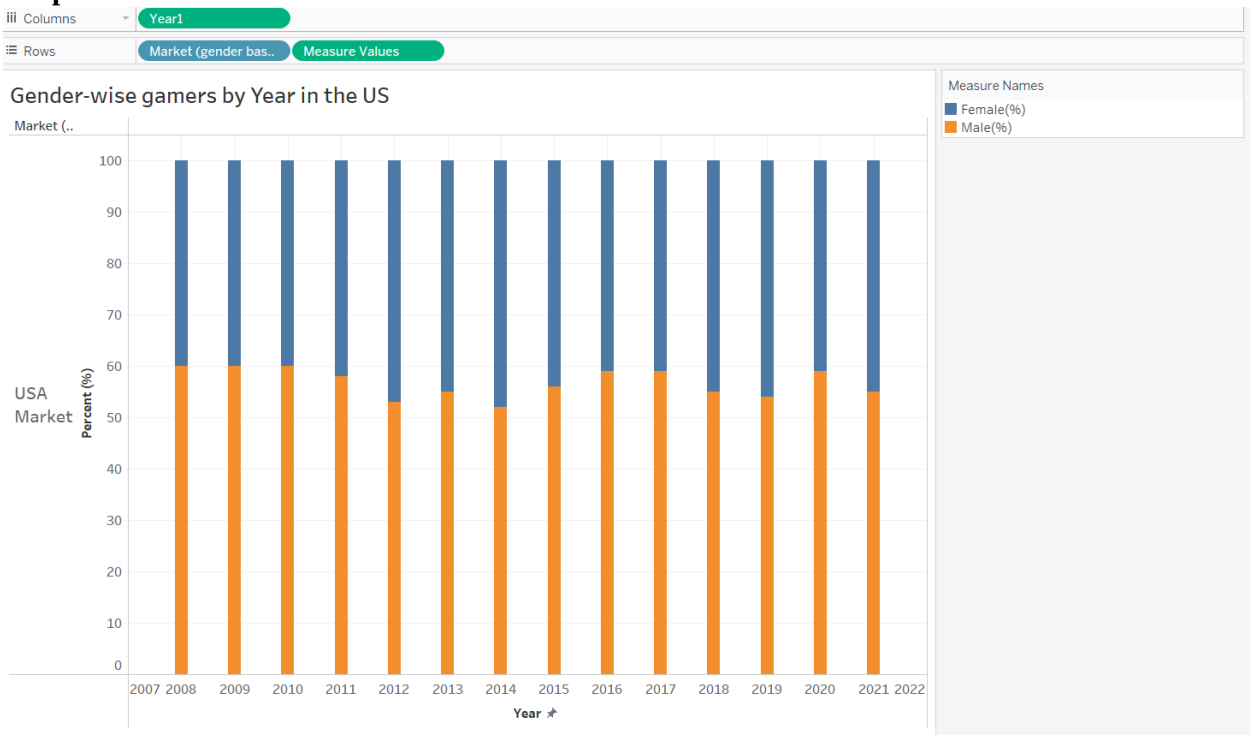


Figure 10. Gender wise gamers by year in the US

From the graph above, it can be seen that the number of female gamers has increased over the years in the US market. Being a girl who is fond of video games, this makes me quite happy. There is not much

difference in male and female gamer count for the year 2021. Thus, gaming companies can market the games to the specific gender if they see a noticeable trend. I'd like to research more on this in the future.

11. Simple Visualization:

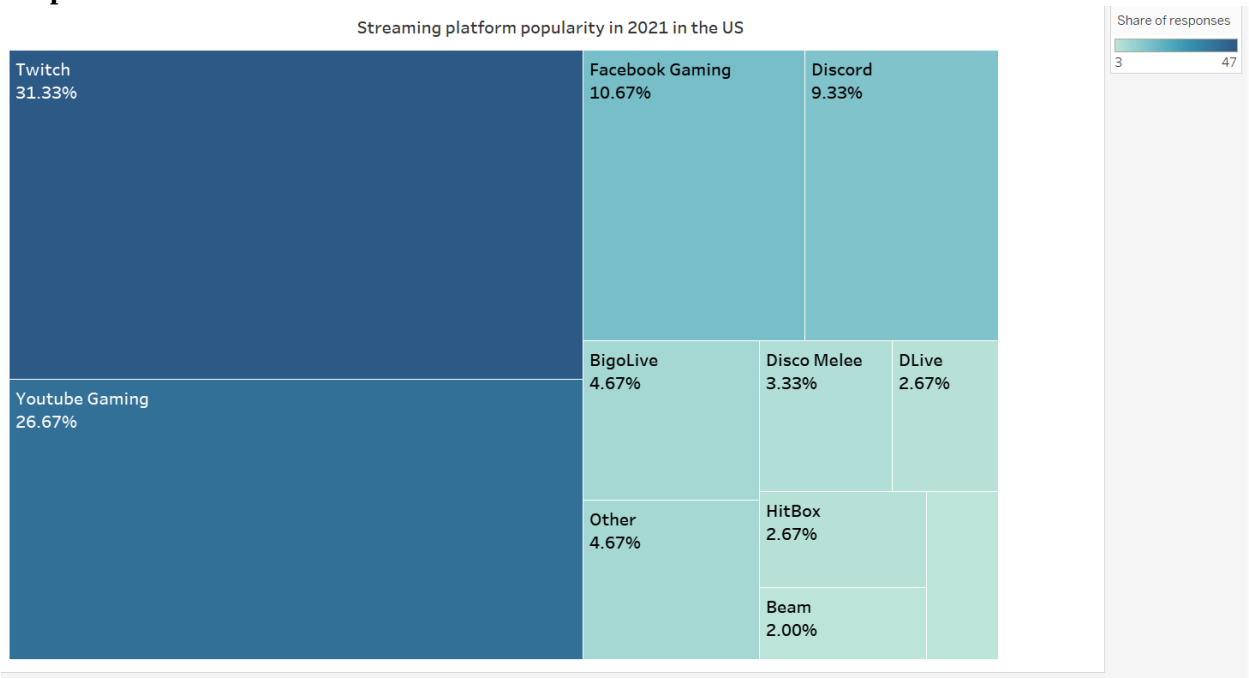


Figure 11. Streaming platform popularity in 2021 in the US

From the graph above, it can be seen that the Twitch was the most popular streaming platform followed by YouTube Gaming and Facebook Gaming. A lot of streamers are switching to YouTube as YouTube has always been famous and preferred tool. This answers the third question.

12. Advanced Visualization (Animation):

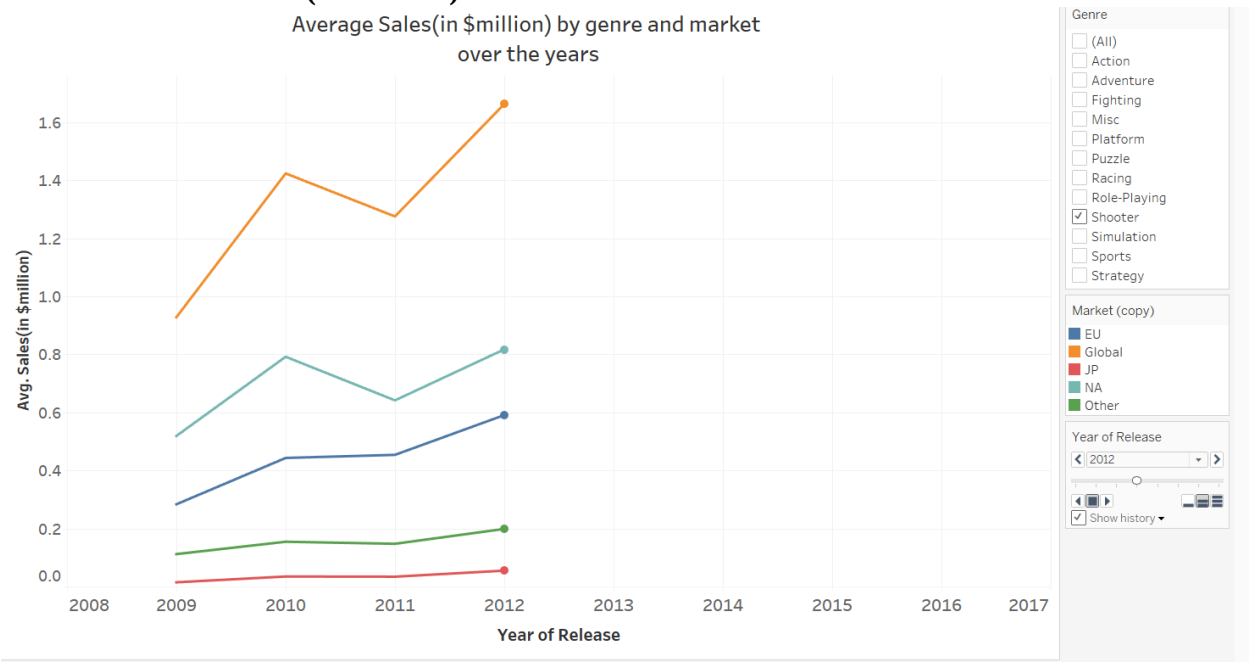
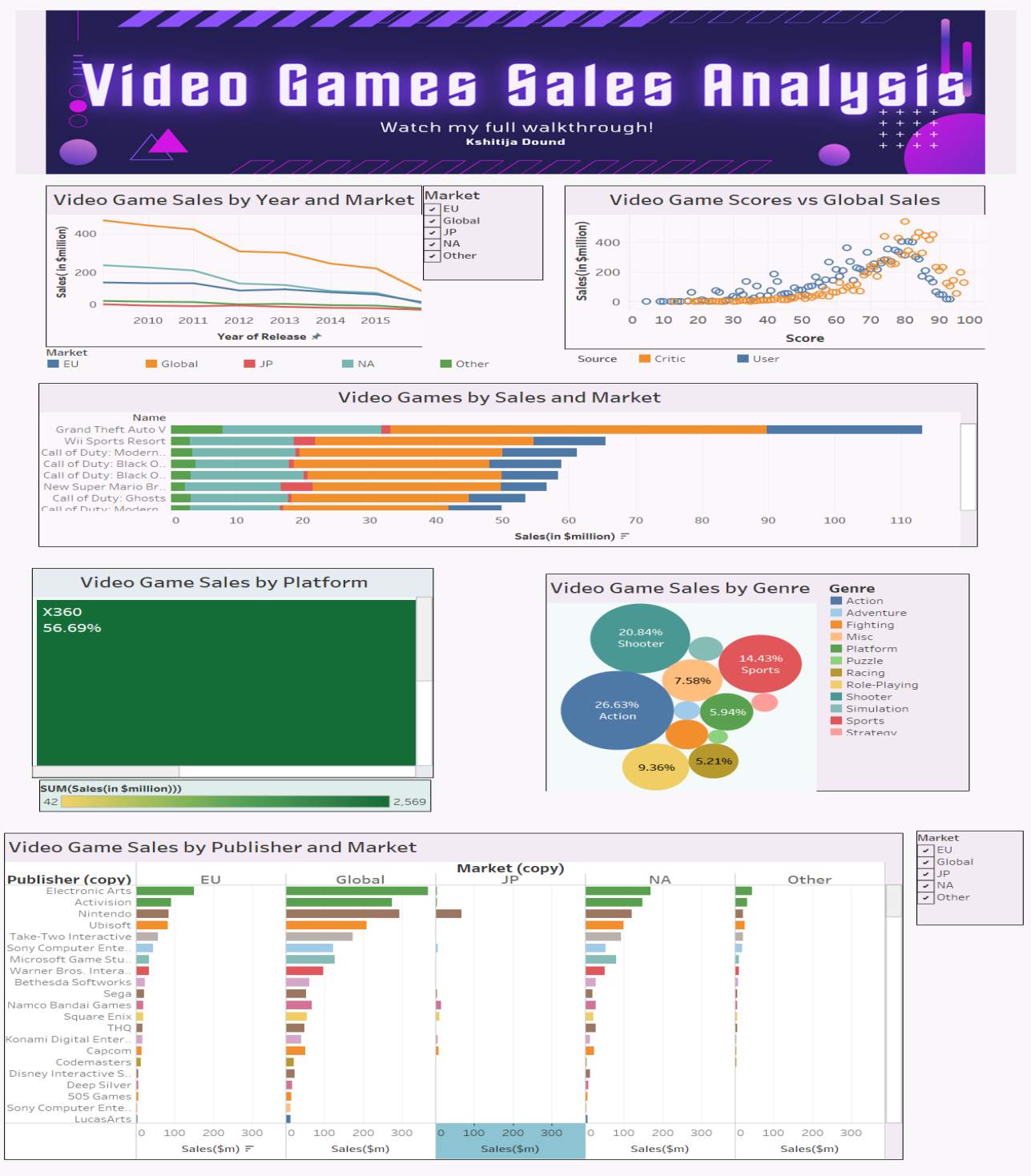


Figure 12. Average Sales(in $million) by genre and market over the years

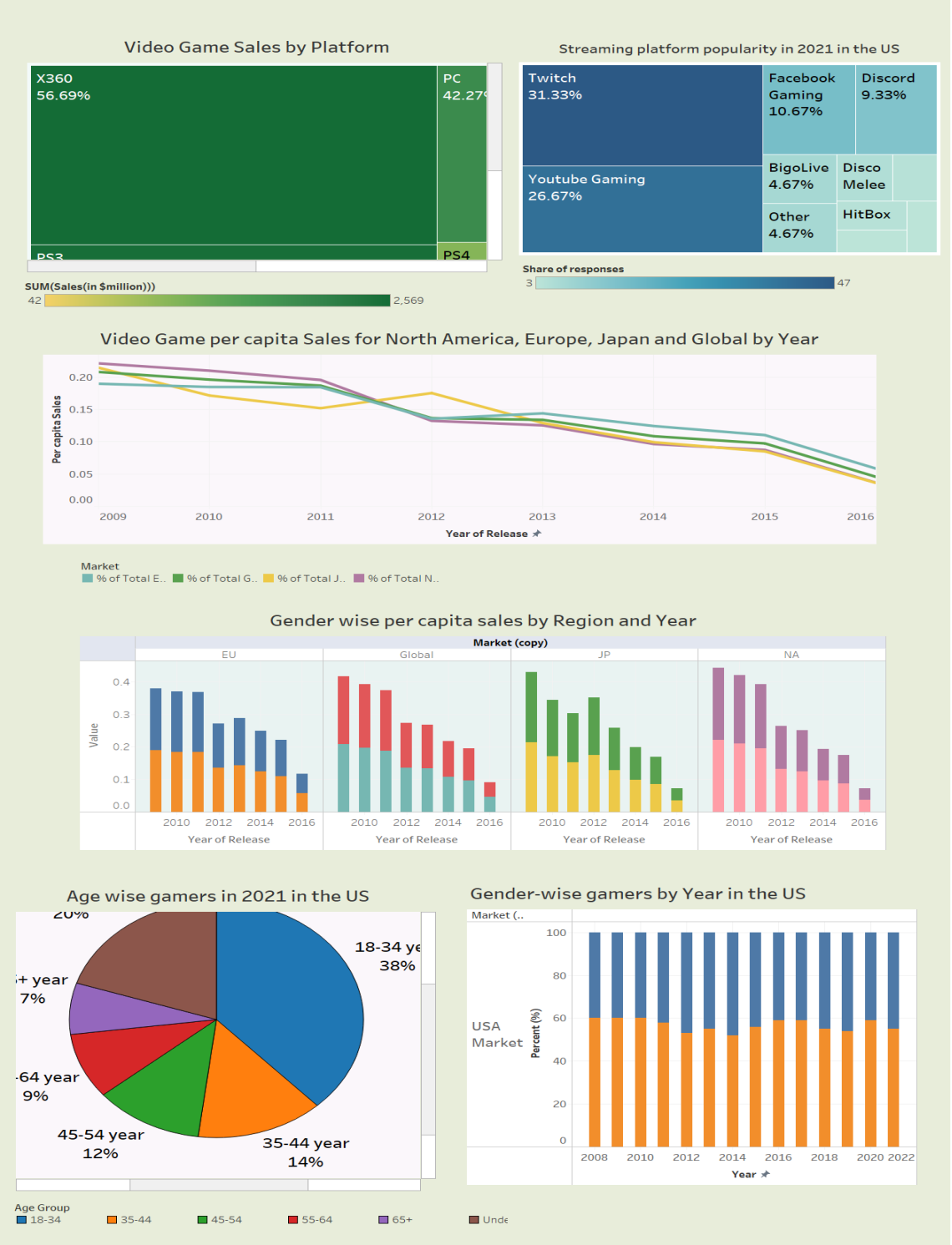
Figure 12. shows the animated visualization for the video game sales by country over the years. There is a filter made available to filter the visualization by the genre of the game. As animations are interactive, it makes it easier to understand the trend.

Dashboard 1:



I have created this dashboard to give viewers a quick view of what's going on in the data. This dashboard covers different parameters to compare the video game sales in different markets, platforms and genre over the years.

Dashboard 2:



This dashboard essentially gives a gist of research question 2 and 3. It gives a quick walkthrough of how the video game sales are distributed in different market based on the gender, and also shows the trends only for the United States by age and gender.

Conclusion:

Thus with the help of different visualizations, animation and dashboards, I have tried to answer the research question I chose for this analysis. This project gave me an opportunity to view the trends in video game sales. I have answered the research questions in the analysis above with the help of visualizations I created. To summarize :

1. Who made the most sales? Is it true that the game market is booming? – Europe was leading in sales in 2016. The game market appears to be shrinking but the possible explanation could be the lack of different markets like China in the dataset, and the absence of mobile platform games.
2. How do gaming companies target their users and distribute or market their products? To answer this question, I'll be researching factors like user and critic score, genre, gender, age group, and countries where they were released. – Different companies can use different marketing strategies based on how well they do in different markets based on the gender, age, genre and the scores. I have explained this in detail in the analysis above.
3. Which streaming platforms are more preferred for streaming? Twitch has always been more popular, but Youtube can soon take over Twitch.

There are some questions which I would like to address in the future as the data I found didn't cover all the questions I had. I would like to see the latest trend as soon as I am able to find the latest dataset. I would like to see the impact of covid on the video game market. I would also like to see how different the visualizations would look with the inclusion of Asia-Pacific market, and Mobile platform games.

Sources, Reference and Acknowledgements:

- Marston, Hannah. (2012). Older Adults as 21st Century Game Designers.. The Computer Games Journal. 1. 90-102. 10.1007/BF03392330.
- Smith, Gregory. "Video Game Sales 1980-2017." http://www.bmtfx.com/video-games-sales-1980-2017-2/?lang=en
- Video Game Industry Statistics- https://www.wepc.com/news/video-game-statistics/
- Kaggle Dataset Source: Kaggle
- Statista Data Source: Statista
- Population Data Source: United Nations World Population Prospects
- Visualizing Video Game Data 2007-2016 with Tableau: https://studentwork.prattsi.org/infovis/visualization/visualizing-video-game-data-2007-2016-with-tableau/
- Gaming Market – Growth, Trends, COVID-19 Impact, and Forecasts(2022-2027) by Mordor Intelligence- https://www.mordorintelligence.com/industry-reports/global-gaming-market
- The rise of the Virual Empire: Video Game Indusry Statistics by fortunly- https://fortunly.com/statistics/video-game-industry-statistics/#gref