CA2 Growth of Digital Games

# Introduction

The gaming industry has experienced a massive growth over the past few decades, with the rise of Steam playing a pivatol role in shaping how games are published and distributed. Steam hosts a large library of games spanning over 90,000 with many genres, developers and price points. With this vast choice for players and developers understanding the trends and patterns in the data can be beneficial for both parties. This report leverages the Steam Games Dataset 2025 which includes all key information such as release dates, genre tags, developer and publisher information, and user reviews. The dataset is used to explore the growth of digital games on Steam and to identify the key trends and patterns in the data.In this report I will be exploring these trends using visualisation techniques in R, examples of things I will explore are how pricing can effect user reviews and how certain genres have grown and become more dominant over time. The findings will show a broader understanding of how game developers and publishers strategise their releases and how the pricing model may impact player perception and how the genre lines up with current trends. This report will provide meaningful insights for game developers and publishers to make informed decisions on their game releases and marketing strategies.

# Research Questions / Rational

## 1. How does the frequency / number of DLC’s for a game effect user reviews?

Some developers focus on keeping their games running for years by adding gameplay mechanics with regular updates and DLC I want to investigate if this has a positive or negative effect on user reviews. I aim to see if the number of DLC’s released for a game has an impact on the user reviews and how this has changed over time. I think this is interesting as some players may feel that the game is being milked for money and that the DLC’s should have been included in the base game. By analysing this data I can see how the number of DLC’s has impacted the user reviews.

## 2. How does the pricing of games affect user reviews?

The price of games has always been a crucial factor in players decision making when buying a game but in this research question I want to figure out how the pricing of games affects user reviews. I aim to see if there is a correlation between the price of games and the user reviews and how this has changed over time. I think this is interesting as higher priced games tend to be reviewed more harshly by players due to the higher expectations they have for the game. Players could also be more forgiving of lower priced games as they are more likely to take a chance on a cheaper game. By analysing this data I can see how the pricing of games has impacted the user reviews.

## 3. What genres have had trending growth over the years?

The third research question aims to understand what genres have had trending growth over the years. I aim to see if certain genres have become more popular over time and how this has impacted the sales of games in those genres. This is important as it can help developers and publishers understand what genres are popular and how they can capitalise on this trend. I want to analyse the number of releases per genre and see how this has changed over time by picking a few popular genres and seeing how they have grown over the years compared to other genres.

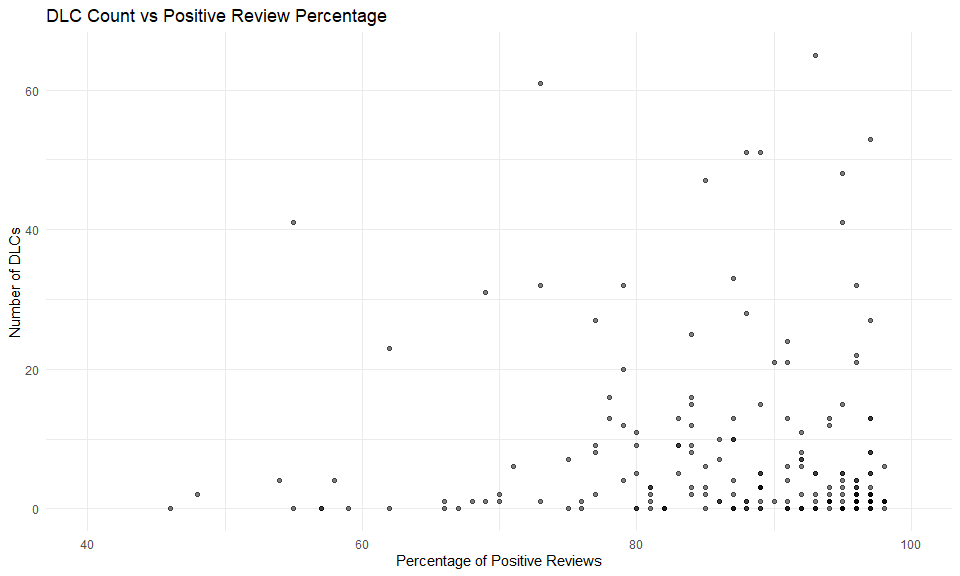
## 4. How has the AAA landscape changed over the years?

The fourth research question aims to understand how the AAA landscape has changed over the years. I aim to see how the number of AAA games has increased over time and how this has impacted the sales of games in this category. I also want to examine the price increase of these games and the user reviews since these price changes have come into effect.I want to understand the shift in the AAA landscape and see how the evolution of these studios and game quality with regular price increases has changed the publics opinion of AAA games and studios.

# Load the dataset  
steam\_games <- read.csv("games\_march2025\_cleaned.csv")  
steam\_games <- steam\_games %>%  
 mutate(  
 release\_year = year(ymd(release\_date)),  
 genres = str\_replace\_all(genres, '\\[|\\]|\"', ""), # remove brackets and quotes  
 genres = str\_trim(genres) # remove leading/trailing spaces  
 )

# Filter the data for pct\_pos\_total from 40 to 100  
steam\_games\_filtered <- steam\_games %>%  
 filter(pct\_pos\_total >= 40 & pct\_pos\_total <= 100, dlc\_count <= 75) %>%  
 arrange(desc(num\_reviews\_total)) %>%  
 head(200)  
  
# Find the highest number of DLCs  
max\_dlc <- max(steam\_games\_filtered$dlc\_count, na.rm = TRUE)

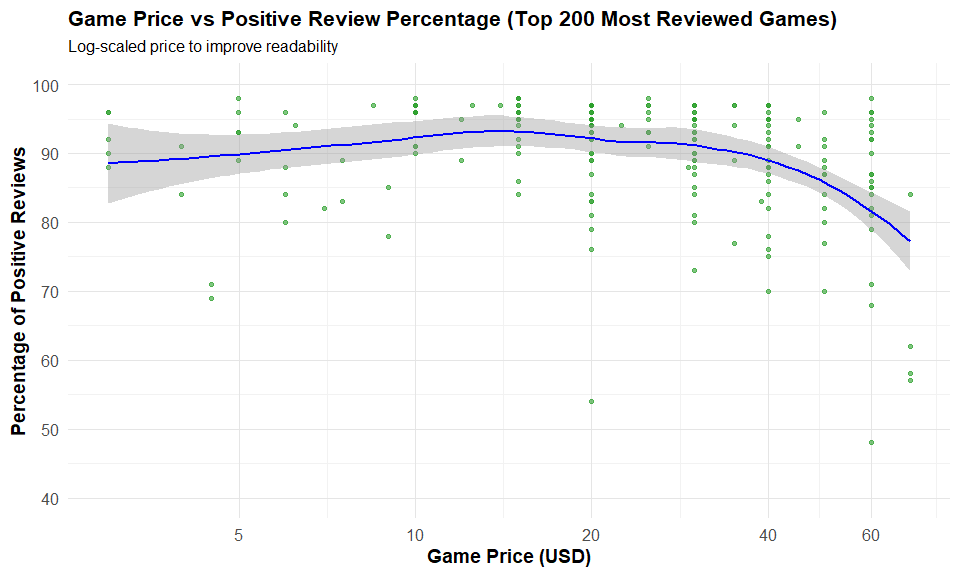
# VisualiSe the relationship between DLC count and average positive reviews  
ggplot(steam\_games\_filtered, aes(x = pct\_pos\_total, y = dlc\_count)) +  
 geom\_point(alpha = 0.5) +  
 scale\_x\_continuous(limits = c(40, 100)) + # Limit the x-axis from 40 to 100  
 scale\_y\_continuous(limits = c(0, max\_dlc)) + # Limit the y-axis from 0 to max DLC count  
 labs(title = "DLC Count vs Positive Review Percentage",  
 x = "Percentage of Positive Reviews",  
 y = "Number of DLCs") +  
 theme\_minimal()



# How does the frequeny / number of DLC’s for a game effect user reviews?

The graph shows a general disconnect from the number of DLC’s compared to the number of positive reviews. The graph does show us some valuable points though, we can see that for games with a very high review percentage there is a general look that we can see that these games tend to recieve small numbers of DLC’s which indicates to me that the developers of these games continued support for these games after release due to positive perception. We can also see some games with a lot of DLC’s that have a lower review percentage which indicates to me that these games are being milked for money and that the developers are not putting in the effort to make the game better. This is a very interesting trend as it shows that the number of DLC’s does not always correlate with the positive reviews and that some games can be successful with a small number of DLC’s while others can be unsuccessful with a large number of DLC’s.

# Filter the dataset:  
# - Remove free games (price > 0)  
# - Ensure positive review percentage is between 40-100  
# - Take the top 200 games based on total reviews  
steam\_games\_filtered <- steam\_games %>%  
 filter(price > 2, pct\_pos\_total >= 40 & pct\_pos\_total <= 100) %>%  
 arrange(desc(num\_reviews\_total)) %>%  
 head(200) # Take the top 200 games based on total reviews  
  
# Create the plot  
ggplot(steam\_games\_filtered, aes(x = price, y = pct\_pos\_total)) +  
 geom\_point(alpha = 0.6, color = "#2ca02c") + # Set point transparency and color  
 geom\_smooth(method = "loess", color = "blue", se = TRUE) + # Add trend line  
 scale\_x\_continuous(trans = "log10", breaks = c(1, 5, 10, 20, 40, 60, 100)) + # Log scale for price  
 scale\_y\_continuous(limits = c(40, 100), breaks = seq(40, 100, 10)) + # Keep review % between 40-100  
 labs(  
 title = "Game Price vs Positive Review Percentage (Top 200 Most Reviewed Games)",  
 subtitle = "Log-scaled price to improve readability",  
 x = "Game Price (USD)",  
 y = "Percentage of Positive Reviews"  
 ) +  
 theme\_minimal(base\_size = 14) + # Set base font size for better readability  
 theme(  
 plot.title = element\_text(face = "bold", size = 16), # Bold and size up the title  
 plot.subtitle = element\_text(size = 12), # Size the subtitle  
 axis.title = element\_text(face = "bold", size = 14), # Bold and size up axis titles  
 axis.text = element\_text(size = 12), # Size up axis text  
 panel.grid.major = element\_line(color = "gray90"), # Lighten major grid lines for cleaner look  
 panel.grid.minor = element\_line(color = "gray95") # Lighten minor grid lines for a softer look  
 )

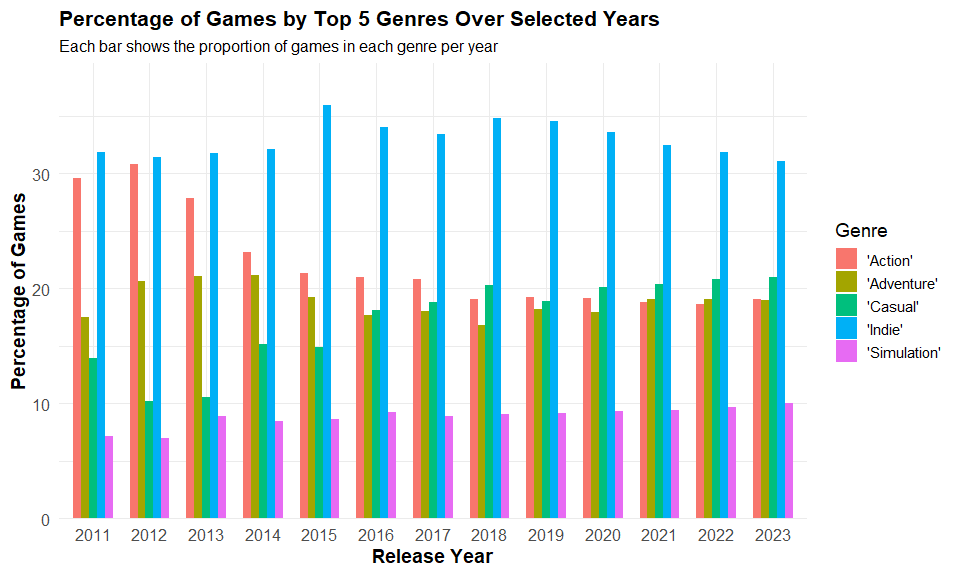


# How does the pricing of games affect user reviews?

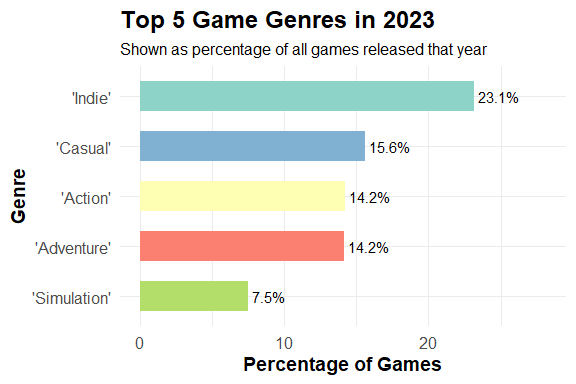
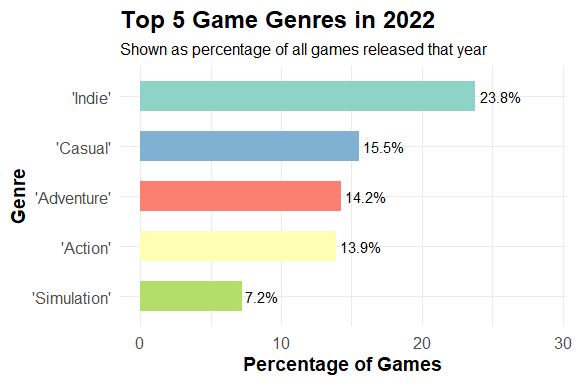
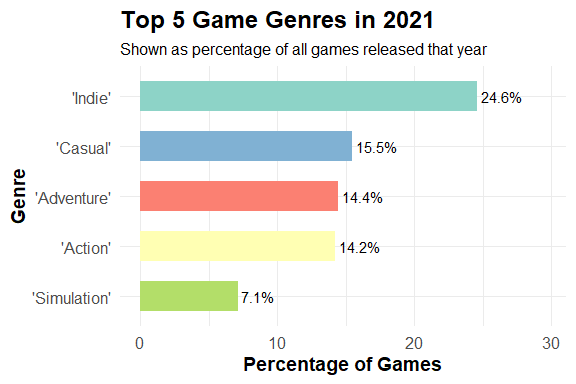
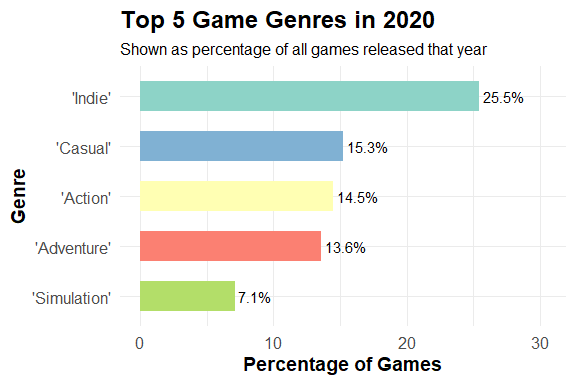
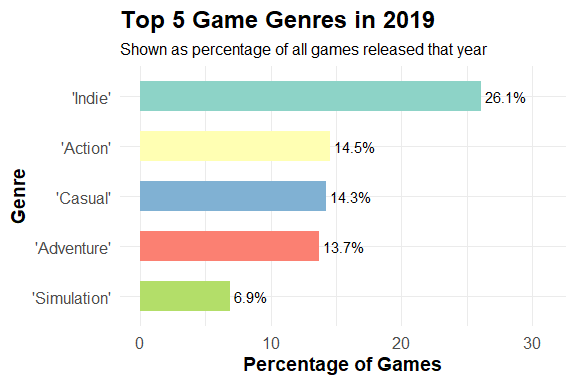
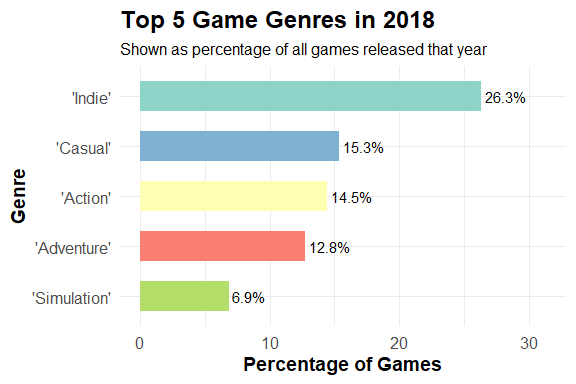
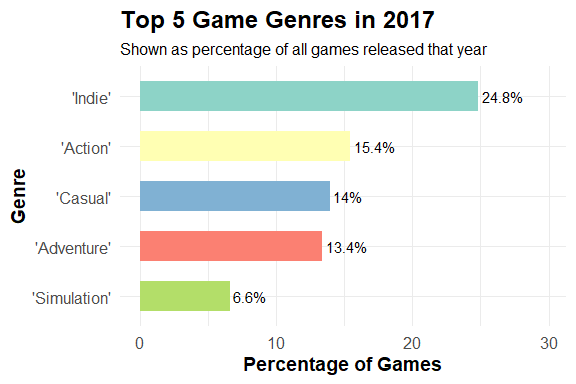
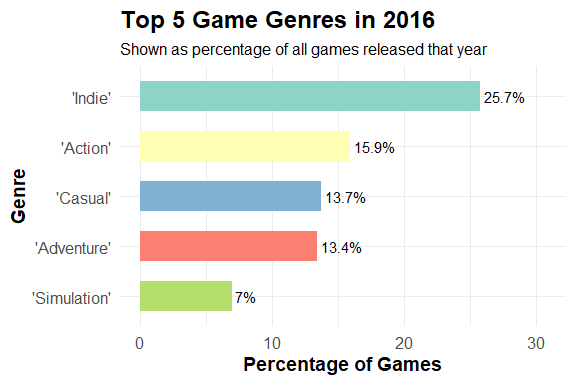
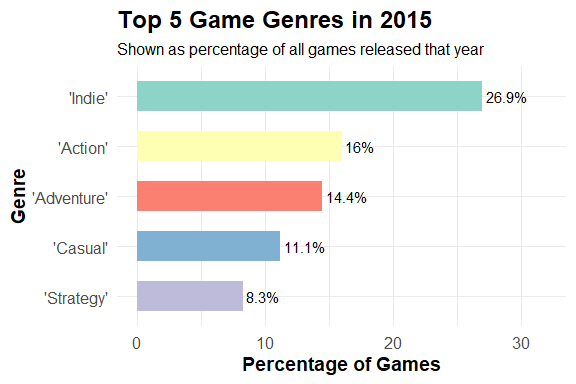
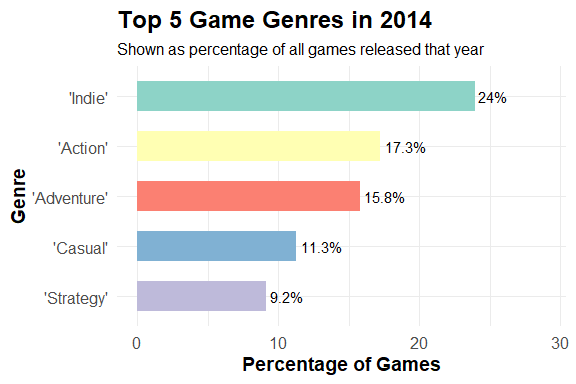
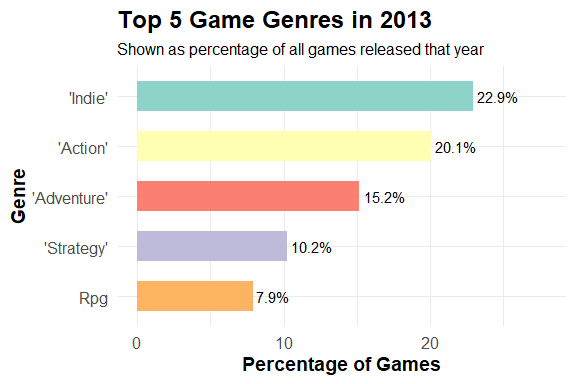
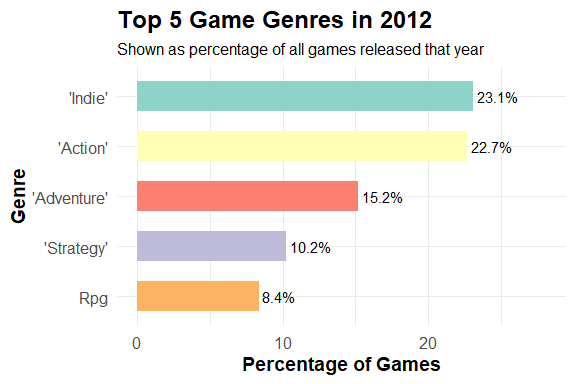
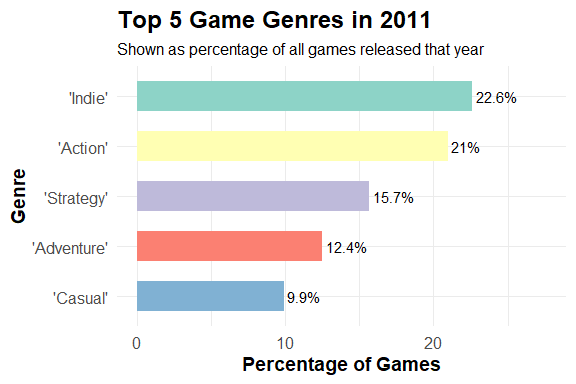
Taking a deeper dive into the meaning of this graph I can see that there is a non-linear relationship between the price of games and the percentage of positive reviews. The graph shows the very cheap games have a slightly lower positive percentage than the games in the range of about 20-30. But we also see as games get above this range players seem to be more harsh in the reviews as I think that these players have a higher expectation of the game based on the price that they paid. We can also see that there is outliers among the data as there is low rated games in in all ranges which shows that no matter the price games can be rated low. But in general we have seen that the price of games does have an impact on player perception and forgiveness for flaws within games.

# Convert release\_date to year  
steam\_games <- steam\_games %>%  
 mutate(release\_year = year(ymd(release\_date)))  
  
steam\_games <- steam\_games %>%  
 mutate(  
 release\_year = year(ymd(release\_date)),  
 genres = str\_replace\_all(genres, '\\[|\\]|\"', ""), # Remove brackets and quotes  
 genres = str\_trim(genres) # Trim whitespace  
 ) %>%  
 separate\_rows(genres, sep = ",") %>%  
 mutate(  
 genres = str\_to\_lower(str\_trim(genres)), # Standardize case  
 genres = if\_else(str\_detect(genres, "rpg"), "RPG", genres), # Group all RPG-related genres  
 genres = str\_to\_title(genres)   
 )  
  
# Get top 5 most common genres overall  
top\_genres\_raw <- steam\_games %>%  
 filter(!is.na(genres)) %>%  
 separate\_rows(genres, sep = ",") %>%  
 mutate(genres = str\_trim(genres)) %>% # trim again after splitting  
 group\_by(genres) %>%  
 summarize(total\_games = n(), .groups = "drop") %>%  
 arrange(desc(total\_games))  
  
# Remove duplicates and take the top 5 unique ones  
top\_genres <- top\_genres\_raw %>%  
 distinct(genres, .keep\_all = TRUE) %>%  
 slice\_head(n = 5) %>%  
 pull(genres)  
  
# Define selected years  
selected\_years <- c(2011,2012,2013, 2014, 2015, 2016, 2017, 2018, 2019, 2020, 2021, 2022, 2023)  
  
# Filter and reshape for genre percentage per year  
steam\_games\_filtered <- steam\_games %>%  
 filter(release\_year %in% selected\_years, !is.na(genres)) %>%  
 separate\_rows(genres, sep = ",") %>%  
 filter(genres %in% top\_genres) %>%  
 group\_by(release\_year, genres) %>%  
 summarize(genre\_count = n(), .groups = "drop")  
  
# Count total games released each selected year (with any genre)  
total\_per\_year <- steam\_games %>%  
 filter(release\_year %in% selected\_years) %>%  
 separate\_rows(genres, sep = ",") %>%  
 filter(genres %in% top\_genres) %>%  
 group\_by(release\_year) %>%  
 summarize(total\_games = n(), .groups = "drop")  
  
# Join and calculate percentage  
genre\_percentages <- steam\_games\_filtered %>%  
 left\_join(total\_per\_year, by = "release\_year") %>%  
 mutate(percentage = 100 \* genre\_count / total\_games)

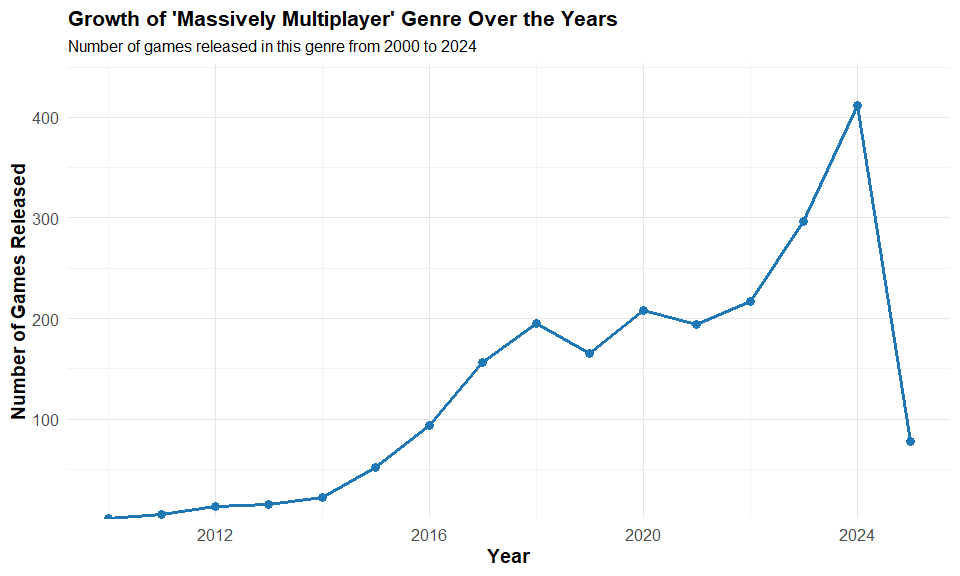
ggplot(genre\_percentages, aes(x = factor(release\_year), y = percentage, fill = genres)) +  
 geom\_bar(stat = "identity", position = "dodge", width = 0.7) +  
 scale\_y\_continuous(expand = expansion(mult = c(0, 0.1))) +  
 labs(  
 title = "Percentage of Games by Top 5 Genres Over Selected Years",  
 subtitle = "Each bar shows the proportion of games in each genre per year",  
 x = "Release Year",  
 y = "Percentage of Games",  
 fill = "Genre"  
 ) +  
 theme\_minimal(base\_size = 14) +  
 theme(  
 plot.title = element\_text(face = "bold", size = 16),  
 plot.subtitle = element\_text(size = 12),  
 axis.title = element\_text(face = "bold", size = 14),  
 legend.position = "right",  
 axis.text.x = element\_text(size = 12),  
 axis.text.y = element\_text(size = 12)  
 )



# Get all top genres across selected years  
top\_genres\_all\_years <- steam\_games %>%  
 filter(release\_year %in% selected\_years) %>%  
 count(release\_year, genres, sort = TRUE) %>%  
 group\_by(release\_year) %>%  
 slice\_max(n, n = 5) %>%  
 ungroup() %>%  
 distinct(genres) %>%  
 pull(genres)  
  
# Assign fixed colours  
genre\_palette <- setNames(RColorBrewer::brewer.pal(n = length(top\_genres\_all\_years), "Set3"), top\_genres\_all\_years)  
  
for (yr in selected\_years) {  
 year\_data <- steam\_games %>%  
 filter(release\_year == yr)  
  
 total\_games <- nrow(year\_data)  
  
 top5 <- year\_data %>%  
 count(genres, sort = TRUE) %>%  
 mutate(percentage = 100 \* n / total\_games) %>%  
 slice\_max(percentage, n = 5)  
  
 p <- ggplot(top5, aes(x = reorder(genres, percentage), y = percentage, fill = genres)) +  
 geom\_col(width = 0.6) +  
 geom\_text(aes(label = paste0(round(percentage, 1), "%")), hjust = -0.1, size = 4) +  
 coord\_flip() +  
 scale\_fill\_manual(values = genre\_palette) + # 👈 Use consistent colors  
 labs(  
 title = paste("Top 5 Game Genres in", yr),  
 subtitle = "Shown as percentage of all games released that year",  
 x = "Genre",  
 y = "Percentage of Games"  
 ) +  
 theme\_minimal(base\_size = 14) +  
 theme(  
 plot.title = element\_text(face = "bold", size = 18),  
 plot.subtitle = element\_text(size = 12),  
 axis.title = element\_text(face = "bold", size = 14),  
 axis.text = element\_text(size = 12),  
 legend.position = "none"  
 ) +  
 scale\_y\_continuous(limits = c(0, max(top5$percentage) + 5))  
  
 print(p)  
}



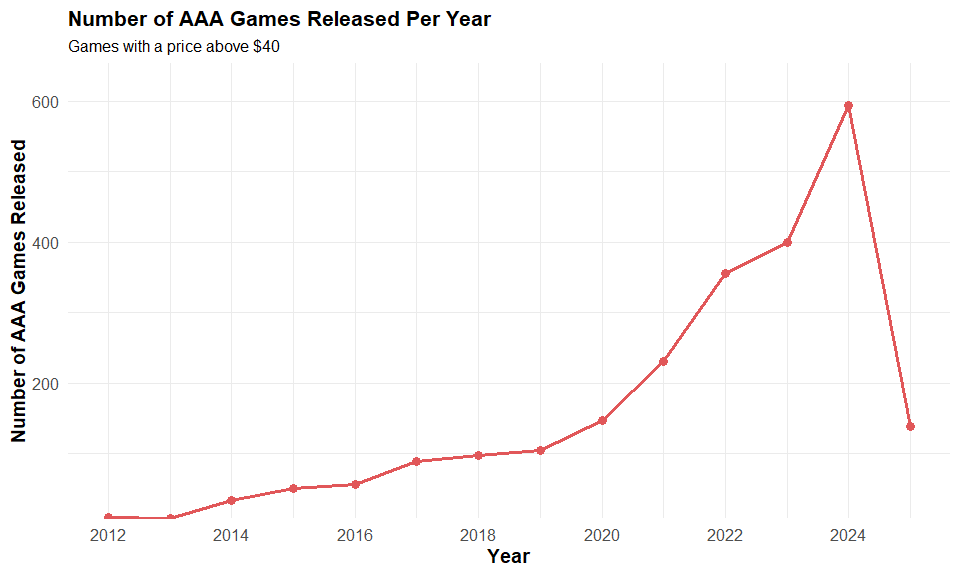
# Convert release\_date to year  
steam\_games <- steam\_games %>%  
 mutate(release\_year = year(ymd(release\_date)))   
  
# Clean genre column: Remove brackets and extra spaces  
steam\_games <- steam\_games %>%  
 mutate(genres = str\_replace\_all(genres, "[\\[\\]']", "")) %>% # Remove square brackets and quotes  
 mutate(genres = str\_trim(genres)) # Remove leading/trailing spaces  
  
# Filter dataset for "Massively Multiplayer" genre  
massively\_multiplayer\_growth <- steam\_games %>%  
 filter(!is.na(genres)) %>%  
 separate\_rows(genres, sep = ", ") %>% # Split genres correctly  
 filter(genres == "Massively Multiplayer") %>% # Only keep this genre  
 group\_by(release\_year) %>%  
 summarize(game\_count = n(), .groups = "drop")   
  
# Create the line chart  
ggplot(massively\_multiplayer\_growth, aes(x = release\_year, y = game\_count)) +  
 geom\_line(color = "#1f77b4", size = 1.2) +  
 geom\_point(color = "#1f77b4", size = 3) +  
 scale\_x\_continuous(breaks = seq(2000, 2024, 4)) +  
 scale\_y\_continuous(expand = expansion(mult = c(0, 0.1))) +  
 labs(  
 title = "Growth of 'Massively Multiplayer' Genre Over the Years",  
 subtitle = "Number of games released in this genre from 2000 to 2024",  
 x = "Year",  
 y = "Number of Games Released"  
 ) +  
 theme\_minimal(base\_size = 14) +  
 theme(  
 plot.title = element\_text(face = "bold", size = 16),  
 plot.subtitle = element\_text(size = 12),  
 axis.title = element\_text(face = "bold", size = 14),  
 axis.text = element\_text(size = 12),  
 panel.grid.major = element\_line(color = "gray90"),  
 panel.grid.minor = element\_line(color = "gray95")  
 )



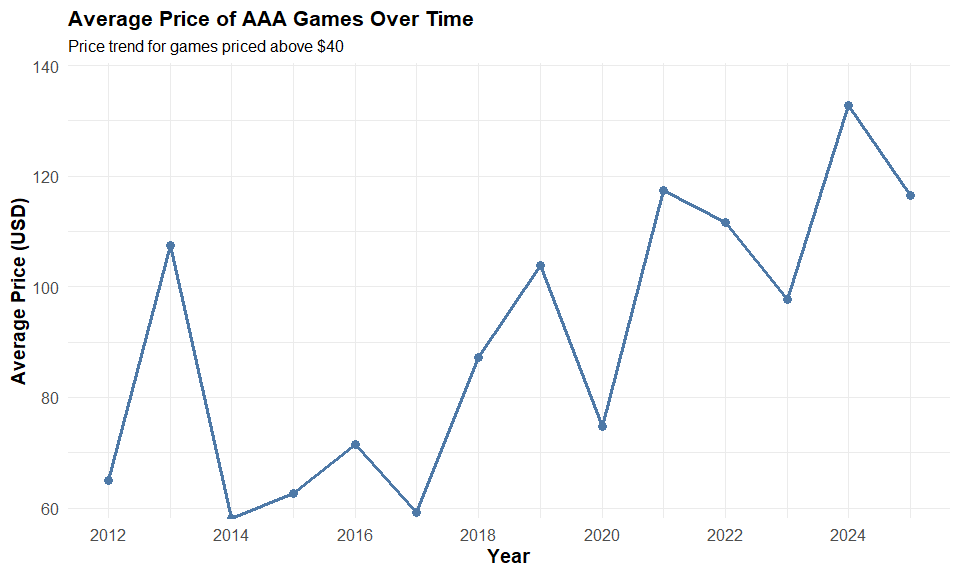
# What genres have had trending growth over the years?

These graphs can show us a few interesting things about the trends of game genres over the years, In the first graph we see the percentage of games released in the top 5 genres over the years. We can see the dominance of action games in the early 2010’s and how the genre has dropped off slightly in recent years. We can also see the rise of casual games which in turn shows the growth of the gaming industry in these years as more and more casual gamers have started playing which has shifted developers focus to jump into the casual markets as shown in the graph. We can also see a large amount of Indie games released more often by smaller developers trying to make a name for themselves which is still a very steady trend in the industry. We can also see the rise of Massively Multiplayer games in the last few years which is a very interesting trend as it shows that with the increase in computational power and internet speeds that these games are becoming more and more popular.I feel these graphs show a very interesting oversight into the world of game genres and the trends that have come and gone over the years.

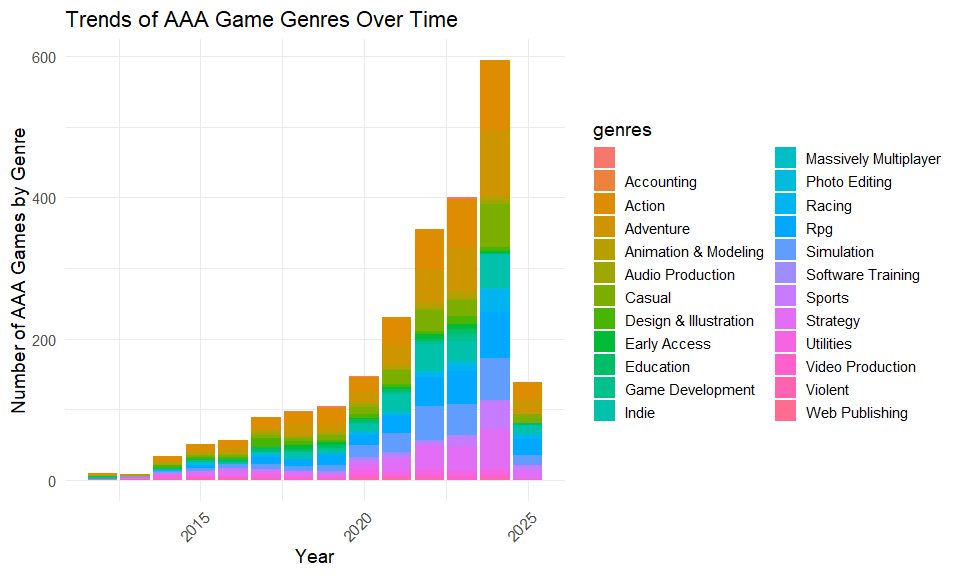
# Convert release\_date to year  
steam\_games <- steam\_games %>%  
 mutate(release\_year = year(ymd(release\_date)))  
  
# Keep full AAA dataset with genre info  
aaa\_games <- steam\_games %>%  
 filter(price > 40, !is.na(release\_year), !is.na(genres)) %>%  
 select(appid, name, release\_year, price, genres)  
  
# Create summary table: number of AAA games per year  
aaa\_summary <- aaa\_games %>%  
 group\_by(release\_year) %>%  
 summarize(num\_aaa\_games = n(), .groups = "drop")  
  
# Create the line chart  
ggplot(aaa\_summary, aes(x = release\_year, y = num\_aaa\_games)) +  
 geom\_line(color = "#e15759", size = 1.2) +  
 geom\_point(color = "#e15759", size = 3) +  
 scale\_x\_continuous(breaks = seq(min(aaa\_games$release\_year), max(aaa\_games$release\_year), 2)) +  
 scale\_y\_continuous(expand = expansion(mult = c(0, 0.1))) +  
 labs(  
 title = "Number of AAA Games Released Per Year",  
 subtitle = "Games with a price above $40",  
 x = "Year",  
 y = "Number of AAA Games Released"  
 ) +  
 theme\_minimal(base\_size = 14) +  
 theme(  
 plot.title = element\_text(face = "bold", size = 16),  
 plot.subtitle = element\_text(size = 12),  
 axis.title = element\_text(face = "bold", size = 14),  
 axis.text = element\_text(size = 12)  
 )



# Calculate average price of AAA games per year  
aaa\_price\_trend <- steam\_games %>%  
 filter(price > 40, !is.na(release\_year)) %>%  
 group\_by(release\_year) %>%  
 summarize(avg\_price = mean(price, na.rm = TRUE), .groups = "drop")  
  
# Create the line chart  
ggplot(aaa\_price\_trend, aes(x = release\_year, y = avg\_price)) +  
 geom\_line(color = "#4e79a7", size = 1.2) +  
 geom\_point(color = "#4e79a7", size = 3) +  
 scale\_x\_continuous(breaks = seq(min(aaa\_price\_trend$release\_year), max(aaa\_price\_trend$release\_year), 2)) +  
 scale\_y\_continuous(expand = expansion(mult = c(0, 0.1))) +  
 labs(  
 title = "Average Price of AAA Games Over Time",  
 subtitle = "Price trend for games priced above $40",  
 x = "Year",  
 y = "Average Price (USD)"  
 ) +  
 theme\_minimal(base\_size = 14) +  
 theme(  
 plot.title = element\_text(face = "bold", size = 16),  
 plot.subtitle = element\_text(size = 12),  
 axis.title = element\_text(face = "bold", size = 14),  
 axis.text = element\_text(size = 12)  
 )



aaa\_games %>%  
 filter(!is.na(genres)) %>%  
 separate\_rows(genres, sep = ", ") %>%  
 group\_by(release\_year, genres) %>%  
 summarize(count = n(), .groups = "drop") %>%   
 ggplot(aes(x = release\_year, y = count, fill = genres)) +  
 geom\_bar(stat = "identity") +  
 labs(  
 title = "Trends of AAA Game Genres Over Time",  
 x = "Year",  
 y = "Number of AAA Games by Genre"  
 ) +  
 theme\_minimal(base\_size = 14) +  
 theme(axis.text.x = element\_text(angle = 45, hjust = 1))



# How has the AAA landscape changed over the years?

The AAA landscape has changed a lot over the years as we can see from the graphs above. The first graph shows the number of AAA games released per year and we can see that there is a steady increase in the number of AAA games released over the years.This shows that more AAA game developers are showing up or else due to the way I have decided to catagorise AAA games that the increase in price in games in recent years has meant smaller studios games have come up in price in line with older AAA titles. We can see from the graph that the price of these games has steadily risen in recent years with the release of the PS5 and Xbox series X we saw a increase in the price of games to 70 euro which is a large increase from the previous generation of consoles. This has led to a lot of backlash from players as they feel that the games are not worth the price that they are being sold for. This is a similar trend I saw previously in the pricing of games and how it affects user reviews. We can see that the price of AAA games has increased over the years and this has led to a lot of backlash from players as they feel that the games are not worth the price that they are being sold for. This is a similar trend I saw previously in the pricing of games and how it affects user reviews. We can also see that the genres of AAA games have changed over the years with the rise of Massively Multiplayer games for AAA studios as they use these games as potential cash cows for the studio as they use a live-service type game model to keep the game alive and keep players coming back to the game. This is interesting as it shows that the motivation for these big companies is to make money and increase profits for shareholders.

# Conclusion

This report provided an in-depth analysis on various digital games development and distribution factors through analysis of the Steam Games dataset — from the impact of downloadable content (DLC), through pricing models, to genre trends and the AAA game landscape evolution.

The results showed that although an abundance of DLCs does not automatically translate to a better reception from users, continued post-launch support in moderation is generally tied to positively reviewed games. When it comes to pricing, though, the relationship is distinctly non-linear: games sold for a little bit more than the average get better average review scores than games sold at the bottom or the very top of the pricing spectrum, while pricier games are held to higher standards and typically received more brutal reviews on average.

The evolution of genres over the years mirrored the broader changes in the gaming landscape. I explored the constant presence of action and indie titles, the proliferation of casual games that could be attributed to new gamers and mobile gamers, as well as the rising popularity of massively multiplayer genres — all signaling the market and technological capability changing in their makings.

Finally, the AAA space has changed a lot. Given the growing tension between player expectations and studio strategies as games continue to sell for higher prices, face more scrutiny, and pivot to evermore live-service work and monetisation, it ought to be no surprise that the data reveals a difference between expectations and reality. This reinforces the need for developers — and particularly those in the AAA market — to toe the line between money-grabbing practices and value-oriented game design.

Overall this shows that with the help of data, developers, publishers and even gamers can better understand what is really going on in our industry, where they can fit themselves into it as we enter an exciting new period.

# References

* Steam Games Dataset 2025: [Steam Games Dataset](https://www.kaggle.com/datasets/nikhilroxtomar/steam-games-dataset-2025)