CS732 - Data Visualization
Assignment 4 Report

Team: Abracadatavis

Kappagantula Lakshmi Abhigna - IMT2019040
Vaibhavi Lokegaonkar - IMT2019090
Vijay Jaisankar- IMT2019525

Introduction

We used the STEAM games dataset for this assignment. In this report, we have analyzed various aspects of video games on the STEAM store using the Tableau Visualization Tool.

Task 1: Developer Statistics Dashboard

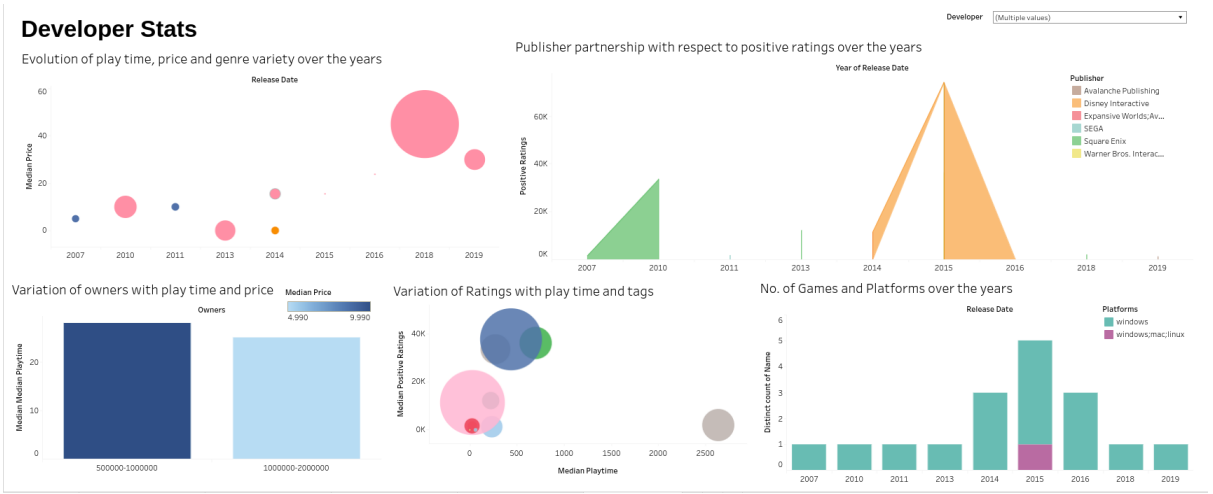


Figure 1: Dashboard Detailing a Developer Organisation's Statistics

For this task, we have developed a Tableau Dashboard for a particular game developer organization, which has the following use cases:

a. A game developer could use the dashboard to pitch their designed game to a publisher organization, which could potentially distribute it on a platform like STEAM.

b. A game developer individual can see the statistics of another game developer who could be an individual or organization. This could help the individual developers

attempt to get hired by the organization or collaborate with an individual developer by facilitating research on the latter progress in the game development field.

c. A game publishing organization looking for developers in order to publish a game can use this dashboard to compare developers and choose the best one for their next rollout.

As shown in Figure 1, the drop-down on the top-right corner can be used to choose developer(s), whose development feats are visualized in the dashboard.

Evolution of play-time and genre over the years

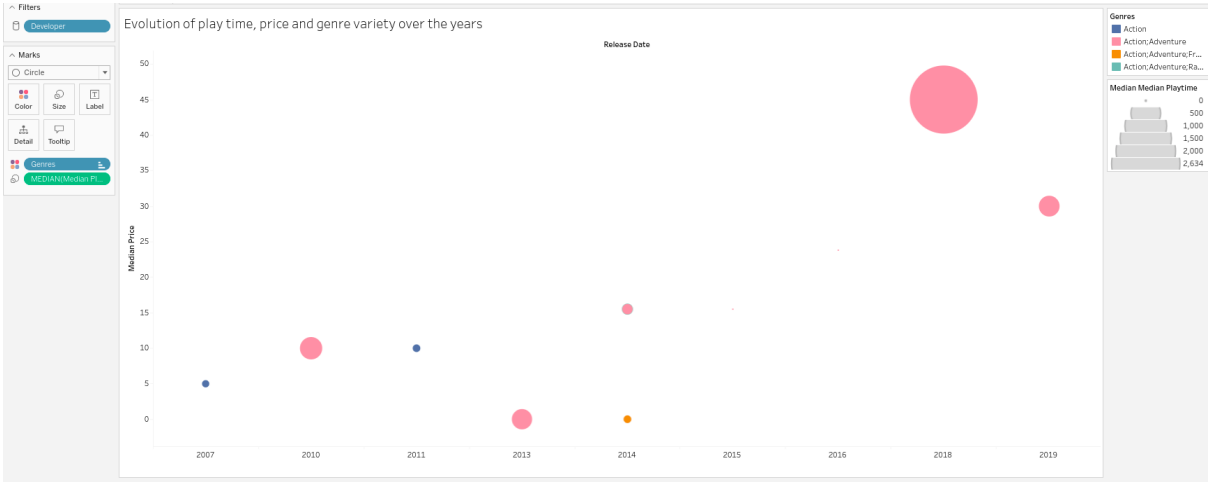


Figure 2: Evolution of playtime, price, and genre over the years for Valve

Figure 2 shows the scattered bubble plot, where the horizontal axis proceeds in time in years, and the vertical axis shows the median price of games developed during a particular year by a developer chosen using the filter. The bubbles' color denotes the genre of games developed while the size denotes the median playtime of the games that are represented by a particular bubble on the plot. The size should convey information having a certain order while color merely can convey different categories. Hence, the denotation was chosen as above.

The above visualization was made by choosing the developer as Valve, the game development organization that also owns STEAM. From the above visualization, we observe the following inferences:

a. Valve, as a game organization, generally develops action-adventure games in the nearer past. While it can also be observed that earlier the game genre was also having a fair share of plain action games.

b. As the years have progressed, the play-time, i.e. the amount of time, gamers have spent playing the game has significantly increased as seen from the increase in the size of the bubbles as we go along the positive horizontal axis.

c. In addition to the play-time, the price of the game has increased. Since both these parameters have increased, we can conclude that the value and quality (referred from the increase in the play-time) have increased over time. This makes Valve, a valued and quality game developer organization.

Publisher Partnership with respect to Positive Ratings over the years

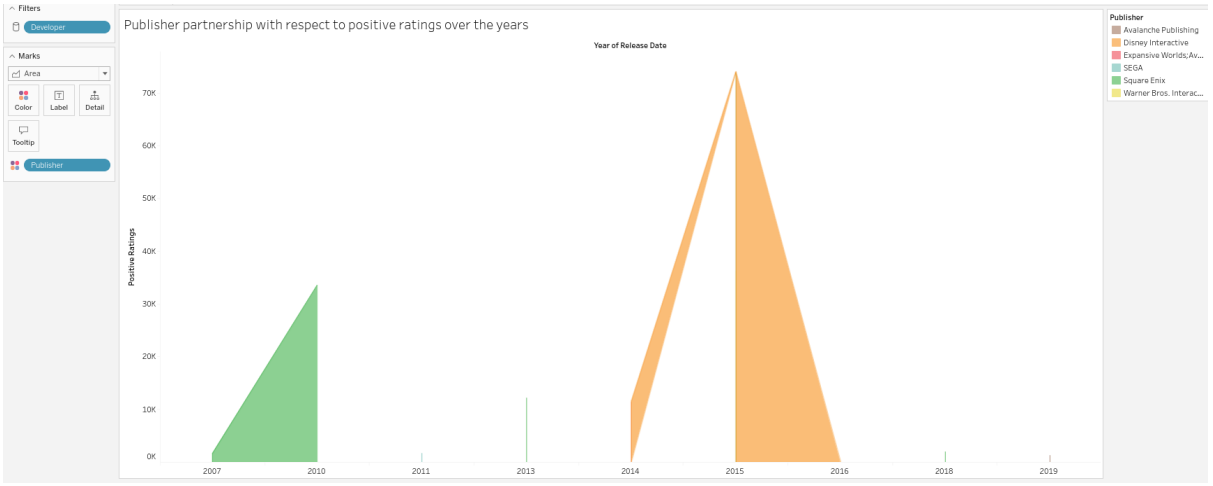


Figure 3: Publisher Partnership with Avalanche has been shown by visualizing positive ratings over the years

An effective partnership between a developer and a publisher organization shows the kind of relationship the developer has with the publisher. This can be an important point for a publishing organization to consider while making a deal with a developer. A consistent partnership, with a very high positive rating, over a long period of time is a good indicator of commitment along with good quality development of games.

The visualization in Figure 3 aims to show the effectiveness of the partnership between Valve and its publishers over the years. The vertical axis depicts positive ratings while the horizontal axis progresses through time in years. For a clearer depiction of ratings, the area covered by the curve joining the points has been colored based on the different publishers Valve has published its games with. The color map used in this visualization is the default setting of Tableau.

The following are the inferences that are taken from the visualization about the game development organization, Avalanche:

a. Avalanche has been associated with several publishing organizations in short time stints over the past year, this shows the commitment of Avalanche, as an organization to its publishing organization.

b. However, the positive ratings of all the games developed have been particularly high, hence indicating good quality. The highest positive rating has been achieved during Avalanche's collaboration with Disney Interactive.

Variation of owners with play-time and price

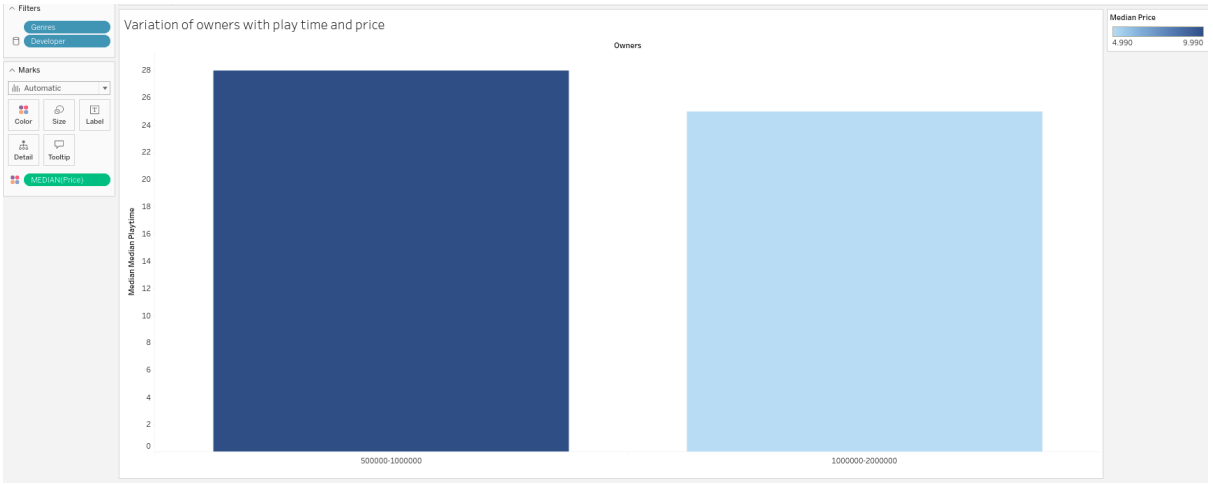


Figure 4: Variation of owners with play-time and price of games developed by Avalanche

Figure 4 shows the visualization of owners with play-time and price. The number of owners has been indicated in the dataset itself as a range of values. We have harnessed those values and put them on the horizontal axis, while the vertical axis indicates the playtime of the games developed by Avalanche. Additionally, the color of the bars is varied by price. The lighter the blue, the lower the price of the game.

The following are the inferences from the above plot:

a. As the number of owners has increased, the price of the games has decreased as is intuitive from general consumer behavior patterns.

b. However, when the price of the games developed by Avalanche is high, the play-time is also high. This shows that the games which are costlier are better in quality and deliver better engagement from their players. This shows that the value of the extra money paid has been well delivered by the games developed by Avalanche.

Variation of ratings with play-time and tags

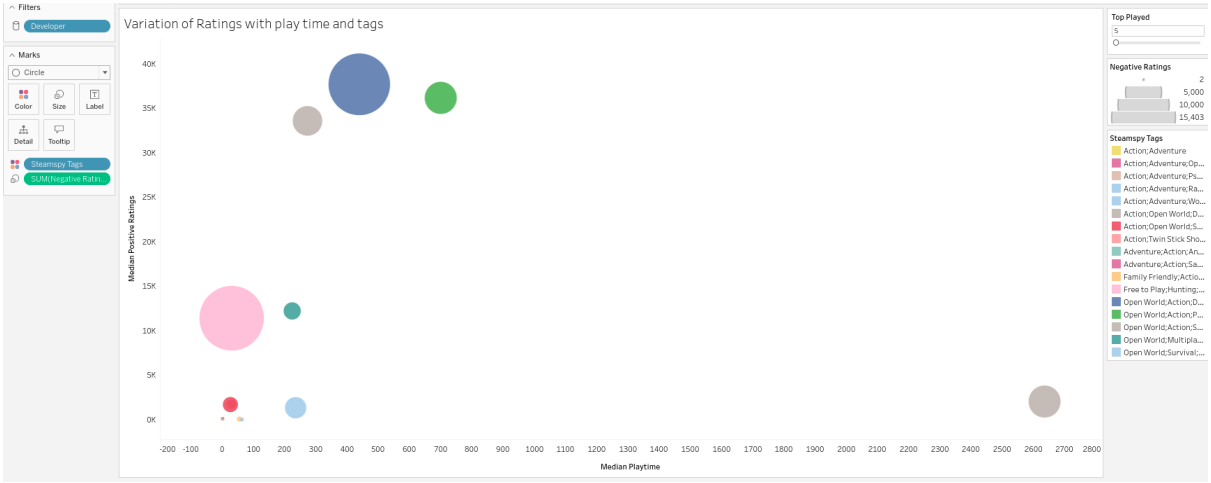


Figure 4: Variation of ratings with play-time and tags given by STEAM of games developed by Avalanche

The above scatter plot shows the variation of ratings with play-time and STEAM allocated tags. The horizontal axis shows the median play-time and the vertical axis shows the positive ratings of games. The color of the scatter bubbles shows the various tags allocated by STEAM to the games, while its size is associated with the negative ratings received by the game.

We have made the following inferences from the above visualization:

a. In all of the games developed by Avalanche, the playtime has been around the same range of values except for one which has been the outlier. However, the number of positive ratings as well as negative ratings is not that high. This implies that the game is not popular enough.

b. In most of the games, we observe the number of positive ratings is almost similar to that of negative ratings. There are also exceptions to this rule, where the number of negative ratings significantly surpasses the number of positive ratings. This implies that the gamers who play games developed by Avalanche have, in general, a neutral response to the game.

Number of games and platforms over the years

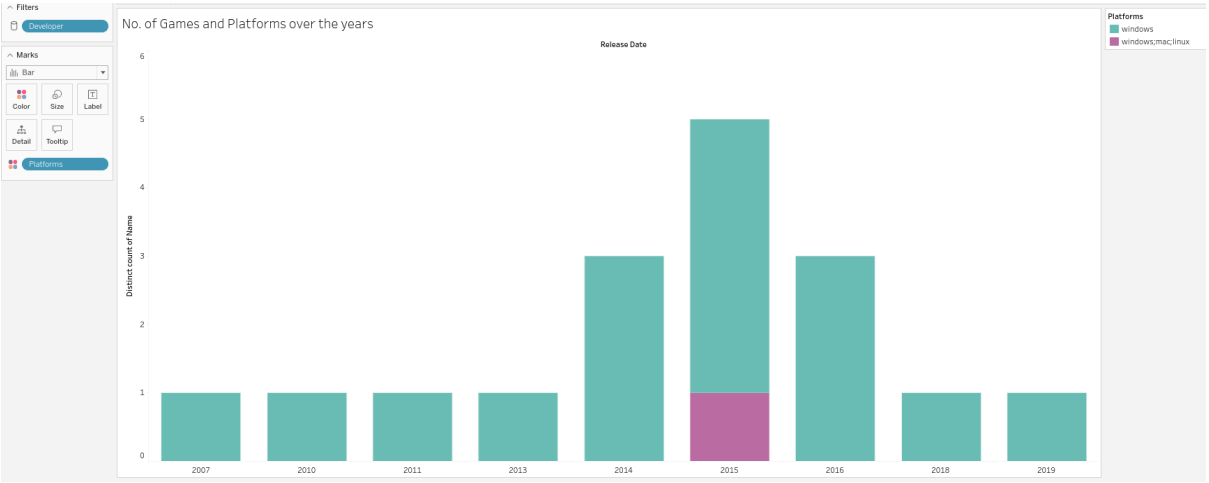


Figure 5: The variation of games and their platforms over the years

Figure 5 shows the variation of the fraction of games developed for each platform over the years. The horizontal axis goes through the years while the vertical axis denotes the number of games developed by Avalanche. The color is used to depict the different platforms.

The following are the inferences made from the above visualization:

a. We observe that the games developed by Avalanche are only for gamers who prefer gaming on the Windows Operating System.

b. Although, there has been a rare occurrence of Mac compatibility in games developed in the year 2015.

Task 2: Steam executives checking which developer to work with

In this task, we look from the perspective of STEAM executives who want to decide which developers they want to work with next based on the games created by them. Before staring with the visualisations, the following observations were made.

a. The ownership field in the file ‘steam.csv’ is not a numerical value, but instead consists of ranges. Also, to find the revenue generated by each game from its time of release to 2019 when the data was scraped, we need the number of owners of the game. So, we considered the midpoints of the range given as the number of owners.

b. The revenue generated by a game from it’s time of release was calculated using:

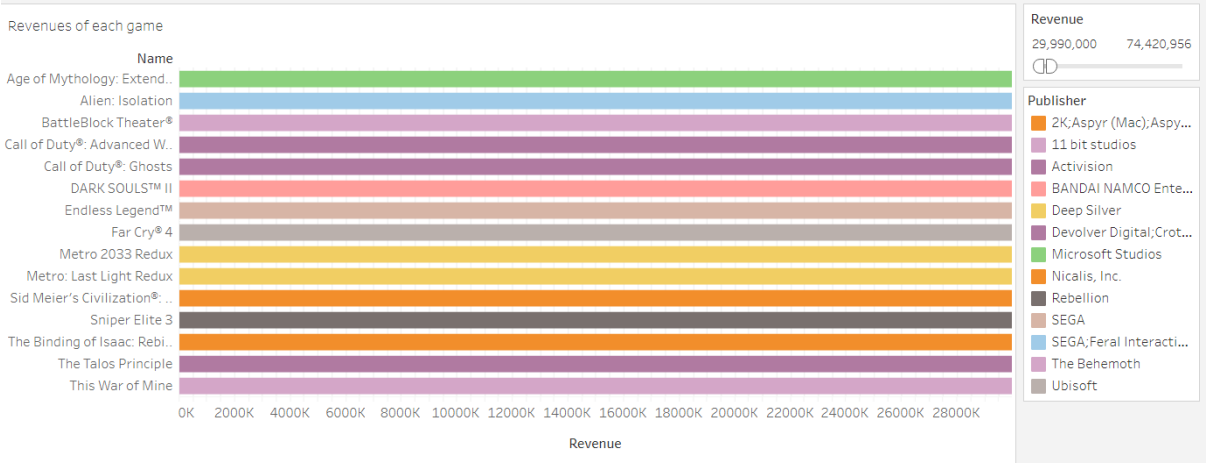
*Revenue = number of owners * price of the game*

Where, as discussed in the previous point, the number of owners is the midpoint.

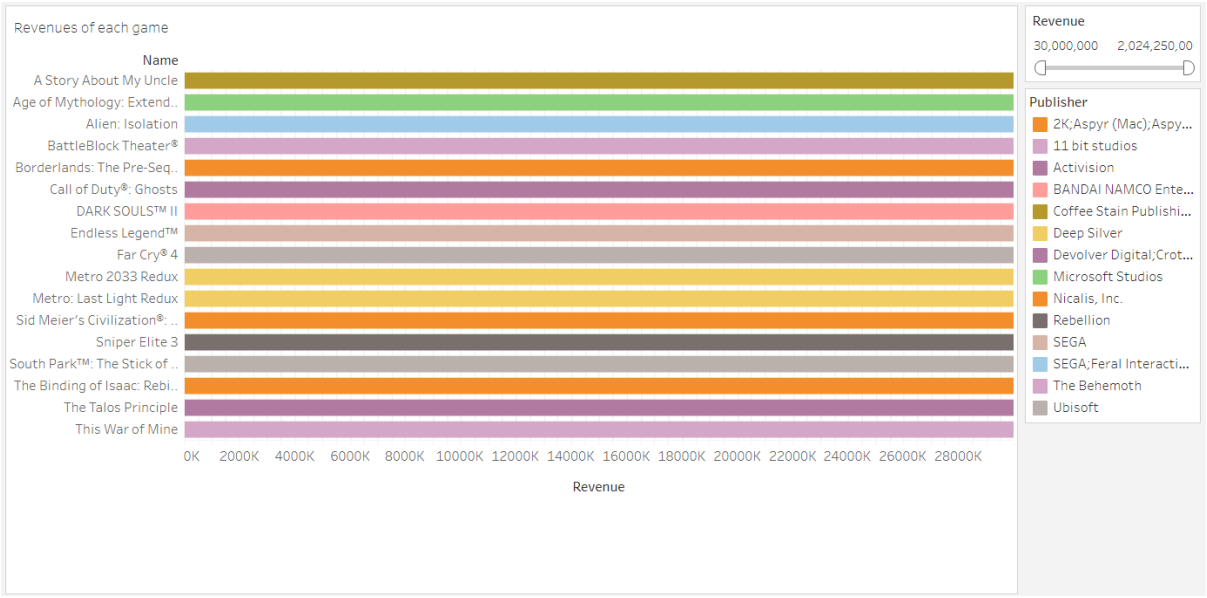
c. We attached this column to the original ‘steam.csv’ and created a new ‘steam_new.csv’ which was uploaded to tableau.

Revenues of all the games

To have an understanding of the values we were working with, we created a visualization taking the game name and the revenue generated on the x-axis. The color denoted the developer of the game. The visualization is shown below.



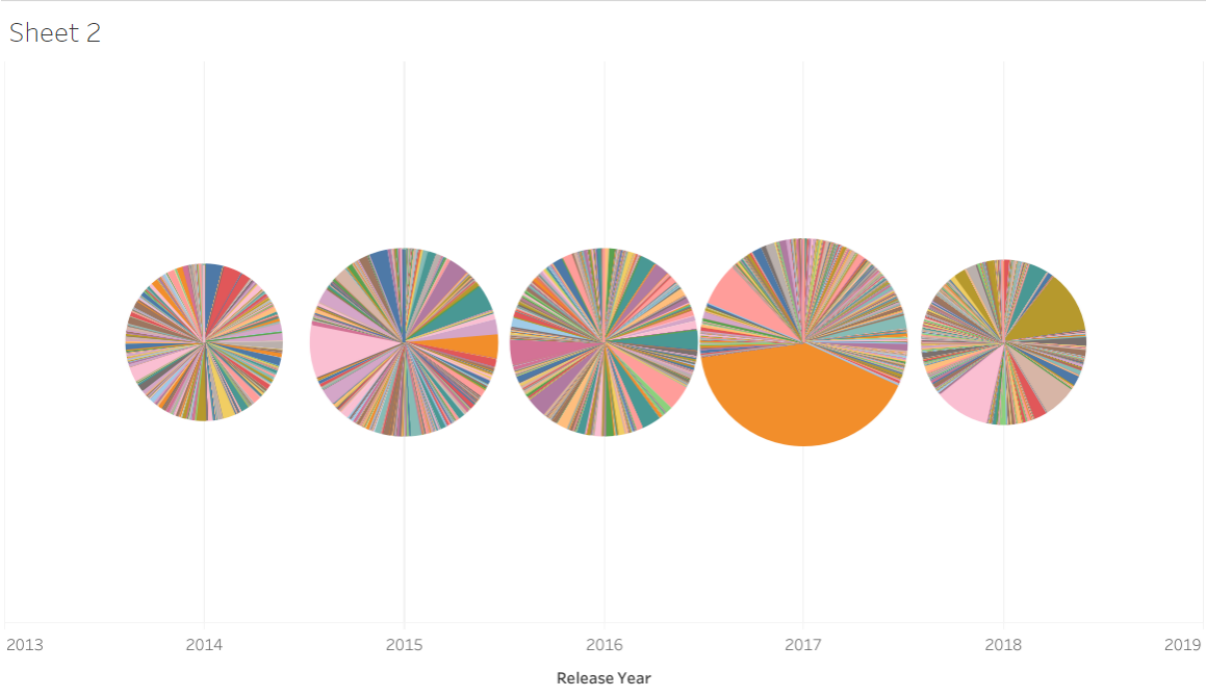
As a company executive looking to make higher revenues, we decided to filter on the basis of games having a revenue greater than $30 M. This is shown below:



This view does show us the highest revenues, if we filter along with developers encoded using colors, but those are not the only parameters involved. Also, there is no year involved here, so this is just to have a general idea of the revenues so that we can use the proper filters in the next few visualizations.

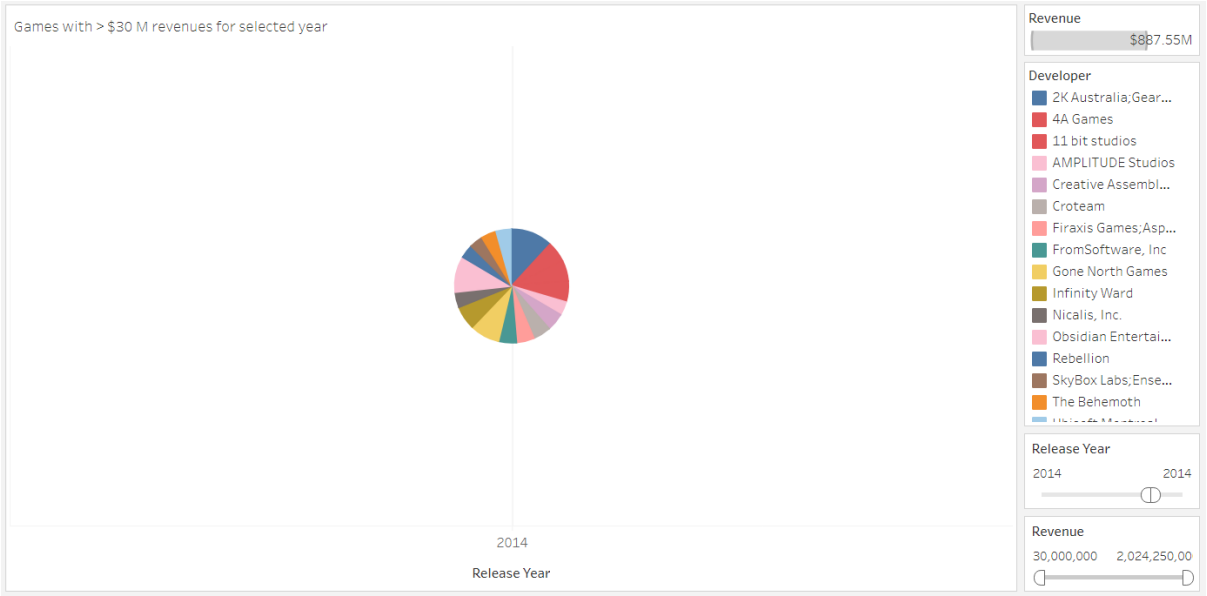
Game revenues over the years

The below plot shows revenues of several games over the years 2014 - 2018. The pie chart is based on the values of the revenues and the colors denote the developers. Here, a user can hover over any game they want to and see the revenue, developer, publisher and name of the corresponding game. Again, this is a general one to help choose a price range for the next visualizations. (This has not been included in the dashboard).



Games with revenues > $30M in the year 2014

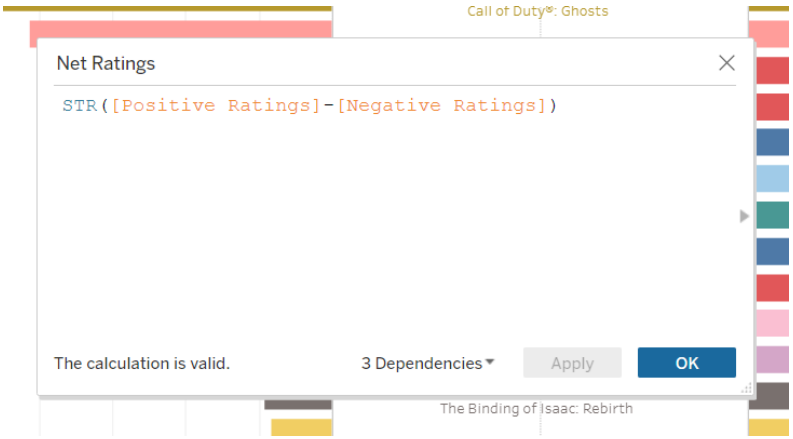
By filtering and plotting the revenues (> $30M. This value was chosen by observing the revenues from the games in the previous plot) of games released in 2014, we obtain the following:



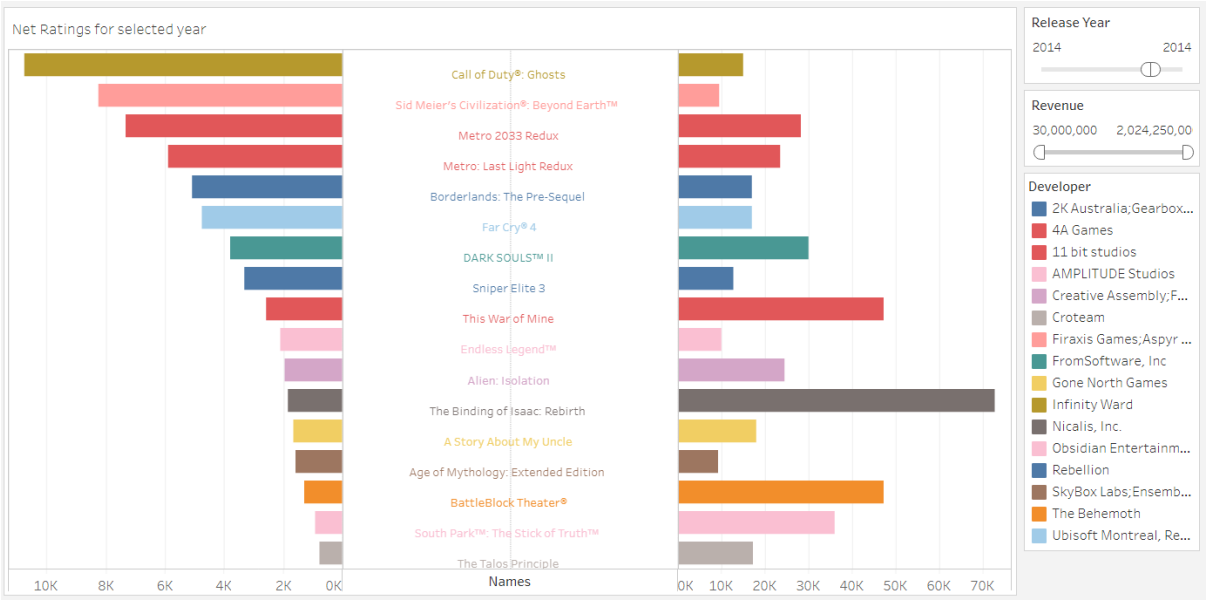
Here we can see that the game developed by 11 bit studios has the highest revenue in that year. But then again, the only way to measure a game's popularity is not the revenue. We also need to check the ratings. Here, in our dataset, we were given two columns: 'positive_ratings' and 'negative ratings' that denoted the number of positive and negative ratings given to each game respectively. But what we needed was the net positive i.e., the net ratings which we tried to visualize.

Games with revenues > $30M in the year 2014 and their ratings

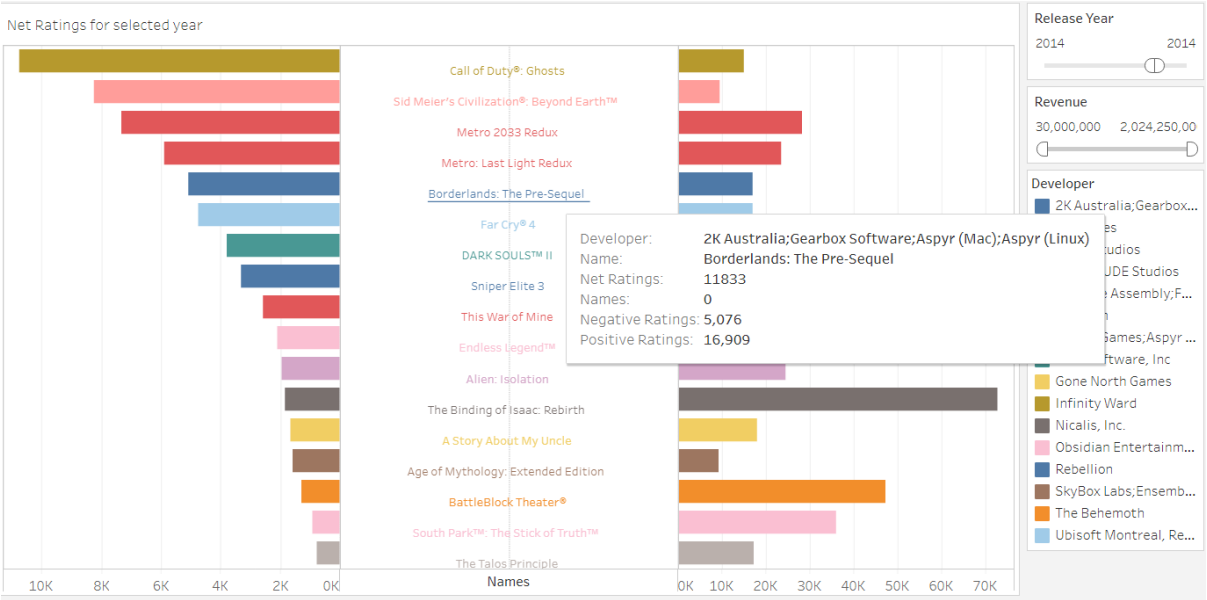
So we had to find the difference in the number of positive and negative ratings for each game. This was done by defining a calculation parameter:



The graph that can be used to observe the ratings of the games (filter revenue > $30M and year 2014 applied) is below.



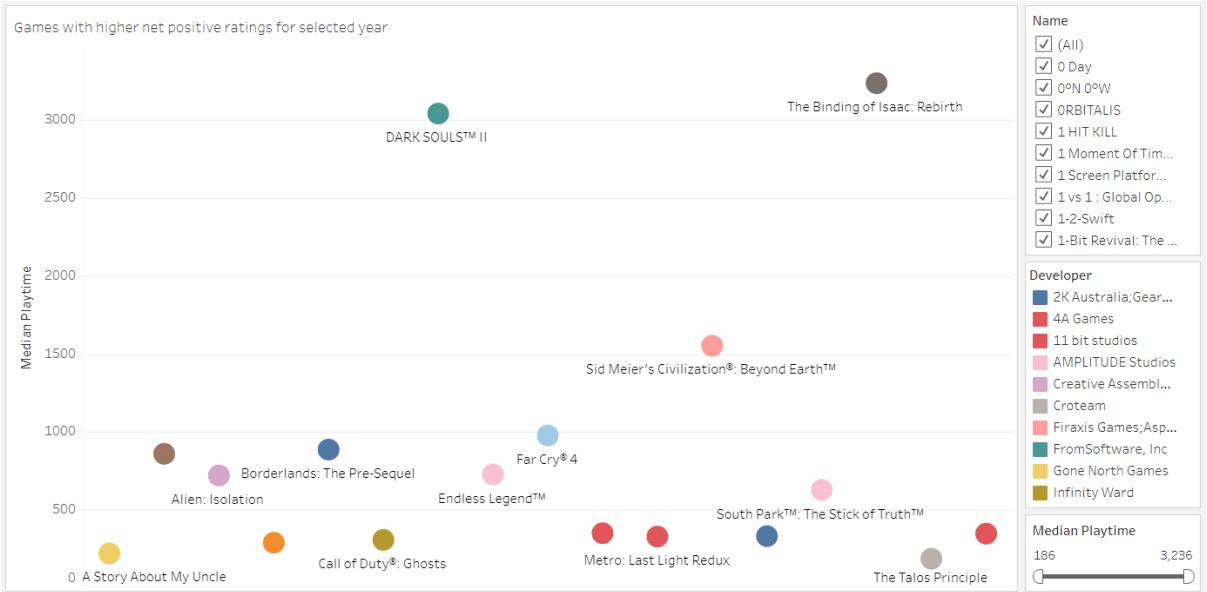
Since we did not want to crowd the visualization, we decided not to directly display this net result on the graph. Instead, by hovering over the name of the game, the net ratings are displayed.



So, the STEAM executive can see the net ratings of each game and decide on the basis of the number of net positive ratings given and the revenue. The next visualization further builds on this, offering an option to select each game they want to, one by one (filter revenue > $30M and year 2014 already applied) based on the ratings they see fit. Again, the color here is developer.

Games (with all previous conditions) chosen based on playtime

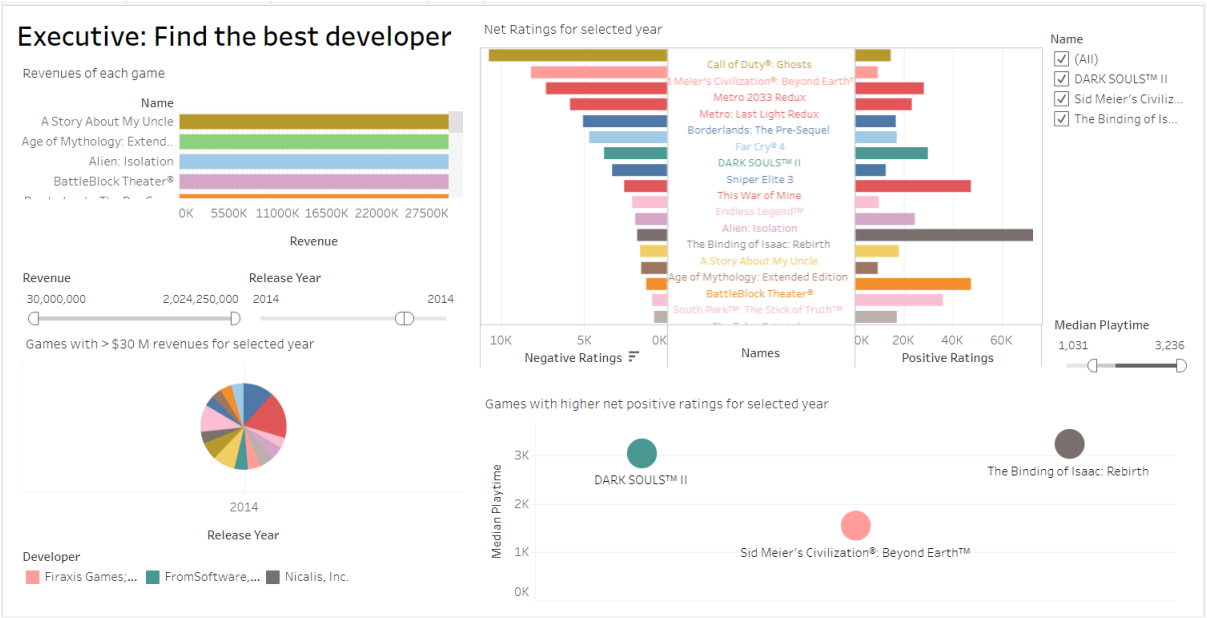
Another important parameter to gauge the popularity of a game is the playtime.



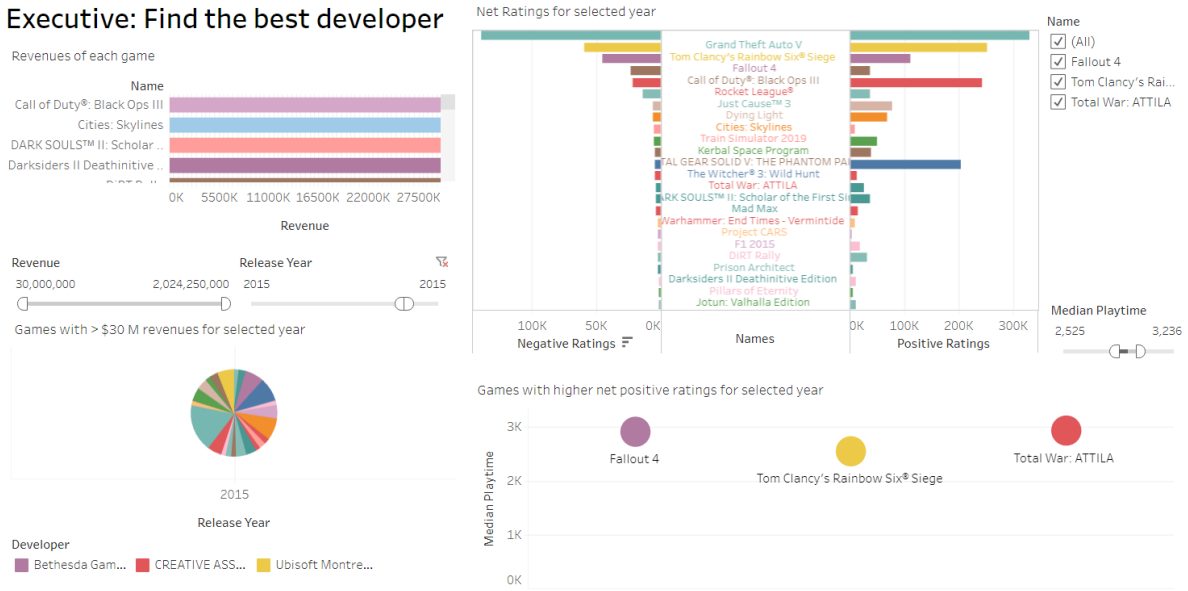
Again, considering that the STEAM executive prioritizes popularity based on several factors (revenue, rating, year, etc.) they would want to choose a game with the highest median playtime. A filter is provided in case they want to choose a range of playtimes.

Here, we can see that we have found the top three most popular games of 2014 (revenue > $30M) which have a median playtime of > 1031 hours. Based on this, the top three games are The Binding of Isaac: Rebirth (by Nicalis, Inc.) Dark Souls II (by From Software, Inc.) and Sid Meier's Civilization: Beyond Earth (by Firaxis Games). Based on this, the executive can choose to work with those specific developers.

Dashboard

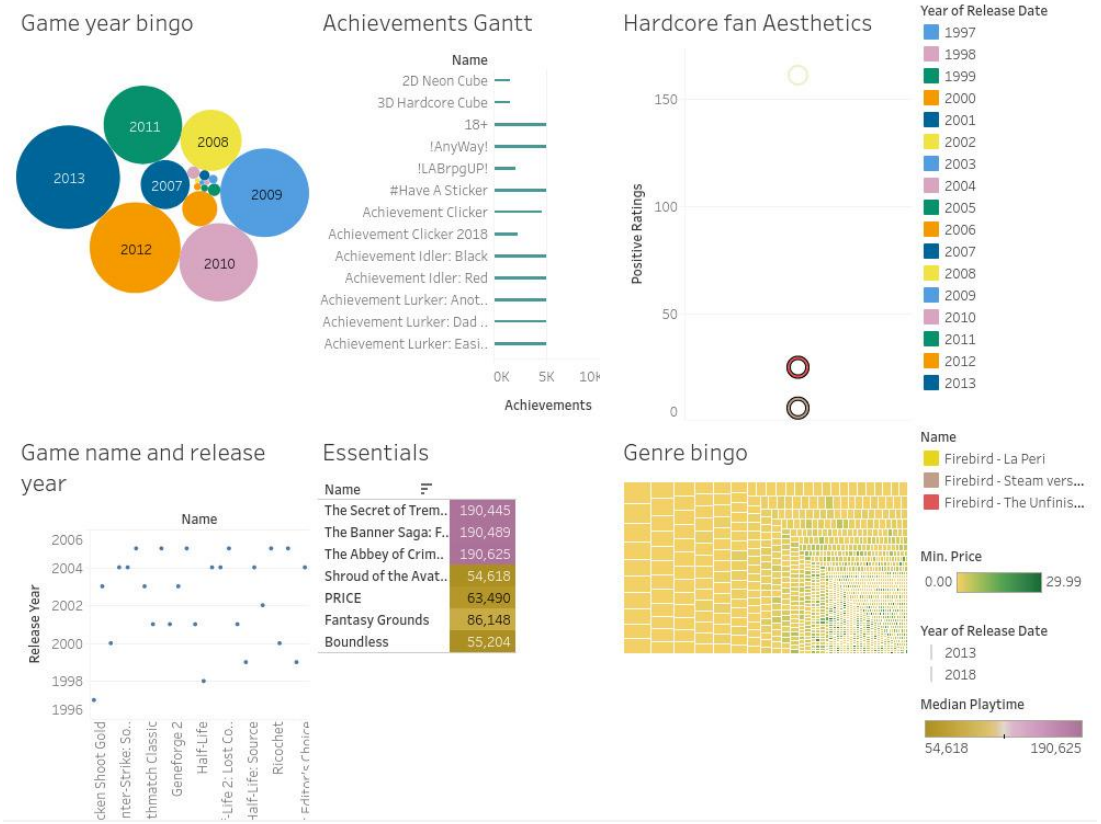


Above is a screenshot of the dashboard. The filters for revenue and release year enable the executive to choose a threshold for what they're looking for. We go from left to right, finally coming to the conclusion reached at the end of the previous section. As discussed above, the top three games in 2014 are The Binding of Isaac: Rebirth (by Nicalis, Inc.) Dark Souls II (by From Software, Inc.) and Sid Meier's Civilization: Beyond Earth (by Firaxis Games). Similarly, if we do the same for games released in 2015,

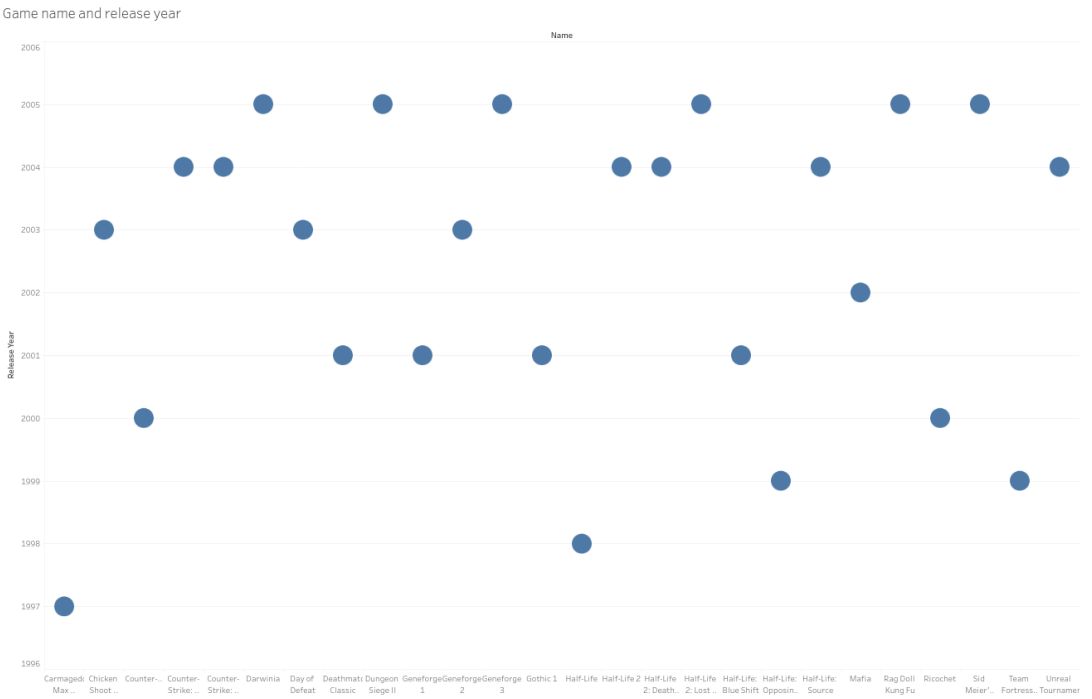


The top three games in 2015 are Total War: ATTILA (by CREATIVE ASSEMBLY), Fallout 4 (by Bethesda Game Studios) and Tom Clancy's Rainbow Six Siege (by Ubisoft Montreal). Based on this, the executive can choose to work with those specific developers.

Task 3: A game collector looking to filter according to their interests



Collectors' dashboard

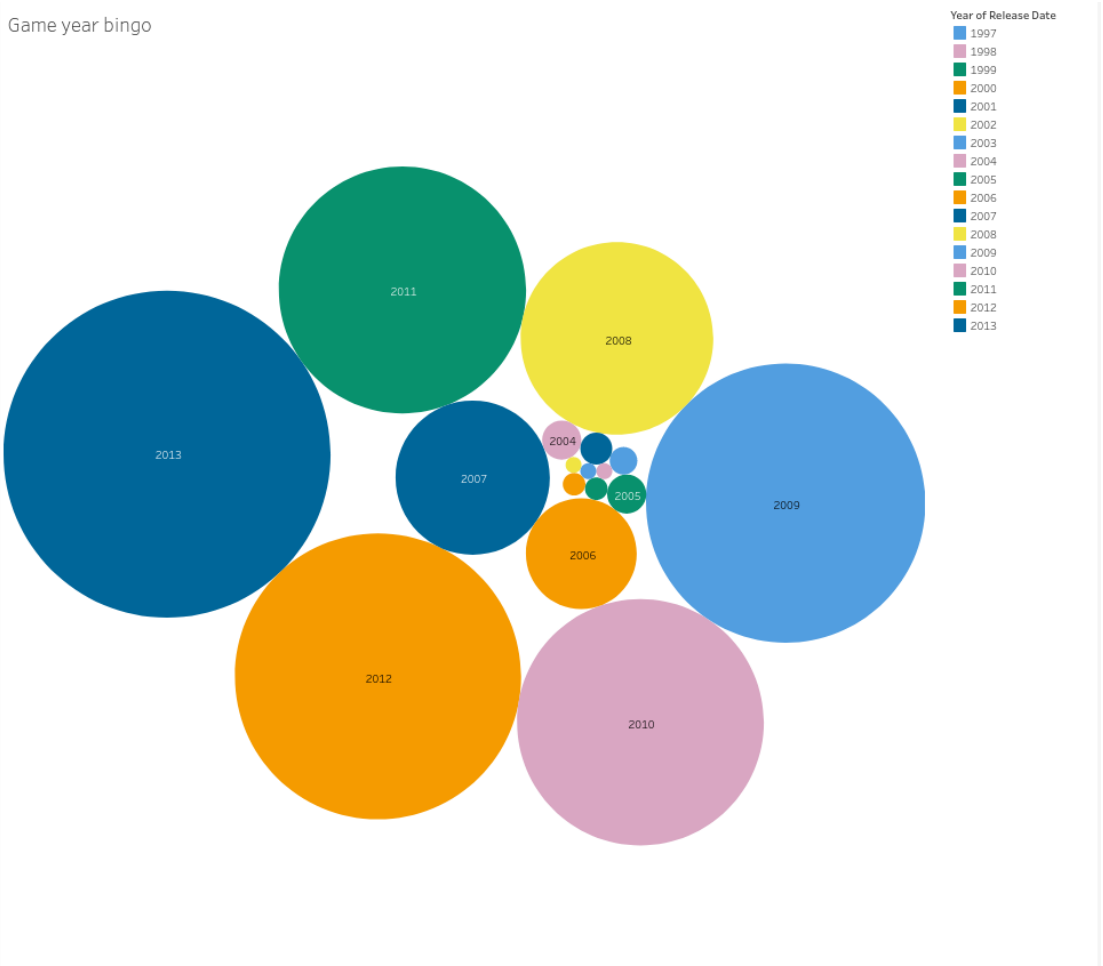


Finding similar games and product lines

a. A collector can find out games to buy by filtering the games based on their release dates, this will help them find out retro games.

b. Also, we have sorted the results accordingly, this will enable them to progress through product lines and franchises, enabling a complete playthrough.

c. This visualization can also be repurposed to find out the most recent games that have better gameplay and will aid in adding value to their collections.



Game year bingo

a. This visualization will aid the viewer in adding more niche games to their collections - we have showcased the number of games released in a calendar year through the area of the circles.

b. A collector can aim for games released in the year with the least number of games for a challenging yet rewarding game hunt.

c. They can also take the accessible route by looking at the years with the most number of games released, to easily start their collections.

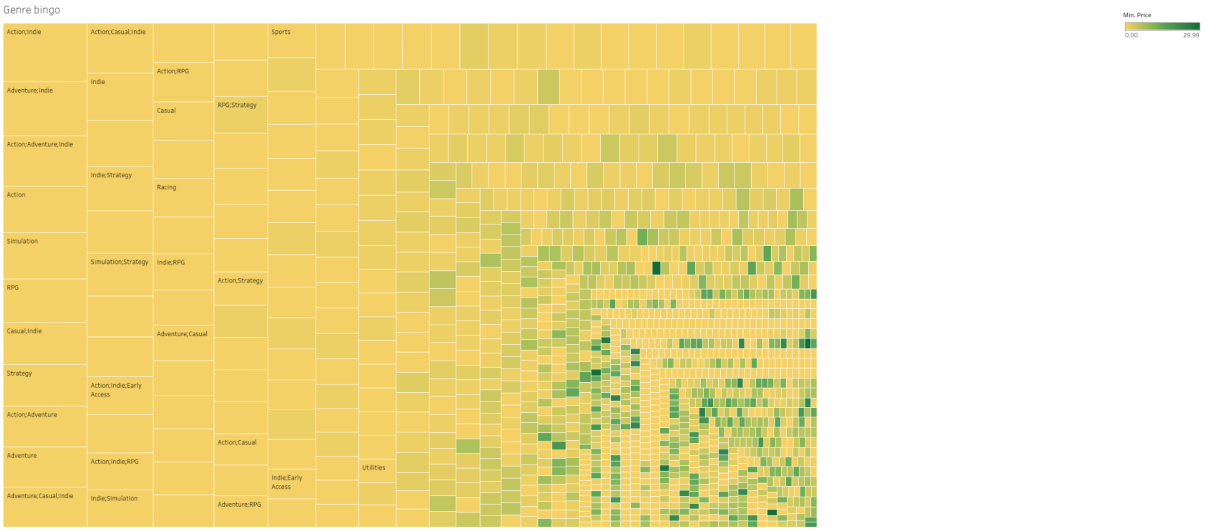
d. The year range can also be customized based on the year ranges which they wish to find games.

Hardcore fan aesthetics

a. This visualization is meant for experienced collectors who have a razor-sharp focus on buying new games, i.e, they have a clear idea on the approximate names of prospective new games.

b. By defining wildcards for string matching, the collector can look at all matching games and the number of their positive ratings.

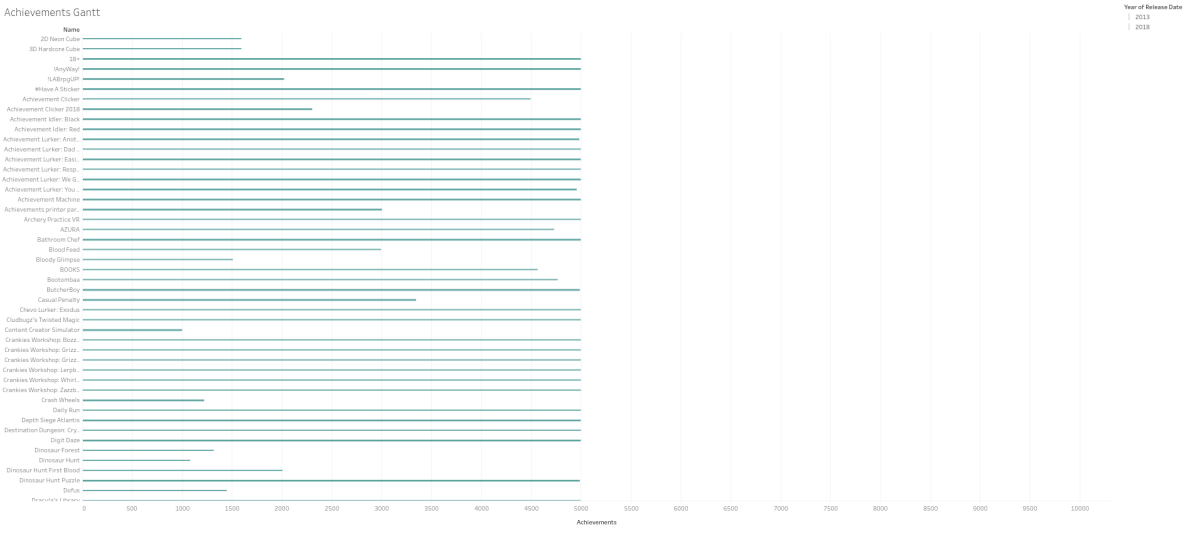
c. For fans of the game, we have also included a script that harnesses the power of machine learning and creates a custom color palette from the cover image (using KMeans clustering) of the given game, and use that as the color map for the visualization.



Budget-based genre finder

a. In this visualization, a collector can find the best genre to invest in, based on their defined budget.

b. The area of the rectangle refers to the number of the games in the corresponding genre, the collector can hence make an informed decision based on the relative availability and the cost of the games of the genres.



Achievements' Gantt Chart

a. This visualization is aimed at speedrunning collectors and achievements-focused gamers, where the width of the lines represent the number of achievements, and the darkness of the line represents the relative recency of the game.

b. They can also filter the results based on the number of achievements desired.

Dashboards

Please find the dashboards for tasks 1 and 3 mentioned below. The credentials for the access are:

- Emails: vijayjaisankar@iiitb.ac.in
- Password: Courier1@

Note: Task 2 is public and can be accessed without signing in.

The dashboards are as follows:

- Task 1: Developer Dashboard
- Task 2 : Executive Dashboard
- Task 3: Collector Dashboard