**Description of the Dataset Used**

**Introduction:**

Steam is a global application with massive user base that serves as the biggest platform for digital distribution of video games. Users can purchase games online and install on their machines seamlessly. This dataset contains an extensive collection of data points associated with games available on the platform and is chosen as the dataset provides both game-specific details and metadata, including aspects related to pricing, features, developer/publisher details, and user reviews.

**Dataset Details:**

This dataset titled “**Steam Games Dataset.xlsx**” is obtained on the 27th of October 2023 and the latest update to the dataset was on the 20th October 2023 as of the time of obtaining the dataset. This dataset encompasses the following attributes in sequence:

|  |  |
| --- | --- |
| **Attributes** | **Description of Attributes** |
| **Title** | The name of the game. |
| **Original Price** | Initial listing price of game upon release. |
| **Discounted Price** | Price after discounts. |
| **Release Date** | Game's release date on Steam. |
| **Link** | URL to the game's Steam webpage online. |
| **Game Description** | Description provided by developers or publishers, often summary of the game |
| **Recent Reviews Summary** | Categorical summary of user reviews in recent times. |
| **All Reviews Summary** | Categorical summary of all user reviews since release date. |
| **Recent Reviews Number** | Number of reviews in the past 30 days. |
| **All Reviews Number** | Total user reviews since release. |
| **Developer** | Game's developer name // Company creating the game |
| **Publisher** | Company// Entity publishing and distributing the game. |
| **Supported Languages** | Languages the game is available in. |
| **Popular Tags** | Keywords indicating genre or features of the game. |
| **Game Features** | Specific features available in the game, such as multiplayer capability or VR support. |
| **Minimum Requirements** | The operating system the game supports, and its minimum system requirements needed to run the game. |

**Purpose and Relevance:**This dataset provides rich and valuable insights into the gaming market on Steam. As the platform is the largest digital distribution of games with a huge significance in the PC gaming world, the dataset offers insights on trends, preferences, and patterns on both game developers and the gaming community. User reviews on Steam plays a big role on game sales and developers' reputations. Therefore, understanding these dynamics is vital for industry stakeholders like developers, marketers, and analysts.

**Initial Questions Description**

Given the dynamic nature and the ever evolving and revolutionising gaming industry, my primary aim is to discern patterns and gain insights that can guide stakeholders in making informed decisions. The initial set of questions I aim to address with this dataset are as follows:

|  |  |  |
| --- | --- | --- |
| No | Initial Set of Question | Initial Questions Description |
| **RQ1** | **Games Released per Year and Pricing Evolution** | Are more games released every year compared to the previous years and how game prices evolved over time? |
| **RQ2** | **Price and User Reviews** | How do user reviews correlate with game prices, both original and discounted and that do higher pricing lead to more critical reviews? |
| **RQ3** | **Developer and Publisher Analysis** | Which game developers or publishers have the most titles on Steam? What games’ primary genre do they focus on and how are their games generally received in terms of user reviews? |
| **RQ4** | **Game Features and Popularity** | Do game features and requirements relate to their popularity or rating? For instance, are games with specific features (e.g., multiplayer, VR support) more popular or better reviewed? |
| **RQ5** | **Genres and their Popularity** | Are there certain genres (derived from popular tags) that are more popular or receive better reviews on Steam platform? |

**\*\* Disclaimer: The discounted prices of games in the dataset reflect the discounts available as of 27th October 2023. It's worth noting that Steam frequently has sale events, and the game prices obtained might be influenced by any ongoing or recent sales. The specific sale events around this date are not detailed in the dataset, therefore this context should be considered when interpreting any insights related to pricing. This applies to when answering RQ1 and RQ2\*\***

Each of these questions is designed to dive into a specific aspect of the gaming industry, be it pricing strategies, the impact of reviews, or the importance of game features. These questions lay the groundwork for analysing and exploring this dataset. Once these initial questions are explained, there are more refined questions that will be raised and answered with the same format of explanation.

**-For each question, a description of visualization strategies used, including data cleaning, transformation, visual encoding, etc.**(Before critical discussion) \*\*5. Further refine/propose questions\*\*:

- \*\*RQ6\*\*: Supported Languages and Global Reach: How do games' supported languages affect their popularity or global reach? Do games supporting more languages have broader appeal or better reviews?

- \*\*RQ7\*\*: Indie vs. Major Developers: Given the increasing importance of indie games, how do indie developers' games perform in comparison to major developers in terms of pricing and reviews?  
  
-\*\* RQ8\*\*: System Requirements and Audience Reach: How do games' minimum system requirements relate to their popularity and user reviews? Is there a correlation suggesting that games with lower system requirements garner a wider audience or better reviews?

**- An explanation of the exploratory process of generating new questions and visualizations.**

RQ1: Dataset Limitations: Begin by acknowledging the limitations of the dataset. Mention that while the dataset provides a vast amount of information, it does not capture every game released on Steam, especially seminal releases like "Half-Life 2" in November 2004.

Steam's Inception: Highlight that Steam was launched in 2003. This can explain the lack of data for the years before 2003. If your dataset doesn't show releases for 2003 and 2004, mention this as a data gap.

Significance of "Half-Life 2": Point out that "Half-Life 2" was not just another game release. It played a crucial role in the evolution of Steam. Valve, the developer of "Half-Life 2" and the owner of Steam, required players to install Steam to activate their copy of the game. This move was initially met with backlash but eventually helped establish Steam as a leading platform for digital game distribution.

Exponential Growth: The plot, as you've presented, shows an exponential growth in the number of games released on Steam over the years. Discuss potential reasons for this growth, such as the increasing popularity of indie games, Steam Greenlight (and its successor, Steam Direct), and the ease of digital distribution compared to physical copies.

Quality vs. Quantity: While the number of games released on Steam has increased, it's worth discussing that not all these games were hits or even of high quality. Steam's open platform approach has led to concerns about the quality of games, especially with the advent of Steam Direct.

Data Source: If possible, provide information on where the dataset was sourced from. Was it scraped from the Steam store? Is it from a third-party source? Being transparent about the origins of the dataset can help your audience understand potential biases or gaps.

Call for Feedback: End by encouraging your audience to share if they notice any other significant omissions or anomalies in the data. This invites collaborative analysis and can help improve the accuracy and completeness of future datasets and analyses.

Remember, every dataset will have limitations, and it's crucial to be transparent about these when presenting your findings. This not only enhances your credibility but also provides a more holistic view of the data's narrative.

**Critical discussion of visualization design**

**Reflection on the development process**