**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. Once these initial questions are explained, there are more refined questions that will be raised and answered with the same format of explanation. 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 dataset is not fully reflected on the exact numbers of games released on steam. It is worth noting that games who are released for example 1990s when steam wasn’t available could be added into the steam library. 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\*\***

**-Description of visualization strategies used**

|  |  |
| --- | --- |
| No | **Visualization Strategies Used** |
| **RQ1** | **Data Loading:**  Used readxl library to load the steam game dataset from the Excel file.  **Data Cleaning:**  Removed the "UTC" format from the "Release Date" column due to having the months as string eg:( 13-Jul-23) and converting the column to the date format using the lubridate and gsub functions.  **Data Transformation:**  Filtered out games released before 2003 due to steam application is only rolled out on 12th September 2003.  **Data Aggregation:**  Grouped the dataset by year of release date to summarize the number of games released in each year. The data is joined with a reference frame (all\_years) to ensure the years from 2003 to 2023 are represented.  **Visual Encoding:**  Created a bar chart with years as the x-axis and the number of games released on that year as the y-axis with each bar represents the number of games released in that year.  **Theme and Aesthetics:**  Added a minimal theme with filled the bars represented in a "steelblue" colour for visual appeal.  **Interactivity:**  The static ggplot chart is converted into an interactive chart using the plotly library for more interactivity. |
| **RQ2** | How do user reviews correlate with game prices, both original and discounted and that do higher pricing lead to more critical reviews? |
| **RQ3** | 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** | 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** | Are there certain genres (derived from popular tags) that are more popular or receive better reviews on Steam platform? |

(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.**

|  |  |
| --- | --- |
| **No** | **Exploratory Process** |
| **RQ1** | **Initial Inspection:**  Upon initially inspection of the dataset, it was observed that the "Release Date" column had "UTC" appended to the dates, which would prevent accurate date conversions and calculations.  **Defining the Scope:**  Given that steam was released in 2003 and the dataset might contain games released before the year 2003 and developers can add games into Steam library regardless of release date, I decided to narrow down the scope and focus on games released on Steam after 2003.  **Visualization Choices:**  I chose a bar chart as it’s effective in representing the frequency of the number of games released per year as bar chart is clear in visualising trend over the years and allows for easy comparison between years.  **Iterative Process:** Initially I wanted to add the average price of each game and comparing the prices however this question is straighter to point and hence I chose number of games released per year. After the initial visualization was created, I made it more interactive for easier usability. |
| **RQ2** |  |
| **RQ3** |  |
| **RQ4** |  |
| **RQ5** |  |

**Critical discussion of visualization design**

**Reflection on the development process**