SteamVis: A Tool for Collecting and Analyzing Data of the Game Distribution Platform Steam

Alexander Kainz
Graz University of Technology
Graz, Austria

Johanna Pirker
Ludwig-Maximilians-Universität, Germany
& Graz University of Technology, Austria

[1]

***Abstract*—In recent years, the gaming industry has experienced rapid growth, further bolstered by the pandemic. Today, games are being utilized beyond entertainment, finding applications in education, health, training, and various industries. Understanding trends and analyzing games and their user base is therefore crucial for game developers, stakeholders, business analysts, and researchers alike. With over 60,000 games, Steam stands out as a major game distribution platform that collects valuable data about games and their users. To simplify the analysis of current trends in the gaming world, we have developed SteamVis: an analysis and visualization tool that efficiently crawls, displays, and analyzes game-related information from Steam. In this paper, we present the tool's details, conducted case studies, and a qualitative evaluation by domain experts. SteamVis facilitates the study of games and game-related information by providing powerful analytical and visualization tools, enabling the discovery of important trends and insights.**

***Index Terms*—visualization, analysis, tool**

I. Introduction

One of the industries experiencing the fastest growth is the video gaming market. According to a report by Juniper [1], the market was valued at approximately $159.89 billion in 2020 and is projected to reach $268 billion by 2025. Among the most prominent platforms for game distribution is Steam, a service provided by Valve Corporation [2]. With a library of over 50,000 games and a continuous influx of new releases each year [2], Steam's expansive nature is evident. The increasing number of peak concurrent users over time further highlights its popularity. In September 2021, a record was set with roughly 26.09 million peak concurrent users, compared to about 6.61 million in January 2013. The substantial user base and the constant release of new games on Steam generate a significant amount of data. Researchers, business analysts, and game developers can leverage this data for experimentation, analysis, and exploration. Steam's inclusion of user reviews is particularly valuable for gaining insights into player behavior. Steam data has been utilized in numerous studies, covering a wide range of topics such as sentiment analysis, investigation of review helpfulness [3] [4] [5], and recommendation systems [6]. While different techniques and processes have been employed to gather data from Steam, there is currently no standardized tool available for this purpose. Hence, the objective of this work is to develop a tool for stakeholders interested in games and game-related environments, including game researchers, developers, and analysts, to gather and investigate data from Steam. By providing predefined procedures, the tool aims to streamline the data collection process and save time and effort. The tool should allow users to gather general game information such as title, publisher, developers, genres, user-defined tags, and other relevant details. Additionally, it should facilitate the compilation of game reviews. Filters should be implemented to limit data collection to specific groups of video games, and users should have the option to gather reviews that meet predefined criteria for language and production time. Analytical capabilities are a crucial aspect of the tool, enabling researchers to delve into their topics of interest promptly. On one hand, the tool should provide a concise summary of games with minimal user input. On the other hand, it should offer advanced features that allow for the design of unique analysis tasks. Furthermore, the tool should support the visualization of data through tables or charts, with the ability to save the content of tables or charts as CSV files.

Based on these requirements, the following main objectives of this work can be summarized:

- Development of a tool for an effective analysis of game and player-related data from Steam
- Proof of concept of the tool by creating use cases
- User study of the tool with stakeholders

Details about this work can also be found in the following thesis: [7].

II. Background and Related Work

Steam, which was launched by Valve Corporation in 2003, stands as the largest platform for distributing video games. As of October 1, 2021 [8], it boasts a collection of over 50,000 games. Each game listed on Steam includes essential information such as its title, creators' and publishers' names, price, rating, related genres, user-defined tags, and a description. Furthermore, users have the ability to rate games as favorable or negative and indicate whether reviews are helpful

[1]Pre-print copy of: A. Kainz and J. Pirker, "SteamVis: A Tool for Collecting and Analyzing Data of the Game Distribution Platform Steam," 2023 IEEE Conference on Games (CoG), Boston, MA, USA, 2023, pp. 1-8, doi: 10.1109/CoG57401.2023.10333255.

[2]URL: https://www.valvesoftware.com/de/

or unhelpful. Researchers have leveraged data from Steam to investigate various aspects of game attributes and player behavior, employing diverse methodologies and approaches. In this section, we provide a concise overview of studies conducted on the Steam platform, with a particular emphasis on different methods of game analysis and the examination of user reviews' significance.

A. Game data analysis and visualization on Steam

In the literature, numerous examples of analysis using Steam data sets can be found. Here, we describe several noteworthy examples from various categories, each utilizing different data elements from Steam.

B. Review Helpfulness and Review Sentiment

Review helpfulness and sentiment analysis are two commonly researched areas using Steam review data. Regarding review helpfulness, Eberhard et al. [3] conducted a binary prediction experiment to explore the distinctions between helpful and unhelpful reviews. They considered various review data elements, including structure, formatting, readability, sentiment, and game-related information such as playtime and Early Access state. Similarly, Barbosa et al. [4] analyzed helpfulness using a Multilayer Perception Artificial Neural Network (MLP ANN) and extracted features such as author information (friends, reputation) and metadata (average game rating, review date). In another study by Kang et al. [5], two data mining approaches, namely the decision tree-based algorithm (CART) and MLP ANN, were compared for helpfulness prediction. In the realm of sentiment analysis, Ji et al. [9] applied Natural Language Processing (NLP) techniques to identify positive and negative Steam reviews, employing methods like naive Bayes and Support Vector Machines (SVMs). They also incorporated meta-features such as review helpfulness, funny votes, and the number of hours played by the reviewer. Bais et al. [10], on the other hand, compared five different approaches for sentiment analysis. These examples highlight the diverse methodologies and data elements utilized in analyzing Steam data, shedding light on the valuable insights that can be extracted from the platform.

C. Genres and Tags

Genres on Steam are predetermined by the platform, while users have the ability to add tags to games. In a study by Windleharth et al. [11], user-generated tags on Steam were analyzed. The authors discovered a total of 294 distinct tags in March 2015 and categorized them into 29 main categories, providing valuable insights into the tagging behavior of Steam users. Another significant area of analysis revolves around "Early Access" games, which are titles that are not yet fully developed but are made available for players to experience and review. In an empirical study by Lin et al. [12], the authors aimed to characterize Early Access games, relying heavily on user review data to gain insights into this unique category of games. These examples demonstrate the diverse dimensions of analysis that can be pursued using Steam data, encompassing genres, user-generated tags, and the evaluation of Early Access games through user reviews.

D. Recommendation Systems

One popular application of game data analysis is the development of recommendation systems, which aim to assist players in discovering their next favorite game or aid distributors in effectively promoting products. Wang et al. [6] propose a recommendation system called STEAMer, which utilizes user data combined with a deep autoencoder. The deep autoencoder is trained using various game and user-related features, including genre, rating, developer ID, publisher ID, playtime, and friends. In another study by Saaidin et al. [13], three different recommendation systems were implemented and compared. The first system is based on genres and playing history, while the second system employs game descriptions instead of genres. The third system takes a hybrid approach, combining multiple factors. Notably, in their analysis, the authors also incorporate geographic features. These examples highlight the importance of recommendation systems in leveraging game data to enhance user experiences and support business strategies. The incorporation of various data elements and methodologies underscores the versatility of analysis approaches in the domain of game recommendation systems.

E. Popularity

An essential objective of game data analysis is to examine the popularity of games, genres, and other game metrics across different user groups. In a study by Toy et al. [14], the authors explored the geographical distribution of users, genres, and genre popularity using heat maps. Their analysis provided insights into the popularity of genres across different countries, revealing geographic preferences and trends. Similarly, Clement [15] conducted a similar analysis, focusing on the popularity of genres among different user groups. These studies highlight the significance of analyzing game data to understand the popularity of games and genres, and how these preferences may vary across different regions or user demographics. The use of heat maps and geographical analysis offers valuable visualizations for gaining insights into these patterns.

F. Playability

Steam data is frequently utilized to gain a deeper understanding of game quality. In a study by Li et al. [16], the authors present an approach that leverages data from a vast collection of game reviews, including user opinions. They classify the reviews into different perspectives related to playability, such as gameplay, functionality, and usability. This methodology allows for a comprehensive analysis of game quality from multiple angles, providing valuable insights into various aspects of the player experience.

G. Summary

In summary, there is great potential for conducting studies on Steam, given the abundance of available features and data related to games, users, and reviews. Game-related data, including genre, tags, rating, and description, along with user-related data such as friends, popularity, and geographical information, have proven to be valuable assets in gaining a better understanding of the gaming landscape. To gather data from Steam, analysis of related work has identified five main data sources, each providing distinct types of data: the Steam website, the Steam API, Steam Spy, Steam DB, and the database of O'Neill. These sources offer different sets of information that can be leveraged for analysis purposes. Overall, the field of data analysis in gaming, particularly in relation to Steam, continues to be an increasingly intriguing area of research. However, currently, no tools have been presented that effectively facilitate the process from data collection to data visualization.

III. Tool

A. System Overview

We propose the development of a comprehensive tool to streamline data analysis from Steam. This tool simplifies the workflow by offering predefined analytical jobs in the form of charts, providing an immediate overview of the data. Users can also perform custom SQL queries and integrate the results into charts or tables. The tool surpasses the limitations of SQL by allowing complex analysis tasks like sentiment analysis and natural language processing. For more advanced analysis, users can copy the dataset for use with other tools or save query results as CSV files or images. The tool targets game researchers and developers, with basic analysis easily performed using predefined jobs. Advanced analysis requires coding skills with Python and SQL queries. Our tool is developed using Python, SQLite for data storage, and PyQt5 for the user interface. Figure 1 shows an overview of the information collection process.

To streamline data collection, we adopt a two-step process that separates general game data from reviews, which are often large datasets. Reviews are collected using the Steam Web API method, retrieving them based on game IDs. Our tool offers predefined analysis tasks for both general game data and optional review data stored in the local database. Users can also create custom analysis tasks as needed. Figure 2 illustrates the process of running predefined analysis tasks on the collected data, enabling users to derive valuable insights. By separating data collection and providing various analysis tasks, our tool facilitates efficient data analysis and empowers users to effectively explore the collected data.

The user interface of our tool incorporates a combo box that allows users to select the specific project for which they want to analyze the data. When a project is changed or selected, the tool dynamically retrieves information from the local database. In this case, checkboxes are generated to display genres, tags, and game information that are assigned to at least one of the

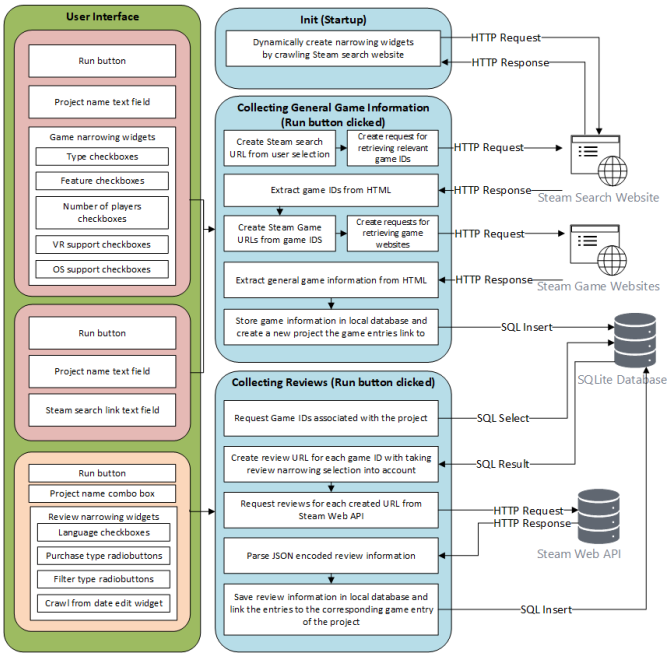

Fig. 1. High-level overview of the information collection process

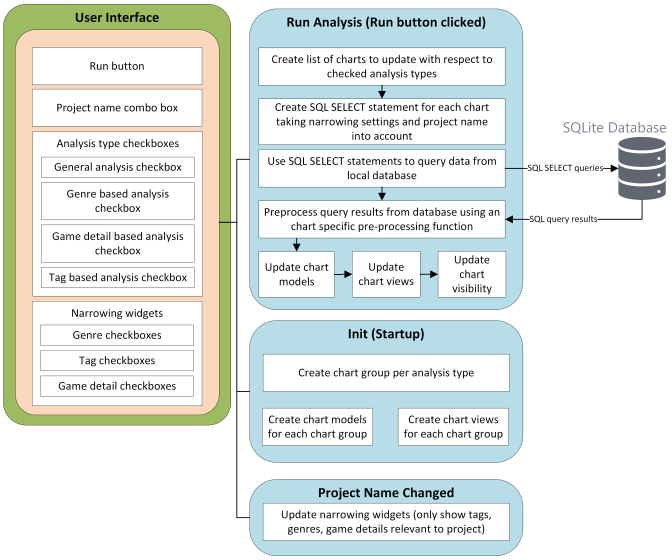

Fig. 2. High-level overview of the general game data analysis process.

games within the selected project. This feature provides users with a convenient way to filter and focus on the relevant data for their specific project of interest.

B. Data Collection

Figure 3 shows an overview of a Steam website, highlighting extracted information.

In our tool, the data collection process is divided into two steps: collecting general game data and collecting reviews. The initial step involves gathering game IDs and associated game data. Figure 4 displays the user interface view designed for collecting general game data. Users can input parameters and

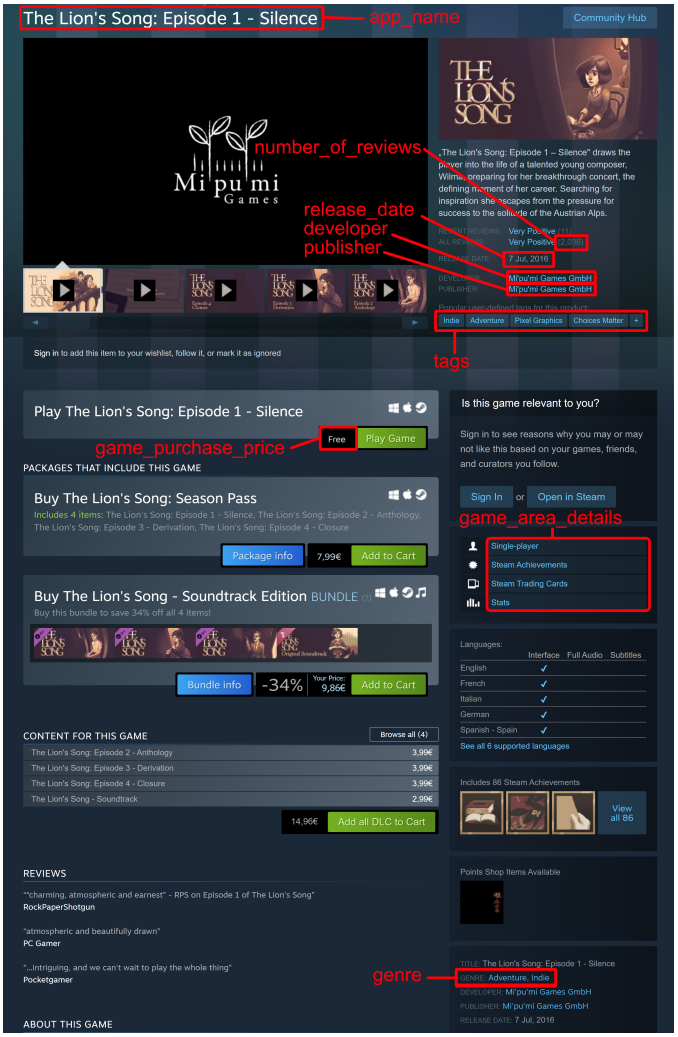

Fig. 3. Steam game website screenshot of the game "The Lion's song: Episode 1 - Silence" of Mi'pu'mi Games GmbH with highlighted extracted information.

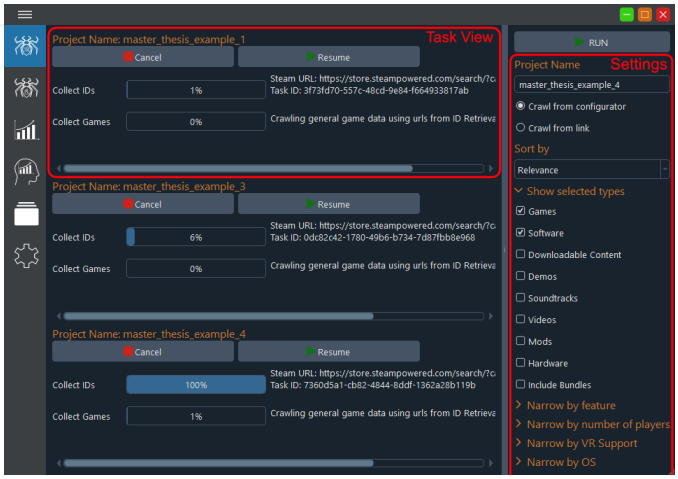

Fig. 4. View for collecting general game data.

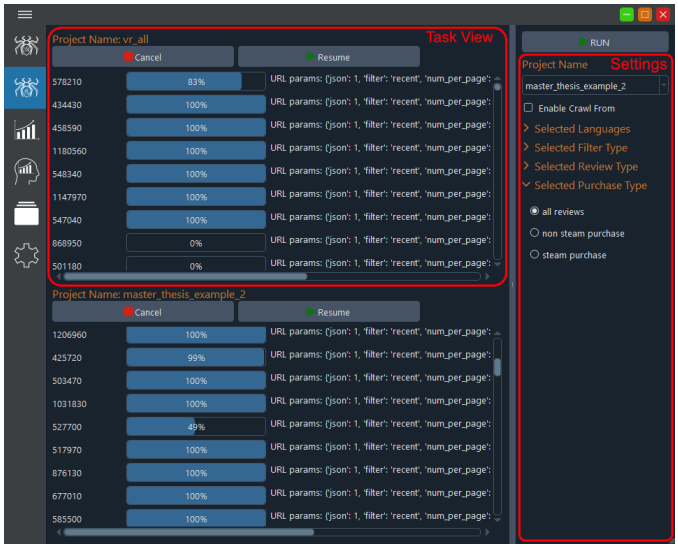

Fig. 5. View for collecting reviews.

initiate the data collection process for their selected project. By separating general game data collection from reviews, our tool enhances data management and analysis efficiency for users.

Once the game IDs and related game data have been collected for a project, users can proceed to gather the corresponding game reviews. Figure 5 illustrates the user interface feature dedicated to this functionality. To collect the reviews, our tool leverages a public method of the Steam Web API[3]. This method allows us to retrieve the reviews through HTTP GET requests, facilitating the process of collecting review data efficiently. By utilizing the Steam Web API, users can easily retrieve and incorporate the valuable insights provided by game reviews into their analysis.

The user can perform data analysis using the game data and (optionally) reviews that are stored in the SQLite database.

C. Data Visualization

Each analytic task within our tool is associated with a chart or table, as previously mentioned. To ensure separation between the presentation and logic, a model and view are created for each analysis assignment. The model is updated upon completion of a task, triggering a view update. Presently, our tool supports line charts, overlapping histograms, grouped bar charts, and bar charts as available chart types. Data retrieval is performed through SQL queries. An advanced feature of our tool is the ability to modify the SQL query results using Python scripts, offering enhanced flexibility and customization. For generating the charts, we utilize the Matplotlib plotting library[4]. Matplotlib provides a wide range of highly customizable plots, allowing users to create visually appealing and informative visualizations. These charts seam-

[3]https://partner.steamgames.com/doc/store/getreviews

[4]https://matplotlib.org/

Fig. 6. View for general data analysis templates.

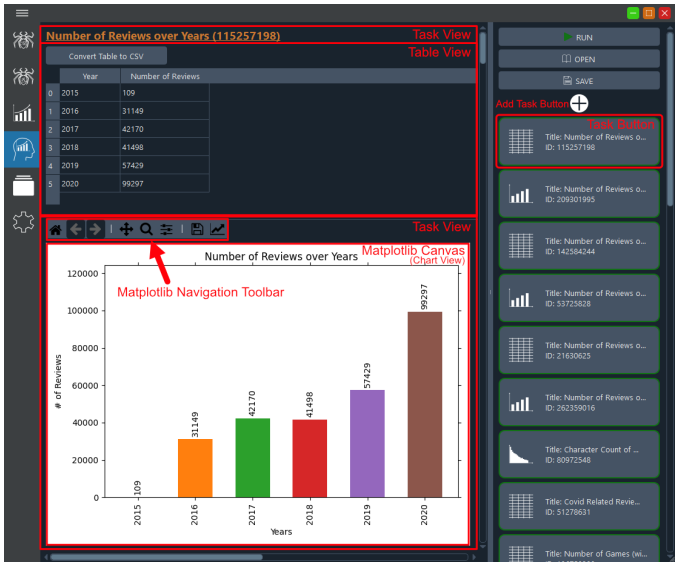

Fig. 7. View for custom data analysis.

lessly integrate into the user interface created with PyQt5, ensuring a smooth and intuitive user experience.

D. Data Analysis

To facilitate the analysis process, the tool offers support for general basic analysis templates, which automatically generate predefined charts based on the selected project. The third item in the menu sidebar corresponds to the general analysis view (refer to Figure 6). Within this view, a settings frame or right sidebar includes a combo box that allows users to choose the project for which they want to perform the general analytic activities. For time-based tasks, the output of the charts can be limited to a specified time period using two line-edit widgets located below the project name combo box. As mentioned earlier, each chart corresponds to a specific task. The tool offers four main areas of analysis, including general analysis, genre-based analysis, tag-based analysis, and game-detail analysis. These areas provide users with focused analytical capabilities to explore different aspects of the collected data.

An additional feature of our tool allows users to generate custom tasks via SQL Select statements. The results can be visualized via charts or tables. Figure 7 illustrates this capability.

IV. Case Studies

In this chapter, we are utilizing the tool in an ongoing study, building upon previous findings [17], [18], to investigate the impact of the Covid-19 pandemic on users' preferences for playing Virtual Reality games available on Steam. Additionally, we aim to evaluate the influence of the pandemic on written reviews and uncover any connections to the virus. To exemplify the potential of our technology, we delve into one of the research questions posed in the study: "Do users write more VR game reviews during the pandemic?"

To address this question, we collected game information and reviews for Virtual Reality games accessible on Steam. A summary of the dataset is presented in the following table.

Two distinct analytical tasks were conducted to examine the temporal distribution of written reviews. The first task

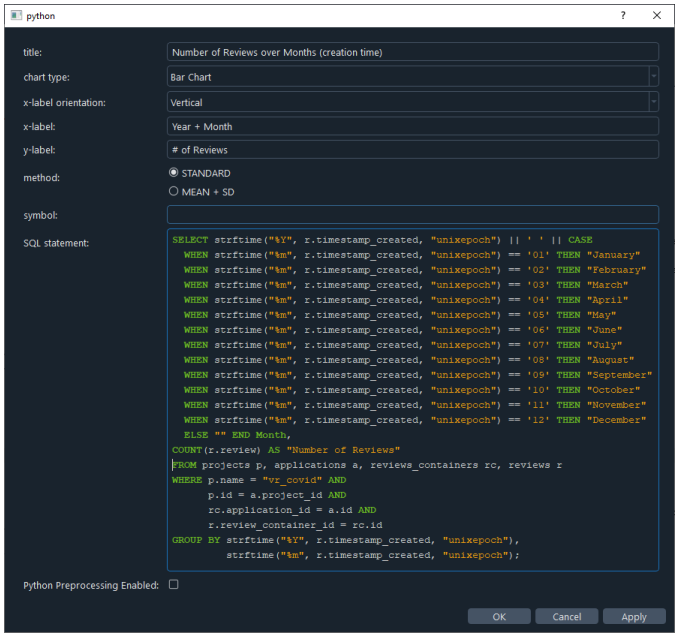

Fig. 8. Custom analysis dialog for visualizing the number of written reviews over months.

visualized the number of reviews across different years, while the second task focused on monthly variations. For both tasks, a bar chart was chosen as the appropriate chart type. In the study dialog, Figure 9 displays the corresponding custom analysis dialog containing the SQL Select statement for the monthly-based analysis. As depicted, the SQL Select statement includes a conversion from the month value to the month name. Alternatively, achieving the same outcome is also possible by utilizing Python to manipulate the SQL result.

Did the users write more reviews during the pandemic?

TABLE I

DESCRIPTION OF GENERAL ANALYSIS CHARTS. THE TYPE COLUMN DEFINES IF THE CHART IS A BAR CHART (B), A GROUPED BAR CHART (GB), A LINE CHART (L), OR A HISTOGRAM (H). THE CATEGORY COLUMN DEFINES IF THE CHART IS PART OF THE *General-Analysis* (GA), *Genre-Based Analysis* (GBA), *Tag-Based Analysis* (TBA) OR *Game-Detail-Based Analysis* (GDBA)

Title	**Type**	**Cat.**
Distribution of Steam Ratings	B	GA
Character Count Comparison of Positive and Negative Reviews	H	GA
Number of Releases per Year	B	GA
Number of Releases per Month	B	GA
Mean Ratio over Months	B	GA
Mean Ratio over Years	B	GA
Mean Ratio over Years	L	GA
Average Character Count over Years - All Reviews	B	GA
Average Character Count over Years - Positive Reviews	B	GA
Average Character Count over Years - Negative Reviews	B	GA
Average Character Count over Months - All Reviews	B	GA
Average Character Count over Months - Positive Reviews	B	GA
Average Character Count over Months - Negative Reviews	B	GA
Number of Games per Genre	B	GBA
Mean Ratio of Games per Genre	B	GBA
Number of Releases (Games) per Genre and Year	GB	GBA
Number of Games per Genre with Respect to Tags	GB	GBA
Number of Games per Genre with Respect to Game Area Details	GB	GBA
Number of Games per Tag	B	TBA
Mean Ratio of Games per Tag	B	TBA
Number of Releases (Games) per Tag and Year	GB	TBA
Number of Games per Tag with Respect to Genres	GB	TBA
Number of Games per Tag with Respect to Game Area Details	GB	TBA
Number of Games per Game Area Detail	B	GDBA
Mean Ratio of Games per Game Area Detail	B	GDBA
Number of Releases (Games) per Game Area Detail and Year	GB	GDBA
Number of Games per Game Area Detail with Respect to Genres	GB	GDBA
Number of Games per Game Area Detail with Respect to Tags	GB	GDBA

TABLE II

DATASET DESCRIPTION.

# of games	4241
# of reviews	386,705
review language(s)	English
review type	all (positive and negative)
purchase type	all (free and non-free games)
retrieval date	1st June 2021
Steam search URL	https://store.steampowered.com/search/?sort_by=Name_ASC&category1=998\%2C994&vrsupport=401

To gain a better understanding of gaming consumption during the Covid-19 pandemic, we utilized our tool to analyze the number of written reviews over time. As depicted in Figure 10, there was a significant increase of approximately 94% in the number of reviews written in 2020 compared to 2019. It is worth noting that the coronavirus outbreak primarily occurred in 2020, which explains the focus on that specific year. The increase from 2019 to 2020 is substantially higher compared to the increases observed in the preceding years (2016-2017, 2017-2018, and 2018-2019). When comparing the numbers, the month of March 2020 stands out. Further investigation into the significance of this month reveals that a considerable number of countries worldwide implemented nationwide lockdowns starting in March 2020.

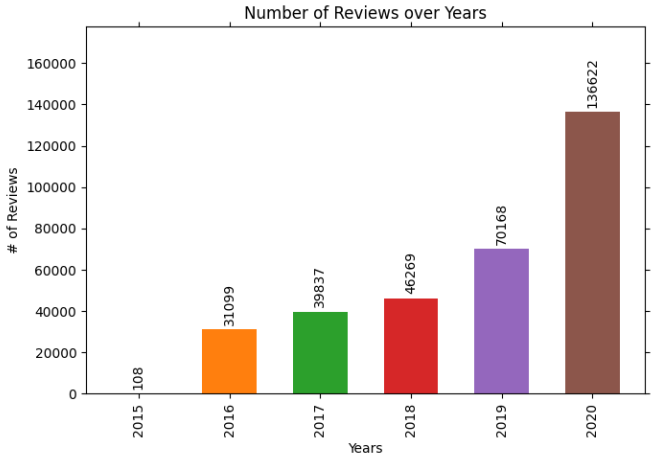

Fig. 9. Number of written reviews over years.

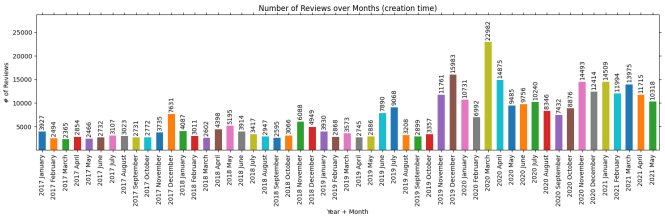

Fig. 10. Number of written reviews over months.

V. User Study

We carried out a qualitative study involving six domain experts to collect their opinions on the tool we developed and make necessary improvements. The study specifically targeted individuals with a background in Computer Science, as the tool is designed for game researchers who possess programming and SQL skills. This decision was made on the basis that individuals lacking these skills would require significant preliminary training to effectively utilize the tool's functionalities.

Our objectives were as follows:

- Assess the usability of the tool.
- Assess the emotions participants have during the usage of the tool.
- Detect and rate the importance of missing tool features.
- Detect errors and bugs.

A. Method

The study was conducted online. Prior to the study, each participant received a unique identifying number and a PDF file containing instructions and clear tasks. Participants were given a one-week window to complete the study at their convenience. The first task for participants was to download a zip file from Google Drive, which included the tool and other necessary materials. Throughout the study, participants were required to complete three questionnaires. These questionnaires included: (1) a pre-questionnaire gathering demographic data and previous experiences, (2) a tasks questionnaire, and (3) a post-questionnaire. The post-questionnaire included sections for overall impressions, features and improvements, as well as standard questionnaires such as CES and SUS.

B. Participants

The study focused on participants with prior knowledge of computer science to also evaluate the enhanced features of the tool. Consequently, all participants rated their understanding of computers very highly. Each participant held either a bachelor's or master's degree in computer science. The study involved a total of six participants, comprising one female and five males. All participants had previous experience purchasing games on Steam, and the majority of them had also posted reviews on the platform. Furthermore, half of the participants had expertise in gathering information from Steam, and one participant mentioned prior usage of the Steam Web API. However, only two out of the three individuals had prior experience in data analysis.

C. Results

The tool garnered positive feedback, achieving a mean rating of 4.00 (SD = 0.63) on a 5-point Likert scale, where the lowest rating is 3 and the highest rating is 5. Three out of the six participants specifically commended the engaging graphical user interface design, while two participants appreciated the tool's flexibility. Additional positive comments highlighted the tool's user-friendly nature and its ability to perform background activities. However, participants noted that there is a learning curve associated with becoming familiar with the tool, and some suggested improvements, highlighted missing features, or pointed out inconsistencies, which will be discussed in the following section.

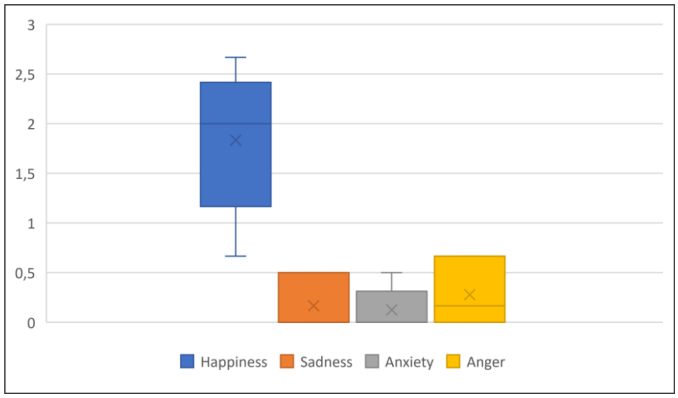

Fig. 11. Visualization of the CES results.

The following table summarizes feature ratings.

Participants were requested to rate the importance of seven features that are planned for implementation in the upcoming version of the tool, using a 5-point Likert scale ranging from 1 (not at all significant) to 5 (very significant). The table displays the mean scores and standard deviations for each feature. The most highly desired feature was the ability to save all charts simultaneously, followed by in-tool documentation.

1) Emotion Scale: Happiness (M = 1.88, SD = 0.72) was reported as the longest-lasting emotion by participants, followed by anger (M = 0.28, SD = 0.33), sorrow (M = 0.17, SD = 0.26), and anxiety (M = 0.125, SD = 0.21). Figure 11 illustrates these findings using a boxplot. Upon examining the three emotions comprising the happy group - satisfaction, excitement, and curiosity - we found that satisfaction (M = 2.17, SD = 0.75) received the highest rating, followed by excitement (M = 1.5, SD = 0.83), and then curiosity (M = 1.83, SD = 1.17). None of the participants in the "anger" group experienced anger (M = 0, SD = 0), but they occasionally felt irritability (M = 0.5, SD = 0.55) or frustration (M = 0.33, SD = 0.52). Throughout the entire investigation, none of the volunteers experienced anger, nervousness, anxiety, or helplessness.

2) Usability: The System Usability Scale (SUS) obtained a mean score of 77.08 (SD = 12.98). According to the adjective scale, the instrument's usability falls between good and exceptional. The lowest individual SUS score recorded was 65, indicating usability that is considered OK to good. In contrast, the highest individual SUS score obtained was 100, representing the best possible usability on the adjective scale.

3) Task Success Rate: The tasks questionnaire was divided into three sets of questions, each focusing on different key features of the tool: Analysis, Custom Analysis, and Data Collection. Participants found the tasks to be appropriately challenging, and all of them answered the questions correctly in the Analysis and Data Collection question sets. In the Cus-

TABLE III

RATING RESULTS BASED ON THE USERS' OPINIONS ON THE USEFULNESS OF NON-IMPLEMENTED FEATURES. THE CORRESPONDING QUESTION ASKED THE USERS IF THEY WANTED THE SEE THE FEATURE IN A LATER RELEASE OF THE TOOL. FOR THE RATING, A 5-POINT LIKERT SCALE BETWEEN (1) *not at all* AND (5) *very much* WAS USED.

Feature Description	**Mean**	**SD**
In-tool documentation	5.00	0.00
Possibility to save all charts and tables at once	4.50	0.55
Possibility to create custom chart designs inside the tool (e.g., contour chart)	3.83	1.47
Option to select which data should be collected from Steam (to save memory)	3.67	1.37
Possibility to change the visibility order of charts/tasks	3.33	1.03
Tabs to structure custom analysis tasks	3.17	1.17
Text-only output (result is converted to a string and visualized in a text field)	2.33	1.21

.

tom Analysis group, only one participant provided incorrect answers to two questions.

VI. CONCLUSION

The video game industry has witnessed significant growth in recent years, facilitated by platforms like Steam that offer online game purchasing and downloading. Steam allows users to rate games and provide optional written descriptions, leading to increased interest in video game distribution and review platforms among game academics and developers. While various methods have been employed to utilize Steam data for different objectives, there is considerable overlap in these approaches. This research introduces a tool that streamlines the process of gathering and analyzing Steam data. It presents a design and conceptual framework outlining the tool's objectives, target audience, specifications, and logical architecture. The implementation enables users to collect game reviews and general information, while also offering data filtering capabilities based on specific features. The tool incorporates analytical tools that provide users with a comprehensive overview of datasets and the flexibility to develop customized analyses using Python or SQL Select queries. A study was conducted to demonstrate the tool's utility, and it successfully met all requirements. Additionally, a small user study was carried out to obtain feedback on the tool's usefulness and user experience. The findings revealed good to exceptional usability, with participants expressing overall satisfaction. The tool received positive reception, with participants expressing their willingness to use it when gathering and analyzing data from the Steam game distribution system.

In conclusion, understanding the dynamics and contexts of the video game industry is crucial for game producers and researchers. Our solution offers valuable insights into games and player behavior, representing a step in the right direction.

REFERENCES

[1] Juniper Research. (2021, 5) Global video game market value from 2020 to 2025 (in billion u.s. dollars) [graph]. [Online]. Available: https://www.statista.com/statistics/292056/video-game-market-value-worldwide/

[2] Steam Spy. (2021, 6) Number of games released on steam worldwide from 2004 to 2021 [graph]. [Online]. Available: https://www.statista.com/statistics/552623/number-games-released-steam/

[3] L. Eberhard, P. Kasper, P. Koncar, and C. Gütl, "Investigating helpfulness of video game reviews on the steam platform," in *2018 Fifth International Conference on Social Networks Analysis, Management and Security (SNAMS)*. IEEE, 2018, pp. 43–50.

[4] J. L. Barbosa, R. S. Moura, and R. L. d. S. Santos, "Predicting portuguese steam review helpfulness using artificial neural networks," in *Proceedings of the 22nd Brazilian Symposium on Multimedia and the Web*, 2016, pp. 287–293.

[5] H.-N. Kang, H.-R. Yong, and H.-S. Hwang, "A study of analyzing on online game reviews using a data mining approach: Steam community data," *International Journal of Innovation, Management and Technology*, vol. 8, no. 2, p. 90, 2017.

[6] D. Wang, M. Moh, and T.-S. Moh, "Using deep learning and steam user data for better video game recommendations," in *Proceedings of the 2020 ACM Southeast Conference*, 2020, pp. 154–159.

[7] A. Kainz, "Steamvis: a tool for collecting and analyzing data of the game distribution platform steam," 2022.

[8] J. Clement. (2021, 6) Number of games released on steam worldwide from 2004 to 2021. [Online]. Available: https://www.statista.com/statistics/552623/number-games-released-steam/

[9] F. Ji, "Sentiment analysis and opinion extraction of game reviews on steam," 2019.

[10] R. Bais, P. Odek, and S. Ou, "Sentiment classification on steam reviews," 2017.

[11] T. W. Windleharth, J. Jett, M. Schmalz, and J. H. Lee, "Full steam ahead: A conceptual analysis of user-supplied tags on steam," *Cataloging & Classification Quarterly*, vol. 54, no. 7, pp. 418–441, 2016.

[12] D. Lin, C.-P. Bezemer, and A. E. Hassan, "An empirical study of early access games on the steam platform," *Empirical Software Engineering*, vol. 23, no. 2, pp. 771–799, 2018.

[13] A. S. B. Kamal, S. Saaidin, and M. Kassim, "Recommender System: Rating predictions of Steam Games Based on Genre and Topic Modelling," *2020 IEEE International Conference on Automatic Control and Intelligent Systems, I2CACIS 2020 - Proceedings*, no. June, pp. 212–218, 2020.

[14] E. J. Toy, J. V. Kummaragunta, and J. S. Yoo, "Large-scale cross-country analysis of steam popularity," in *2018 International Conference on Computational Science and Computational Intelligence (CSCI)*. IEEE, 2018, pp. 1054–1058.

[15] J. Clement. (2018, 4) Distribution of steam users worldwide as of april 2018, by country. [Online]. Available: https://www.statista.com/statistics/826870/steam-distribution-country/

[16] X. Li, Z. Zhang, and K. Stefanidis, "A data-driven approach for video game playability analysis based on players' reviews. information 2021, 12, 129," 2021.

[17] S. Petrosino, E. Loria, A. Kainz, and J. Pirker, "The panorama of steam multiplayer games (2018-2020): A player reviews analysis," in *Proceedings of the 17th International Conference on the Foundations of Digital Games*, 2022, pp. 1–7.

[18] J. Pirker, E. Loria, A. Kainz, J. Kopf, and A. Dengel, "Virtual reality and education–the steam panorama," in *Proceedings of the 17th International Conference on the Foundations of Digital Games*, 2022, pp. 1–11.