A Framework Proposal for Developing Historical Video Games Based on Player Review Data Mining to Support Historic Preservation

Sarvin Eshaghi¹, Sepehr Vaez Afshar¹, Mahyar Hadighi¹
¹Texas Tech University, Lubbock, TX

ABSTRACT: Historic preservation, which is a vital act for conveying people's understanding of the past, such as events, ideas, and places to the future, allows people to preserve history for future generations. Additionally, since the historic properties are currently concentrated in urban areas, an urban-oriented approach will contribute to the issue. Hence, public awareness is a key factor that paves the way for this conservation. Public history, a history with a public audience and special methods of representation, can serve society in this regard. Historical video games, which get their game objectives from the past, can be considered a form of public history, raising people's awareness and letting them gain historical experience. While plenty of studies focused on this content with various approaches, this study takes the topic under consideration with a player-oriented focus. The research uses the Steam video game digital distribution service database to explore the existing historical games based on their genre, player scale, dimension, VR compatibility attributes, and player rating. The database includes 1,148 games published between 2002 and 2022, filtered as historical by Steam tags and sorted by player review. For the data collection, the study uses the Octoparse web scraping tool for data extraction, and for the data analysis, it uses RapidMiner Studio data mining software. The study aims to provide a framework applicable to the development of commercially playable historical video games contributing to heritage preservation by raising public awareness in the context of public history.

KEYWORDS: Historic Preservation, Public Awareness, Video Games, Data Scraping, Data Mining, Virtual Reality

INTRODUCTION

American urban neighborhoods and communities were in danger of destruction in the 1950s and 1960s due to the damaging effects of freeway construction and urban renewal. Hence, the town planners were invited to be more sensitive, maintain, and enhance the neighborhoods' physical fabric. In 1965, the so-called concerns led to the promotion of historic districts in the United States (Hamer 2000). Currently, the National Park Service, the US agency managing national parks, monuments, and other natural, historical, and recreational properties, defines historic preservation as humans' communication with their past regarding the future. Through historic preservation, people can figure out the important parts of history and the factors which could be preserved for the future. It is a prominent way to convey people's understanding of the past to the next generations. These understandings may include the people, events, and ideas, as well as places. The professions embracing the historic preservation field are as varied as architects, landscape architects, historians, archeologists, curators, and many other experts (NPS 2022). Hence, the historic preservation field expanded its spectrum of expertise towards being an urban occupation since the beginning of the current century through its intersection with urban revitalization.

Given the high concentration of historic buildings and neighborhoods in urban areas, preserving historic properties and districts requires an urban-oriented approach to inclusion and economic development (Ryberg-Webster & Kinahan 2014). According to De Guichen and d'Iteren (2009), the survival of this heritage relies to a great extent on public awareness and enlightened decision-making on the part of authorities. Hence, in this context, public history—differentiated from traditional academic history in terms of target audience and locations of representation—raises awareness through the urban area itself instead of via written or oral academic discourses (Eshaghi et al. 2021; Plummer 2001). Due to Howe and Kemp, public history is multidisciplinary in its nature and requires well-defined objectives. Multiple, ever-changing, and diverse audiences accounted for its venues. These audiences may have various socio-economic classes, cultural settings, and educational levels. The public historians' task is to advocate expansive public history knowledge and awareness. However, the public history contents, including visuals, texts, and oral data, rarely demonstrate the comprehensive research behind it (Plummer 2001). Awareness of the importance of heritage buildings and neighborhoods can engender a public effort towards preservation—an effort that should come with opportunities for the enjoyment of the focal heritage (De Guichen and d'Iteren 2009; Vaez Afshar et al. 2021; Afshar et al. 2022).

In this regard, historical game studies with a multitude of themes, flourishing year by year (Wright 2018), can provide both education about the historical issues at stake and provide entertainment focused on preservation. Following the popularity of the games, the historical game studies field recognized the importance of outreach within the domain.
Hence, within the last fifteen years, video games experienced an explosion of enthusiasm towards including history as their content (Bierstedt 2022). Additionally, since the appearance of historical game studies, in general, game studies, the pedagogical potential for learning has been pointed out within video games. Also, people have adapted to entertain, learn, and acquire skills via interacting and engaging with a variety of interactive contexts and multimedia content (Redder and Schott 2022).

In historical video games, which are a game genre whose gaming objectives are based on the past, it is assumed that the player gains key historical experiences of the setting, cultures, societies, and periods. While the traditional written or visual historical contents are not directly replaceable with mentioned historical video games, their interactiveness and multimodality allow the players to freely explore the historical places, spaces, artifacts, and inhabitants and experience the practices and rituals through hearing sounds and languages. Since historical video games are beyond linearly moving images associated with a narrative, players need to understand them (Redder & Schott 2022; Varinlioğlu et al. 2022). The puzzles, challenges, and obstacles required to be solved for proceeding within the game are the gamic features of the 2D or 3D worlds within the video games (Adams & Rollings 2010; Vaez Afshar et al. 2022). Hence, the game system's parameters still shape a player's agency despite video games being highly interactive (Aarseth 2007). As historical video games both function as a descriptive historical text and a systematic form of play guiding attention, learning, and activity, generating an inevitable efficient platform for presenting and exploring various histories. Since the emergence of historical video games, they have reproduced and allowed a chance to experience a wide range of histories (Redder & Schott 2022).

The carried research in the historical video games field demonstrates the concerns of scholars exploring the intersection of history and games about the form of historical games, the effect of false storytelling, and the gamic representation analyses focusing on the accuracy and the authenticity of the historical content provided within the games (Elliott 2017; Morley 2020; Rollinger 2020). However, this study explores various aspects of historical video games via web scraping and data mining, focusing on the players' notions. The mined features include the games' compatibility with head-mounted display (HMD) devices providing a virtual reality (VR) experience, player scale, dimension, and genre. It aims to figure out the characteristics effective in the player rating within the analyzed games. The ultimate goal of the analysis is to provide a framework applicable to the development of commercially playable historical video games contributing to heritage preservation by raising public awareness in the context of public history. Bellow follows a literature review exploring the existing historical video games and the investigations that scholars have conducted in this field.

1.0 HISTORICAL VIDEO GAMES

Many studies in the literature are investigating the developed and commercially played historical video games to figure out the contents carried in them, the way they convey these contents, the reasons these games got popular among gamers, their educational aspects, and so on.

Redder and Schott (2022), in their study, mentioned several historical video games before their main case study. Due to Redder and Schott (2022), current hit historical games like Assassin’s Creed: Origins (2017), Red Dead Redemption 2 game occurring in America’s wild-west period setting (2018), and Total War: Three Kingdoms (2019), sold respectively 155 million copies till 2020, 38 million till 2021, and over one million in the first week following their release dates. These games carry an educational value beyond barely written content by demonstrating authentic and periodically credible historical elements. Civilization series (1991-2016) and Humankind (2021) are samples of historical games reflecting histories involving resource management and global strategy in civilization development. As another type, Kingdom Come: Deliverance (2018) and Titanic: Honor and Glory games provide vast explorable worlds to the players as role-playing games to experience the documented histories by scholars. On the other hand, Wolfenstein ii: The New Colossus (2017) is a sample touching history indirectly in alternate timelines. In their main case study, A Plague Tale: Innocence (2019), they used a multimodal approach exploring both the presented historical content in game design and the gameplay. They focused on imaginative history in their research, including fiction occurring within period-accurate elements. The researchers introduce this game as valuable for its capability in history education regarding the re-mediation of past concerning beliefs, i.e., the Black Death.

In another attempt, Bierstedt (2022), in his study, in addition to mentioning the common existing outreaches focusing on consultancy, education, and new game developments in the historical video games field, noted the importance of the fourth category, digital criticism, in the form of Let’s Play. Due to him, while a historian can analyze historical video games using this method as a tool, its potential is yet underestimated. Hence, he tried to fix the issue from two points of view; first, by highlighting the requirement of various outreach methods via case studying the recent high-cost games, and second, by emphasizing the advantages of digital platform outreaches theoretically and practically with some experiments. He also discusses the streamer-historian category definition. These historians critique historical video games by playing and commenting in the form of videos or texts.
2.0 RESEARCH

2.1. Data collection
As previously mentioned, this research will investigate historical video games based on their attributes focusing on the players’ ideas and reviews regarding the games. The research undertaking is based on the extensive video game digital distribution service database, Steam (Steam 2022). Since this research is player-oriented, we used the tagging system and the user review information pool of Steam. Tags are a feature of this database that either the developer, the players, or the Steam moderators can apply to a game. These tags, which include the themes, terms, and genres, allow them to markup games for a better description of the game to others (Steam Tags 2022). Additionally, Steam users who have playtime in a specific game are allowed to add reviews to it and share their experiences with others. Steam uses a combination of positive and negative reviews to calculate a review score (Steamworks 2022). These reviews have a spectrum including overwhelmingly positive, very positive, positive, mostly positive, mixed, mostly negative, negative, very negative, and overwhelmingly negative, which are also related to the review number (Gamedeveloper 2022). Hence, by implementing the historical tagging filter and sorting them based on user review, we ended up with a sample of 1,148 games produced in the period of 2002 to 2022, which are historical based on the developers or the players.

In the next step, the data gathering, we applied web scraping using Octoparse (Octoparse 2022), a modern visual web data extraction software. Via Octoparse, both experienced and inexperienced people can extract bulk information from various websites with almost no coding needed in a time and effort-efficient manner. In this software, by entering the required website URL, it is possible to click on the target data and run the extraction to get the web page’s information as a spreadsheet (Ahamad et al. 2017). To do so, we first extracted basic info like the game title and release date. Then, we checked the most used tags within these games to decide which attributes’ information we wanted to gather. Among those tags, we selected the genres. Game genres are understandable categories identifying the gameplay, and they can be both combined with other genres and can be divided into sub-genres (Clearwater 2011). On the other hand, due to Heintz and Law (2015), while it is hard to distinguish the difference between game genre and game type, the first one is defined as gameplay style and artistic design, and the other as gameplay mechanics and functionality and architecture. However, in practice, they may be used conversely. Hence, to determine the genres among the tags, we both used Steam’s categories (Steam Tags 2022) and the intersection of several papers defining the genres (Apperley 2006; Arsenault 2009; Clearwater 2011; Elliott et al. 2012; Heintz & Law 2015). Consequently, we chose RPG, Action, Adventure, Simulation, and Strategy as the investigating genres, defined in Table 1.

**Table 1**: Genre definitions. Source: (Apperley 2006)

<table>
<thead>
<tr>
<th>Genre</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Action</td>
<td>It is usually extensively performative, differing from other performative genres, requiring the player’s engagement in extreme nontrivial actions to make the ergodic traversal. Additionally, the player should select the appropriate inputs to perform a required action for the game.</td>
</tr>
<tr>
<td>Adventure or RPG</td>
<td>They are closely connected to the literary genre of fantasy. However, in these games, the players take a character model and act in that environment without real role-playing. They try to suggest something to the computer that is not in the program.</td>
</tr>
<tr>
<td>Simulation</td>
<td>It simulates flying, driving, sports, and the town, city, and small community dynamics.</td>
</tr>
<tr>
<td>Strategy</td>
<td>Divided into real-time and turn-based strategy subgenres, it has a general god’s eye-view towards the taking placed actions generated from more photorealistic depictions. It remediates the strategy tabletop board games playing.</td>
</tr>
</tbody>
</table>

Additionally, we gathered data regarding the games’ player scale as single or multiplayer, dimension as 2D or 3D, VR compatibility, and rating as a percentage and a positive and negative range. Finally, we combined all the Excel sheets and prepared the final data set. However, due to the inaccuracy of percentage data, in which the number of people rated is effective, we removed the percentage data from the final list (Table 2).
2.2. Data mining

The current era faced insufficient data gathering with highly accelerated data aggregation growth. Through data mining, it is possible to figure out rules or patterns among this huge bunch of data leading to the detection of new information (Andry et al. 2022). Hence, in this study, during the analysis stage of the gathered information set, we used a data mining and machine learning tool, RapidMiner Studio (Rapidminer 2022), for data preprocessing and visualization. The tool has both a Free Open Source and a commercial edition. RapidMiner, which can also be integrated with different programming languages like Python and R, includes standard data mining features ready to use, with no coding required.

In their study, Dol and Jawandhiya (2022) compared several free, open-source data mining tools like KNIME, Keel, Weka, RapidMiner, Orange, and Tanagra. They also described the tools mentioned and their use in various papers via applied data mining techniques such as classification algorithms like Decision Tree, Naïve Bays, and Random Forest, clustering algorithms like K-means, Agglomerative, and association rule algorithms like Apriori, besides the performance parameters. In other researches, the authors used RapidMiner and its data-mining techniques to analyze insurance data, sales transactions in a company, and work accidents data to achieve critical information and predictions (Andry et al. 2022; Ginting & Iqbal 2022; Sani et al. 2022).

The current research uses data mining of the player-generated historical game information in Steam to figure out attributes influencing rating patterns. Firstly, to examine the correlation between the social rating data and all other attributes, the Mutual Information Matrix model is used (Figure 1).

![Figure 1: Mutual Information Matrix process. Source: (Author 2023)](image)

Mutual information between two attributes shows their dependence on each other as a quantity. It measures how much data each attribute gives about the others and is a dimensionless quantity. This model’s result demonstrates that the dimension and the strategy genre attributes affect the rating most among the others (Figure 2).
Figure 2: Mutual Information Matrix model result. Source: (Author 2023)

Then, for data modeling, we applied a Decision Tree to the dataset after assigning the label role to the rating attribute using the Set Role model (Figure 3).

Figure 3: Decision Tree Process. Source: (Author 2023)

A decision tree is generated from the collection of nodes in the form of a tree to make an estimate of a numerical target value or a decision on values affiliation to a class. However, to do so, one of the attributes should be defined as a label attribute allowing the Decision Tree to generate the results based on that. In other words, the role of an attribute explains how other applied Operators handle that specific Attribute. Hence, the label role causes the rating attribute to act as a target Attribute for learning Operators (Figure 4).

Figure 4: Decision Tree model result. Source: (Author 2023)

Analyzing the results, the Decision Tree shows that while most of the games with very positive ratings are strategy games released after 2020 (Figure 5), the others are games published between 2007 and 2020, compatible with VR technology (Figure 6).
However, most of the games with mixed rating results appeared as single-player simulation games with no VR compatibility, which were released after 2007 (Figure 7).

Regarding the above analytic applications, we visualized the data to see how the attributes are associated with the rating and release date attributes. To have a more understandable result, we converted the rating data to numerical data ranging from 1 for overwhelmingly negative and 9 for overwhelmingly positive. Due to the results, in general, each year, the number of historical video games is raised from 2002 to 2022. Also, we can see a general enhancement in the rating of the games from the past till now.

Figure 8 depicts that while Simulation, Strategy, and Action genres have been used in historical video games since the early 2000s, the RPG genre entered the historical video game field later, after 2014. Additionally, due to the ratings, Action and Strategy historical games are more popular among the players despite the PRG and Simulation ones.
Additionally, the results regarding the dimension, player scale, and VR compatibility show that the HMD feature allowing the VR experience is emerged since 2014 and is not well-settled yet in the field (Figure 9). Additionally, single-player games started to be developed a few years after multiplayers and are remarkably more popular and better rated than multiplayer. 2D and 3D games have been distributed almost equally over the years. However, surprisingly, the 2D ones are slightly trendier.

CONCLUSION
To conclude, historic preservation allows people to find the important parts of history valuable to be preserved for the next generations and conserve them. Due to the high concentration of the historic properties within the urban areas, historic preservations require an urban-oriented approach. Additionally, public awareness and enlightened decision-making enhance this process. Hence, public history, differing from the academic one, in terms of audience and presentation could serve to this issue. Historical video games are a great sample of public history implementation. These games which their objectives are based on the past, let the player gain historical experience regarding the setting, culture, society and periods.
This study explores the existing 1,148 historical video games published between 2002 and 2022 within the Steam database regarding their genres, player scale, dimension, and VR compatibility through web scraping with Octoparse the data and mining in via RapidMiner Studio to figure the effect of these attributes on the player rating (Figure 10).
Due to the results, since the general number of historical video games and their rating amount is getting more as time goes on, it is assumed that historical video games are getting better in terms of quality, or it could also be related to the rise of people's interest in playing historical games. Also, the late emergence of VR games within the historical video game field and their low rating means that VR technology has great potential within this field and needs to be paid more attention to by historical video game developers. Finally, field experts can use this study's results to enhance the quality and popularity of these games among players. Consequently, when used as a tool for education and public awareness, architects and designers can help to preserve and promote the architectural heritage of different cultures and periods. Therefore, the results of this study can be of great value to architects and designers interested in the history and preservation of architectural importance. By understanding the trends and potential of historical video games, they can create new and innovative approaches to architectural design that reflect the historical context and cultural significance of different buildings and structures.

REFERENCES

Dol, S. M., and Jawandhiya, P. M. 2022. Use of Data mining Tools in Educational Data Mining. In 2022 Fifth International Conference on Computational Intelligence and Communication Technologies (CICIT) (380-387). IEEE.


